Does VaR Tell The Truth?

by
Chong Lio [1]

An essay submitted to Department of Economics
in partial fulfillment of the requirements for
the degree of Master of Arts

Queen's University
Kingston, Ontario, Canada
August 2009

[1]M.A. Student, Department of Economics, Queen's University, Kingston, Ontario, Canada. I have benefited the comments from Professor Wulin Suo and Yin Ho Chow. All errors are my own.

Abstract

This paper explores three models to estimate volatility: historical simulation (HS), exponential weighted moving average (EWMA) and generalized autoregressive conditional heteroskedasticity (GARCH). The volatility estimated by these models can be used to measure the market risk of a portfolio of assets, called Value at Risk (VaR). VaR depends on the volatility, time horizon and confidence interval for the continuous returns under analysis. For empirical assessment of these models, we used a sample based on four stock indices, two commodities and one treasury to specify the GARCH, EWMA and HS models. Additionally, we adjusted these models by violation Kupiec backtesting for one-day VaR, to compare the efficiency of the HS, GARCH and EWMA volatility models. The results suggest that VaR calculated considering HS performs relatively better at high confidence intervals compared to the other two approaches as it considers the extreme values that falls out of the normal distribution. GARCH approach with normal distribution constantly produces too many VaR breaks to be accepted by the Kupiec test, but GARCH with t-distribution performs better both on the 95% and 99% confidence level. Lastly, EWMA approach produces very similar results as GARCH since mean reversion does not play a crucial role in the sampling period.

Keywords: Value at Risk, GARCH, Historical Simulation, EWMA and Backtesting

Contents

List of Tables

List of Figures

1 Introduction

Amid the recent financial turmoil, estimating and controlling risks have become a vital topic in financial institutions. In a dynamic trading environment, they face credit risk, market risk, liquidity risk, solvency risk, operational risk and sovereign risk. Among these different risk issues and categories, market risk is the central piece faced by financial institutions since it estimates the uncertainty of future earnings due to the changes in market conditions and reflects the potential losses caused by the decrease in the market value of the portfolio.

Among the market risk measurement methods, Value at Risk (VaR) has been adopted as the standard measure by financial institutions (BIS, 2001). There are three main reasons to spread the popularity of VaR. Firstly, J.P. Morgan made their RiskMetrics database freely available to the public in 1994. Secondly, the climate created by the derivatives disasters such as Procter and Gamble, Kidder Peabody, Orange County and Barings. Lastly, in 1996, the Basle Committee on Banking Supervision (BCBS) issued an amendment to the Capital Accord of July 1988 to use VaR for measuring market risk to set capital requirements for banks (BCBS, 1996). For instance, the European Union's Capital Adequacy Directive makes the VaR of the market risk in a bank's trading book as part of calculation of their capital reserve requirement. The VaR approach, which was strongly recommended in the July 1993 study by Group of Thirty, has become a benchmark for managing all financial risk. Since then, many national regulatory authorities have adopted the BCBS recommendations. Moreover, VaR has also gained strong support as an industry standard in various forms of academic literature (for example Christoffersen et al., 2001; Jorion, 2001; Heffernan, 1996; Santomero, 1997).

Calculating VaR is a very easy and intuitive concept but its measurement is a very challenging statistical problem. Even though there exists numerous methods in calculating VaR, they all follow a common structure, which can be summarized in three steps: mark-to-market the portfolio, estimate the distribution of the portfolio returns and compute the VaR of the portfolio. The VaR can be described as below:

$$VaR = \sigma \times C \times \sqrt{T} \times dollars \tag{1}$$

To put it in a nutshell, it tells us that "We are C% confidence that we will not lose more than P dollars in the next T days." The variable P is the VaR, T is the time horizon, σ is the daily volatility and C is the confidence level. VaR summarizes the effects of leverage, diversification, and probabilities of adverse price movements in a single dollar amount that is easy to communicate with the senior executives.

Since the introduction of the simplest VaR models a little over 10 years ago, the range of techniques used to obtain VaR estimates has expanded both in number and in complexity. Yet, so far, there is no industry consensus on the best method for calculating VaR (Engel and Gizycki, 1999). As with any statistical model, VaR depends on certain assumptions and it is commonly regarded that there are four main approaches to calculate VaR: parametric (RiskMetrics and GARCH), non-parametric (Historical Simulation and the Hybrid model), semi-parametric (Extreme Value Theory, CAViaR and quasi-maximum likelihood GARCH) and the Monte Carlo simulation. The results of each method can be varied very significantly from each other. Beder (1995) applies eight common VaR methodologies[2] to three hypothetical portfolios and the results show the VaR estimate varying by more than 14 times for the same portfolio. Hence, in order to decide which methodology to choose from, it is necessary to understand the underlying assumptions as well as the mathematical models and quantitative methods used.

Each VaR method has its own set of assumptions and each is a simplification of reality. All methods have their own strengths and weaknesses, and together the differences between the approaches predictably results in different risk perspectives. The differences in common VaRs emphasize the fact that no single set of parameters, data, assumptions and methodology is accepted as the superior approach. Beder's study found that depending on the selection of the time horizon, database and correlation assumptions across asset classes, the same model may produce widely divergent VaR views for the sample portfolio. Therefore, the key is not the model itself, but the right selection of an appropriate method and its suitability to financial institutions' risk objectives.

This paper is organized in six sections. The first section introduces the main concepts of the parametric and non-parametric models. The second section presents the sample data. The third section then details the backtesting to compare the efficiency of the volatility prediction models used in calculating the VaR. The fourth section show some of the major limitations of VaR. The fifth section presents the final conclusion and the sixth section contains the daily volatility of each asset and its histogram.

[2]Extreme value theory and CA ViaR are not considered by Beder (1995) as they have only applied to VaR estimation recently.

1.1 Parametric Method

So far, there is no industry consensus on the best method for calculating VaR. As with any statistical model, VaR depends on certain assumptions and the choice of which method of calculation used is normally dictated by the user's risk aversion to unrealistic or over simplistic assumptions. The variance-covariance method is the simplest in terms of application to financial practices and computer time consumption. This method assumes that the returns on risk factors are normally distributed (or some other distributions such as the t-distirbution) and the correlations between risk factors are constant. For risk management purposes, using the normal distribution assumption is generally considered to be acceptable. Deviation from normality usually does not significantly alter the results of VaR calculations under normal market conditions. Within this method, a Gaussian distribution is essentially assumed and it also assumes that extreme price swings, such as market crashes, occur too rarely to contribute to an accurate picture of the likelihood of future events.

The choice of the confidence level also depends on its use. If the resulting VaRs are directly used for the choice of a capital cushion, then the choice of the confidence level is crucial, as it should reflect the degree of risk aversion of the firm and the cost of a loss of exceeding the calculated VaR numbers. The higher the risk aversion or the greater the costs, implied that a big amount of capital should be set aside to cover possible losses, and this consequently will lead to a higher confidence level. In contrast, if VaR numbers are only used to provide a firm wide yardstick to compare risks among different portfolios and markets, then the choice of confidence level is not that relevant.

There are two popular parametric methods in calculating VaR: Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (1,1) model and exponentially weighted moving average (EWMA) model. The simplest GARCH (1,1) model can be described as follows:

1. $u_t = \sigma_t \epsilon_t$
2. $\epsilon_t \sim i.i.d.(0, 1)$

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2}$$

where $\omega = \gamma V_L$, V_L is the long run average variance rate, u_t is the mean return and σ_t is the volatility. This model has two crucial elements: the particular specification of the variance equation and the assumption of the standardized residuals (ϵ_t) are independent and identically distributed (i.i.d.). The first element was inspired by the characteristics of financial data are leptokurtotic, which have heavier tails and a higher peak than a normal distribution. The assumption of i.i.d. standardized residuals is a necessary device to estimate unknown parameters. A further necessary step to implement any GARCH is the specification of any distribution of the ϵ_t. The most generally used distribution is the standard normal. Only after this extra distributional assumption has been imposed, it become possible to write down a likelihood function and get an estimate of the unknown parameters.

The GARCH (1,1) model is similar to the EWMA model except that, in addition to assigning weights that decline exponentially to past u^2, it also assigns some weight to the long run average volatility. A stable GARCH (1,1) process requires $\alpha + \beta < 1$. Otherwise the weight applied to the long term variance is negative.

On the other hand, the RiskMetrics approach as known as the EWMA can be describe as followed:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) u_{t-1}^2 \tag{3}$$

The EWMA model is a particular case of GARCH (1,1), where $\gamma = 0, \alpha = 1 - \lambda$ and $\beta = \lambda$. According to RiskMetrics it sets λ equal to 0.94 for updating daily volatility. The value of λ governs how responsive the estimate of the daily volatility is to the most recent daily percentage change. A low value of λ assigns more weight to the u_{t-1}^2 when σ_n is calculated. In this case, the estimated produced for the volatility on successive days are themselves high volatile. A high value of λ (i.e., a value close to 1.0) produces estimates of the daily volatility that respond relatively slowly to new information provided by the daily percentage change. The EWMA approach has the attractive feature that relatively little data need to be stored. When we get a new observation, we can calculate a new daily percentage change with the new data and the old value of the market variable can be discarded.

RiskMetrics also assumes that standardized residuals are normally distributed. The general finding is that these approaches (GARCH and RiskMetrics) tend to underestimate the VaR because the normality assumption of the standardized residuals seems not to be consistent with the behavior of the financial returns. In practice, variance rate do tend to be mean reverting. The GARCH (1,1) model incorporates mean reversion, whereas the EWMA model does not. GARCh (1,1) is therefore theoretically more appealing than EWMA model. In certain circumstances where the best fit value of ω turns out to be negative, the GARCH

(1,1) model is not stable and it makes sense to switch to the EWMA model.

1.2 Non-Parametric Method

Historical Simulation (HS) assumes the process generating the profit and loss remains the same over time. For instance, if one has a five year estimation period, one has to make the assumption that the profit and loss distribution remains the same over the whole period. One way to avoid these implication is to discount the observations according to how far back they occurred: the further the observations apart from current, the less weight they have.

Before we implement the model, we must understand its major assumptions and drawbacks. First, this method is logically inconsistent. If all the returns within the window are assumed to have the same distribution, then the logical consequence must be that all the returns of the time series have the same distribution. Second, the length of the window must satisfy two contradictory properties: it must be large enough in order to make statistical inference significant and it must not be too large to avoid the risk of taking observations outside of the current volatility cluster. For instance, assume the market is moving from a period of relatively low volatility to a period of relatively high volatility (or vice versa). In this scenario, VaR estimates based on the historical simulation methodology will be biased downwards (correspondingly upwards) since it will take some time before the observations from the low volatility period leave the window. Moreover, if we want to estimate VaR at a high confidence level (at 99% level), a possible disadvantage is that the approach requires a lot of data to perform well at this confidence levels. When estimating at a 99% confidence level, at least 100 observations are needed in order to have one extreme value in the left tail.

Finally, with this methodology, the heteroskedasticity of the residuals is discounted partly by the use of the recent one year of daily returns, but the implicit equal weight set on the sample does not account properly for the decay of the volatility lagged correlation. In other words, the heteroskedasticity is improperly discounted. Therefore, the apparent simplicity of the historical methodology hides important assumptions that are not justified by the empirical properties of the financial time series. It is likely that even well behaved data will exhibit some systematic changes over time, and this might lead us to prefer a somewhat short estimation period. Volatilities and correlations will change over time. The information contained in more recent observations would be more useful than the information contained in the older observations. A long estimation period would lead the newer, more useful information contained in more recent observations to be drowned out by the older, stale information in the earlier observations. The VaR estimate would be insensitive to new information and reveal little about changes in risk factors over time [3].

2 Data

In this paper, I am testing three different approaches: historical simulation, GARCH (1,1) and EWMA using daily data on four different stock indices, one treasury notes and two commodities between January 2000 and June 2009. The assets are S&P 500, Heng Seng index, Nikkei 225, FTSE, 10 year US Treasury Notes, Crude oil and Gold [4] and each market variable contains around 2400 daily observations.

The VaR will be estimated on a daily basis using two different confidence levels: 95% and 99% to see how the approaches perform on different assets. Different confidence levels fit different purposes and depends on the management's preference. Choosing a higher level of confidence will result in a higher VaR and vice versa.

Different characteristics of the assets returns such as the volatility, skewness and kurtosis can be critical criteria in selecting one approaches over the other on estimating VaR. This is especially true with the parametric approaches that assume the returns to be normally distributed. This is the central question to the paper that what is the relationship between the characteristics of the assets and the estimating methods and how does the relationship affect the estimated VaR?

3 Results and Interpretation

3.1 Descriptive Statistics

Method of moments can be a useful tool to capture the different characteristics of each asset. There are four key moments that we need to pay attention to and they are mean, variance (or standard deviation), skewness and kurtosis. Mean tells us the average return of an asset and standard deviation measures how far the daily returns apart from the mean. Skewness measures the imbalance in a distribution, that is, whether

[3] Hendricks (1996) provides evidence that this is exactly what happens with long estimation periods, i.e., ones of 1250 days

[4] All the data are collected from Yahoo! Finance

observations occur more frequently above or below the mean. The skewness of each profit and loss series to determine whether the skewness is significantly non-zero. Kurtosis is a measure of the ”peakness” of the data sample and how concentrated the returns are around their mean. If kurtosis equals to zero, it means the tails on the distribution follow a normal distribution. If kurtosis is negative, the distribution has thinner tails than a normal distribution. Finally, if it is positive, the distribution has fatter tails (also known as leptokurtosis). Furthermore, in time series analysis, autocorrelation is a major gauge of wether the data correlated with itself over time. Autocorrelation in financial time series data is measured by regression of the observed returns with a lagged version of themselves. Regressing the returns with themselves like this can be described as testing if it is possible to describe returns of today as a linear function of the returns from yesterday. The presence of autocorrelation means that the applied approach will have a poor fit to the actual data which leads the analyst to conclude that the returns of today cannot be accurately described as a linear function of the returns of yesterday.

In order to compare the special properties of the assets, a summary of the statistical characteristics of the asset return and its autocorrelation can be found below:

Table 1: Statistical Characteristics of The Asset Return

	Crude Oil	Gold	Treasury	S&P 500	Hang Seng	Nikkie	FTSE
Skewness	-0.273	-0.097	-0.348	-0.088	-0.029	-0.298	-0.090
Kurtosis	7.207	7.975	11.351	10.689	10.935	9.336	9.216
Daily Volatility	2.746%	1.184%	1.617%	1.418%	1.716%	1.656%	1.354%
Annual Volatility	43.598%	18.80%	25.669%	22.506%	27.234%	26.282%	21.491%
Average Daily Return	0.042%	0.053%	-0.025%	-0.0183%	0.002%	-0.029%	-0.018%

From the result above, all the assets have negative skewness and positive kurtosis which are expected from financial data. The mass of the distribution is concentrated on the right and has a long left tail. From the histograms, we can see that treasury, Nikkie index and crude oil have more extreme negative values than others. When we look at the kurtosis of each assets, all of them are positive which is a double confirmation that our sample data have fat tails. The daily volatility of most assets are range from 1.2% to 1.7% except for crude oil. One of the reason for the excess volatility of the crude oil is that more than one third of the world

Table 2: ACF, PACF and Ljung-Box Q test

Assets	Lag	AC	PAC	Q	$Prob > Q$
Crude Oil	1	-0.0292	-0.0292	12.0104	0.0002
Gold	1	-0.0070	-00070	9.1162	0.0023
Treasury	1	-0.0203	-0.0203	13.9756	0.0000
S&P 500	1	-0.0899	-0.0899	19.1500	0.0000
Hang Seng	1	-0.0174	-0.0174	10.7098	0.0013
Nikkie	1	-0.0337	-0.0337	12.6254	0.0002
FTSE	1	-0.0672	-0.0673	10.71	0.0011

Note: The null hypothesis is that the data are random.

supply is controlled by the Organization of Petroleum Exporting Countries (OPEC). The volatility is mainly due to market manipulation by a small number of players in the market. When the spot price falls below their target price, they can reduce the supply in order to prop up the prices. Since the market participants can predict the actions of OPEC, when OPEC makes the announcement public, the commodity traders and hedge fund managers can take advantage of the announcement and speculate a cut in production in order to make profit. On the other hand, gold has relatively low volatility due to its stability and tradition role in storage of value and inflation hedge. When the market experience extreme price swing, market participants favor the stability of gold. From the histogram, we can see most of gold's daily returns are concentrated around zero and range from -0.5% and 0.5%.

I have also conduct a hypothesis test on wether there are any first order autocorrelation for each asset. A Ljung Box test is used to test whether anyone of them exhibit autocorrelation. The presence of autocorrelation and heteroskedasticity in the data is obvious on the assets. This finding is troubling for VaR models based on normality assumption, as well as for the nonparametric approaches that are based on the IID assumption, such as the historical simulation. This is very indicative for risk managers because elementary assumptions of many VaR models are not satisfied, meaning that VaR figures obtained for such models cannot be completely trusted. All assets are found to be strongly non-normal. Since a variance-covariance VaR model is used by most banks, the finding of non-normality implies that the VaR measure will not correctly estimate the true risk exposure. In particular, since the profit and loss distributions are quite fat-tailed, the VaR model would be expected to significantly underestimate risk.

3.2 Regulatory Backtest

Table 3: Basel Committee's Three Zones

	Numbers of Exceptions (252 days)	Multiplication Factors
Green Zone	4 or less	3.0
Yellow Zone	5	3.4
	6	3.5
	7	3.65
	8	3.75
	9	3.85
Red Zone	10 or more	4.0

Under the capital adequacy arrangements proposed by the Basle Committee, each bank must meet a capital requirement expressed as the higher of: (i) an average of the daily VaR measures on each of the preceding sixty trading days, adjusted by a multiplication factor; and (ii) the banks previous days VaR number. The multiplication factor is to be set within a range of 3 to 4 depending on the regulators assessment of the banks risk management practices and on the results of a simple backtest (Basle Committee on Banking Supervision 1996).

The multiplication factor is determined by the number of times losses exceed the days VaR figure (termed exceptions or VaR break) as set out in the above table (Basle Committee on Banking Supervision 1996). The minimum multiplication factor of 3 is in place to compensate for a number of errors that may arise in model implementation. Over simplified assumptions, analytical approximations, small sample biases and numerical errors will tend to reduce the true risk coverage of the model (Stahl 1997). The increase in the multiplication factor is designed to scale up the confidence level implied by the observed number of exceptions to the 99 per cent confidence level desired by regulators. In calculating the number of exceptions, banks will be required to calculate VaR numbers using a one-day holding period, and to compare those VaR numbers with realized profit and loss figures for the previous 252 trading days. A simple approach to exceptions-based backtesting would be to assume that the selected data period provides a perfect indication of the long-run performance of the model. For example, if a VaR model was supposed to produce 99th percentile risk estimates, observed exceptions on any more than 1 per cent of days could indicate problems with the model. This is not realistic since with a finite number of daily observations; it is quite probable that the actual number of losses exceed VaR estimates will differ from the percentage implied by the models confidence interval, even when the model is in fact accurate. Hence, the Basle approach is to allocate banks into three zones based on the number of exceptions observed over 252 trading days. A model, which truly covers a 99 per cent confidence interval, has only a 5 per cent chance of producing more than four exceptions (yellow zone), and only a 0.01 per cent chance of producing more than ten exceptions (red zone).

3.3 Backtesting

A backtest can be done by comparing the historical VaR forecasts with their associated subsequent returns. It is a formal statistical framework that consists of verifying the estimated losses in line with the actual losses. The simplest form of backtesting is by counting the number of the exceptions (the actual losses are greater than the estimated VaR which is called a VaR break) for a given period and chosen confidence level.

However, backtesting also has its shortcomings. The most fundamental problem arises from the fact that backtesting attempts to compare static portfolio risk with a more dynamic revenue flow. VaR is measured as the potential change in value of a static portfolio, at a specific point in time (typically end-of-day). Hence, the VaR calculation assumes that there is no change in the portfolio during the holding period; the portfolio can be viewed as representing a stock of risk at a given point in time. In practice, banks portfolios are rarely static, but change frequently. Profits and losses are flows accruing over time as a bank takes on and closes out positions reflecting changes in portfolio composition during the holding period.

The difficulties of a dynamic portfolio can be illustrated by considering a trading desk that is not permitted to hold open positions overnight. During the day the desk may take positions and as a result experience large swings in profit and loss, but at the end of each day all positions must be closed out. Hence, an end-of-day VaR will always report a zero risk estimate, implying zero profit and loss volatility, regardless of the positions taken on during the day. More generally, where open positions remain at the end of the trading day, intra-day trading will tend to increase the volatility of trading outcomes, and may result in VaR figures underestimating the true risk embedded in any given portfolio.

To overcome this problem of dynamic portfolios, a backtest could be based on a comparison of VaR (using a one-day holding period) against the hypothetical changes in portfolio value that would occur if end-of-day positions were to remain unchanged. That is, instead of looking at the current days actual profit or loss, the profit or loss obtained from applying the days price movements to the previous end-of-day portfolio is calculated. This hypothetical profit or loss result could then be compared to the VaR based on the same, static, end-of-day portfolio. In such a case, the risk estimate and the profit and loss would directly correspond.

3.4 Kupiec Test

Kupiec (1995) presents a more sophisticated approach to the analysis of exceptions based on the observation that a comparison between daily profit or loss outcomes and the corresponding VaR measures gives rise to a binomial experiment. If the actual trading loss exceeds the VaR estimate the result is recorded as a failure (exception or VaR break); conversely, if the actual loss is less than the expected loss (or if the actual trading outcome is positive) the result is recorded as a success. Jorion (2001) state that the number of VaR breaks is expected to be the same as one minus the level of confidence level. For instance, for a sample of 100 observations where a 95% confidence VaR is calculated, we would expect five VaR breaks to occur. If there are more or less VaR breaks than expected, it implied the deficiencies of a specific approach. The Kupiec test uses the binominal distribution to calculate the probability that a certain number of VaR breaks will occur given a certain confidence level and sample size. The Kupiec test function is:

$$Pr[x|n,p] = (nCx)p^x(1-p)^{n-x} \tag{4}$$

where x is the number of VaR breaks, n is the sample size and p is the confidence level. The binomial function produces the likelihood that a specific number of VaR breaks is to occur. By using the cumulative binominal distribution, it is possible to calculate an interval within which the number of VaR breaks must fall. VaR approaches produce values of n that lies within this range can be accepted. If the approach produces values of n outside the range, the approach is rejected. A rejection means that the confidence level that one used in the VaR approach does not match the actual probability of the VaR break.

3.5 Historical Simulation

The backtesting results for the seven assets, which calculate by using the historical simulation approach, are shown below. The table shows the minimum and maximum values allowed by the Kupiec test, the target number of VaR breaks and the resulting number of VaR breaks from the simulation. If the resulting number of VaR breaks is within the range, the estimating method and number are meaningful. Otherwise, we need to use other method to recalculate VaR.

Table 4: Kupiec Test: Historical Simulation 252 days

	Observations	95%				99%			
		Min	Target	Result	Max	Min	Target	Result	Max
Crude Oil	2356	111	118	137	135	17	24	30	31
Gold	2372	112	119	129	136	17	24	33	31
Treasury	2362	112	118	149	134	17	24	37	31
S&P 500	2366	112	118	129	134	17	24	31	31
Hang Seng	2338	110	117	114	134	16	23	31	30
Nikkie	2310	108	115	121	132	16	23	35	30
FTSE	2377	112	119	113	136	17	24	32	31

Table 5: Kupiec Test: Historical Simulation 504 days

	Observations	95%				99%			
		Min	Target	Result	Max	Min	Target	Result	Max
Crude Oil	2356	111	118	127	135	17	24	44	31
Gold	2372	112	119	127	136	17	24	31	31
Treasury	2362	112	118	184	134	17	24	42	31
S&P 500	2366	112	118	116	134	17	24	30	31
Hang Seng	2338	110	117	109	134	16	23	41	30
Nikkie	2310	108	115	112	132	16	23	33	30
FTSE	2377	112	119	111	136	17	24	35	31

I have computed the results using two window time: one with 252 trading days and another one with 504 trading days. The historical simulation is performing well on the 95% confidence level with both windows, but fewer number of VaR breaks occurs with 504 observation window. At the 99% confidence level, we see mixed results. The tables above show that the approach produces too many VaR breaks and indicates that it overestimates the VaR. The reason why the approach produced too many VaR breaks is that the historical simulation approach assumes the distribution of returns does not change over time. Without this assumption, there would be no reason at all to look at the past returns in hope of predicting the future. In other words, one contributing cause of error in the approach would be that the distribution is nonstationary.

The next thing we need to look at is how the volatility is being calculated. While we are calculating the volatility of the latest 252 and 504 daily returns, there is a tradeoff between the importance of new information and old historical information. If we pick a longer interval, the approach react slowly to new information and changes in the daily return. If a shorter interval is chosen, the would have increased the impact of new observations and put more weight to the most recent values. Among all the assets, the estimates on treasury perform the worst in the lower confidence level for both window size. On the other hand, the results for crude oil with these two observation windows have very different VaR breaks, 44 with 504 days and 30 with 252 days. This can be explain by the volatility clustering. When we are using 504 days, our estimation includes more older observations. Especially during the subprime mortgage crisis starting in August 2008, the price of crude oil experience extreme price movement. It shifted from a period of relatively low volatility to a period of high volatility, in some extreme cases, it can fluctuate more than 10% a day. It is one of the reason to explain this unusual result. The truth is that the non parametric approach could have produced a better result if we had chosen to shorten the amount of days that the volatility was based on. By doing so, the extreme daily volatility in the middle of the 2008 would have had less impact of the result as the historical simulation is an approach that responds very slowly to changes in volatility. Over all this approach performs best on high confidence levels and this is because the chosen historical window size suited for these confidence levels. The reason that the approach performs relatively better at high confidence intervals compared to the other two approaches is that it considers the extreme values that falls out of the normal distribution. This can be clearly seen in the histograms in the graph section showing the daily returns are negatively skewed.

Table 6: Kupiec Test: GARCH(1,1) with normal distribution

	Observations	95%				99%			
		Min	Target	Result	Max	Min	Target	Result	Max
Crude Oil	2356	111	118	130	135	17	24	33	31
Gold	2372	112	119	125	136	17	24	37	31
Treasury	2362	112	118	146	134	17	24	35	31
S&P 500	2366	112	114	116	134	17	24	45	31
Hang Seng	2338	110	117	109	134	16	23	38	30
Nikkie	2310	108	115	129	132	16	23	37	30
FTSE	2377	112	119	120	136	17	24	40	31

Table 7: Kupiec Test: GARCH(1,1) t-distribution with 6 d.f.

	Observations	95%				99%			
		Min	Target	Result	Max	Min	Target	Result	Max
Crude Oil	2356	111	118	126	135	17	24	31	31
Gold	2372	112	119	122	136	17	24	31	31
Treasury	2362	112	118	132	134	17	24	32	31
S&P 500	2366	112	118	122	134	17	24	35	31
Hang Seng	2338	110	117	126	134	16	23	31	30
Nikkie	2310	108	115	119	132	16	23	32	30
FTSE	2377	112	119	115	136	17	24	37	31

3.6 GARCH(1,1)

The GARCH approaches produces good results on the 95% confidence level with exception of the 10 years treasury. However, the results for the 99% confidence level deviate from the range. The number of VaR breaks are much higher than the theory predicted based on the normal distribution of the asset return. If we assume the daily returns follow a t-distribution with 6 d.f., it can capture the fat tail distribution of the financial assets. From the results above, it outperforms the result with normal distribution. GARCH approach constantly produces too many VaR breaks to be accepted by the Kupiec test, but GARCH with t-distribution performs better both on the 95% and 99% confidence level. The Kupiec test reject all of the assets at the 99% level with a normal distribution. The reason why the approach produced too many VaR breaks is probably the assumption of normality is acceptable in the lower confidence level but not on higher one. The further out in the tail of the distribution of the returns we come, the less acceptable the assumption of normality becomes. With a t-distribution assumption, the Kupeic test fail to reject the results on crude oil and gold which can think of a non-normal distribution capture the behavior of the commodity price movement better.

All in all, the GARCH approach does a good job of handling volatility clustering and estimating a variance forecasts. However, the maximum likelihood function does not work proficiently and it has to adjust to better fit the assets return in order to further increase the performance of the GARCH approach. According to Goorberg and Vlaar (1999), the most important return characteristic when calculating VaR is volatility clustering. From the result above, I do not agree that volatility clustering is the most important characteristic. In my opinion the proper assumption of the return distribution would play a more critical in estimating VaR.

3.7 EWMA

Table 8: Kupiec Test: EWMA with Normal Distribution

	Observations	95%				99%			
		Min	Target	Result	Max	Min	Target	Result	Max
Crude Oil	2356	111	118	119	135	17	24	35	31
Gold	2372	112	119	126	136	17	24	30	31
Treasury	2362	112	118	152	134	17	24	37	31
S&P 500	2366	112	118	123	134	17	24	32	31
Hang Seng	2338	110	117	115	134	16	23	35	30
Nikkie	2310	108	115	119	132	16	23	30	30
FTSE	2377	112	119	111	136	17	24	37	31

According to J.P Morgan, the volatility updating factor λ is set equal to 0.94 as this value of λ gives forecasts of the variance rate that comes closest to the realized variance rate. At 95% confidence level, EWMA method produces very similar result as the GARCH(1,1) model because the mean reversion does not play a crucial role in this sampling period. The mean of most assets are very close to zero percent and also the parameter, ω is non zero. The EWMA approach also suffer from the same deficiency as the GARCH approach assumes

Table 9: Kupiec Test: EWMA t-distribution with 6 d.f.

	Observations	95%				99%			
		Min	Target	Result	Max	Min	Target	Result	Max
Crude Oil	2356	111	118	112	135	17	24	33	31
Gold	2372	112	119	129	136	17	24	32	31
Treasury	2362	112	118	149	134	17	24	39	31
S&P 500	2366	112	118	130	134	17	24	36	31
Hang Seng	2338	110	117	121	134	16	23	34	30
Nikkie	2310	108	115	120	132	16	23	29	30
FTSE	2377	112	119	115	136	17	24	36	31

the returns to be normally distributed. This becomes apparent at the 99% confidence level where the EWMA produces too many VaR breaks to be accepted by the Kupiec test. Even if we assume the return distribution follows a t-distribution with 6 d.f., it does not provide superior results than the normal distribution. Although it is surprising the EWMA approach with t-distribution does not enhance the performance of the model, it is important to point out that these results are only indicative for a sample, and by comparing them it is not possible to conclude which distribution is more efficient.

4 Limitations of VaR

After we observe all the statistical results on various methods, we can conclude it is unwise to seek a method of computing VaR to be both accurate and available on a time basis since there is likely to be an inherent trade off between these objectives since more rapid methods tend to be less accurate (Pritsker, 1996). It may be argued that the choice of market risk measures can be examined from hierarchical perspective. Financial institutions would first structure various selection criteria into a multi-level hierarchy, with managers identifying the appropriate selection criteria ranked by those factors influencing their choice. The more appropriate criteria frameworks are used, the better the fit, and the best choice will be determined by which dimensions the risk manager considers most important. Therefore, it should be recommended that knowing the factors that are central considerations in the model choices are of critical important to market risk practitioners.

Secondly, the actual situation and resources available restricts the freedom of selection. For instance, the some cases not enough historical data can be utilized to easily communicate with senior management through a historical simulation VaR report. Therefore, a best selection and use of risk measures should ensure that maximum available resources can be guaranteed. In particular, a strong database should be available, along with intensive IT backing, risk education and training, all of which within the discipline of a well developed risk infrastructure recommended by the project's participants.

Thirdly, the procyclical of VaR contributes to excessive risk taking in the fast growing period prior to a financial crisis. A procyclical capital framework is one that reinforces business cycles. A bank requires less capital when times are good which encourage greater risk taking and more capital as the economy contracts which constraining banks ability to lend and working against economic growth. Alongside this point, there is a demonstration that VaR based on short histories will produce such procyclical capital requirements. One of the solution is to use longer historical periods. The conflict between the needs of capital and the output of VaR has existed since VaR first became part of the capital regime. The regulators have responded to this by placing restrictions on VaR models, in an effort to embed into VaR's properties that are desirable for regulatory capital. The support for longer observation periods in order to eliminate procyclical is a continuation of this mind set. The recent proposal of the Basel Committee advocates adding some turbulent periods into the positions of today, in order to calculate a so-called stressed VaR. Even though this new restriction may be enhance the explain power of VaR, the trading book capital under the current regime is still flawed and further restrictions are unlikely to make it work in the future.

Fourthly, when choosing the criteria framework, financial institutions are further faced with deciding the number of determinants that should be taken into consideration. This is important because as more determinants are included, that fit will improve. However, at some point adding more criteria will result in over fitting, the fit will be better for the observed portfolio, but, on the other hand, will result in worse predictive performance. Thus, the variance of selection is a sum of two components: fit and feasibility. As more criteria are included, the fit increases, but the feasibility decreases. When a financial institution limits the number of determinants used for selection, the feasibility increases, yet the fitness decreases. Thus it is necessary to find the right trade-off between reducing variance and enlarging fitness.

Fifthly, the approach of model selection involves a dilemma between picking a single best model and using multiple measures. Philosophically, this might arise because one feels that there is a single best model, and the goal of selection is to identify that model. In some areas of application this approach may be justified, but in other areas one may doubt that an analyst could ever find an exact best model. For example, one may never find the best fit model to include all of the determinants in the selection framework. However, most selection is still required in order to find the closed approximation to the best model. For practical purposes, a financial institution may not need an exact fit model, but can find a model or models that they feel are sufficiently close to the best model. For instance, a financial institution may seek a model that can be easily reported to the senior manager and that simultaneously captures the fat tailed problem. In this case there will be no single best fit model, but rather a combination of two: one a historical simulation VaR which can be easily understood and the other such as extreme value theory or monte carlo simulation VaR which is dedicated for the fat tailed distribution. Sometimes, those model combinations that could be successful lead to the creation of a hybrid new method, such as the demonstrated in Boudoukh et al.(1998) and the RiskMetrics Technical Document. Even worse, if the combination plan fails or it is hard to find any single fitted model, risk managers have to face the arduous task of developing their own models. For financial giants who have a strong research capacity, they still have an alternative choice between devising new or adopting existing models, even if custom fitted models could be provided.

An obvious issue is the question of whether the model combination will produce a better result than the single method. Previous data analysis demonstrated that VaR should be best supplemented and combined with other market risk measures. Yet, experts are controversial in their views that combination might outperform the single model application. This is not a surprise as the existing literature not only lacks adequate research to this end, but also creates a troubling paradox. Although the necessity and benefits of pursuing the multiple model scheme is sometimes unquestionable, it should be noted that there are a number of problems such as the computation itself and data intensity, demanding great effort, and the difficulties together with the uncertainty as to whether its use can outperform the single model method precludes many risk managers from attempting the multi-model initiative. It may be argued that only through careful use and successful solution of these problems then the multivariate model framework can be a better application.

Another issue that must be addressed for the multiple model approach is how to utilize the multi models' outcomes. To a further extent, how to solve the problem that different models yield different results for an identical portfolio. It will be necessary to further decide about the key and supplementing models, the degree by which the assistant model should be utilized, as well as when and where it should be used. Thus it may be recommended that if a multi-model is to be used then a careful designed plan should be employed.

Last but not least, this paper suggests that the selection framework should be dynamic rather than deterministic. It has been argued that financial institutions should define the selection framework in and according to a given environment. It is obvious that the environment is continuously undergoing change, the change coming from the extra-organizational environment such as technology change, economic and market factors, and changes in the corporate risk culture. These factors require frequent alignment of the model selection framework on a timely basis. Take the emergence of the euro as an example. For a risk manager's favorite historical simulation, this may not be good news as the manager may need to change their risk model since no historical data may be available, particularly since the euro first emerged only in 1998. In fact, adapting to the environment change is itself a determinant for selection. Adapting these arguments, I suggest that the selection framework should be adjusted regularly, contingent on the type and level of environment uncertainty.

5 Conclusion

Empirical risk modeling forms the basis of the market risk regulatory environment as well as internal risk control. This paper identifies a number of shortcomings with regulatory VaR, where both theoretical and empirical aspects of VaR are analyzed. Based on the backtesting results it can be concluded that VaR models that are commonly used are not well suited for measuring market risk in the sampling period. The results suggest that VaR calculated considering HS performs relatively better at high confidence intervals compared to the other two approaches as it considers the extreme values that falls out of the normal distribution. GARCH approach with normal distribution constantly produces too many VaR breaks to be accepted by the Kupiec test, but GARCH with t-distribution performs better both on the 95% and 99% confidence level. Lastly, EWMA approach produces very similar results as GARCH since mean reversion does not play a crucial role in the sampling period. These findings bear very important implications that have to be addressed by regulators and risk practitioners. Risk managers have to start thinking outside the frames set as the results present in these markets may find themselves in serious trouble, dealing with losses that they were not expecting. Contrary to the widespread opinion, it is not enough to blindly implement the VaR models. Every VaR software package that a bank is thinking about implementing should be rigorously tested and analyzed to see if it really provides a correct estimate of the true level of risk a bank is exposed to. Regulators have to take into consideration that simplistic VaR models that are widely used are not well suited for the illiquid and excess volatile stock markets. For these reasons, it is imperative that before permission is given to banks to use internal VaR models that are either purchased or developed in-house, regulators should rigorously checks and analyze the backtesting performance as well as the theoretical framework of such model for any inconsistencies or unwanted simplifications.

6 Appendix

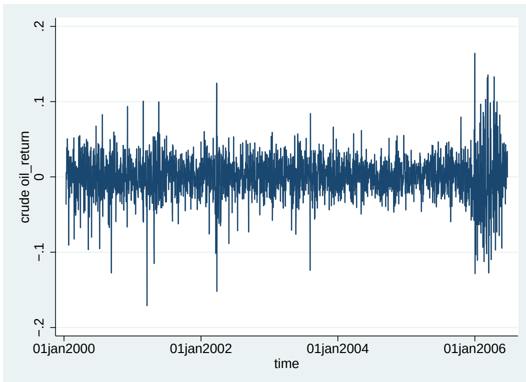

Figure 1: Daily Crude Oil Return: 2000-2009

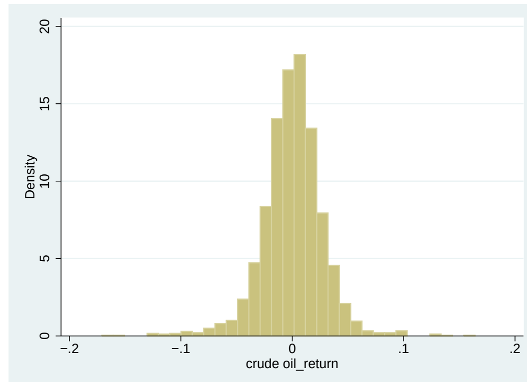

Figure 2: Histogram of Crude Oil return: 2000-2009

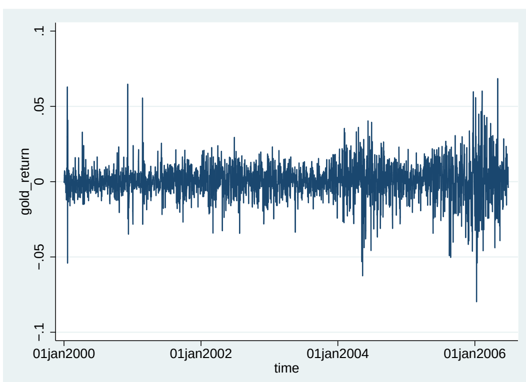

Figure 3: Daily Gold Return: 2000-2009

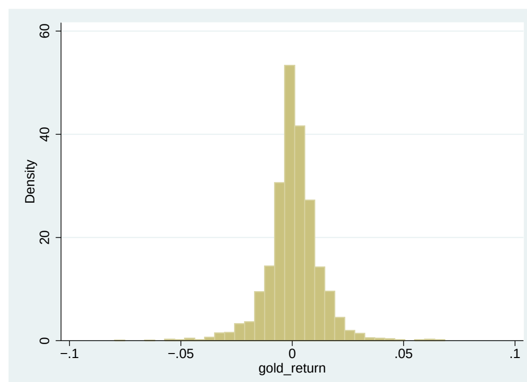

Figure 4: Histogram of Gold Return: 2000-2009

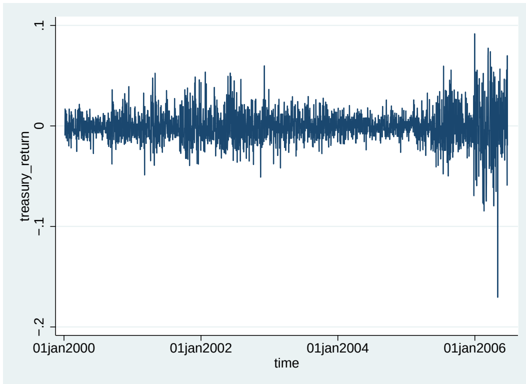

Figure 5: Daily 10 years US Treasury Return: 2000-2009

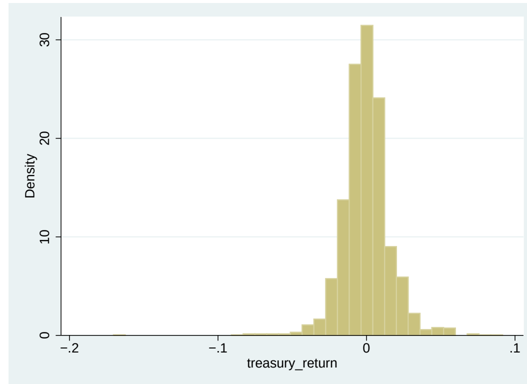

Figure 6: Histogram of 10 years US Treasury Return: 2000-2009

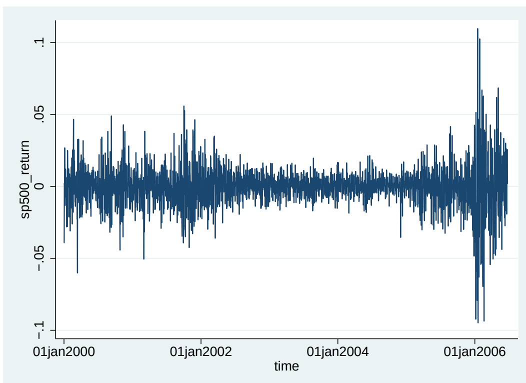

Figure 7: Daily S&P 500 Return: 2000-2009

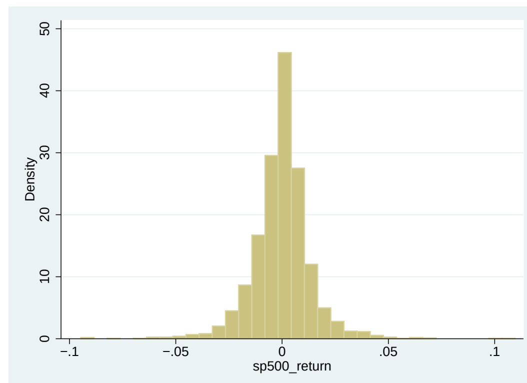

Figure 8: Histogram of S&P 500 Return: 2000-2009

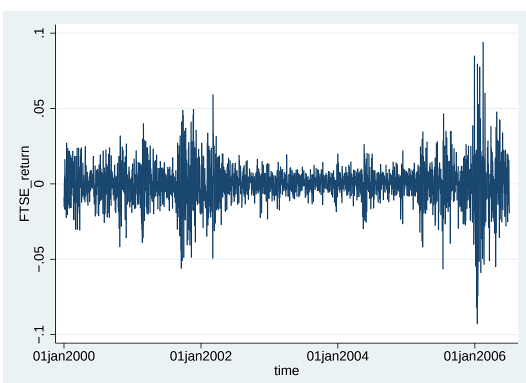

Figure 9: Daily FTSE Return: 2000-2009

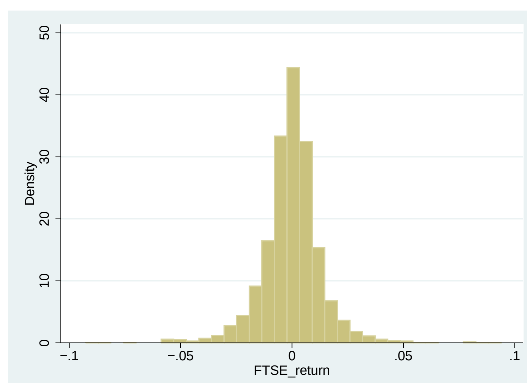

Figure 10: Histogram of FTSE Return: 2000-2009

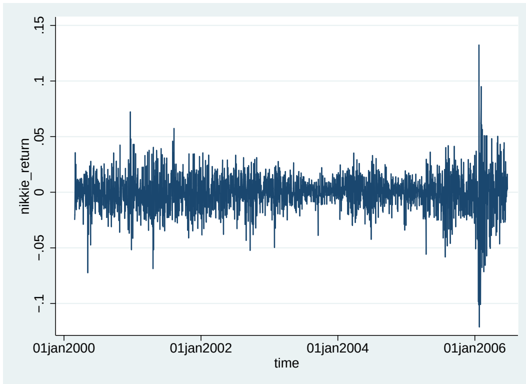

Figure 11: Daily Nikkie 225 Return: 2000-2009

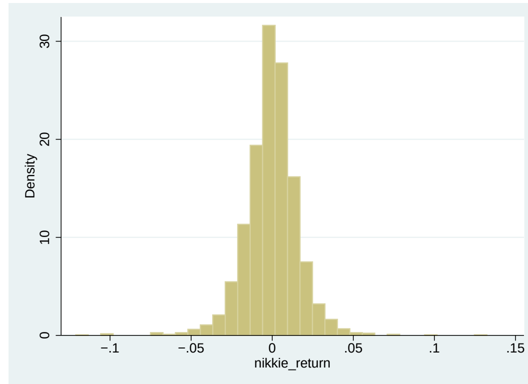

Figure 12: Histogram of Nikkie 225 Return: 2000-2009

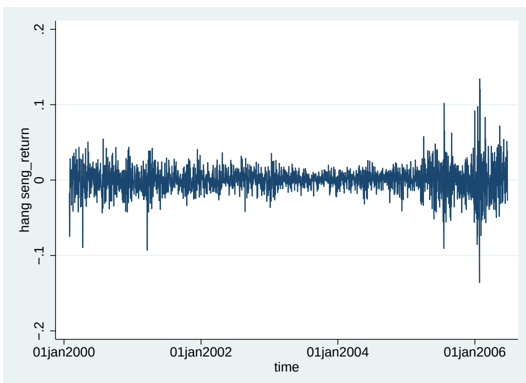

Figure 13: Daily Hang Seng Return: 2000-2009

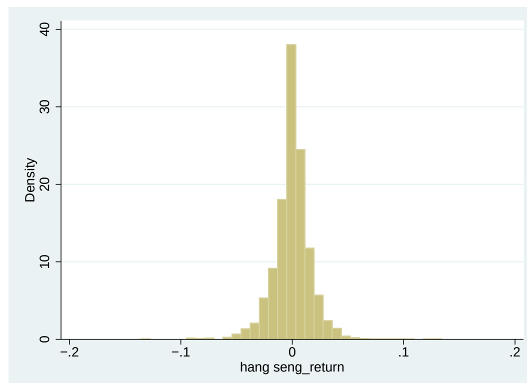

Figure 14: Histogram of Hang Seng Return: 2000-2009

References

Bank of International Settlement (1996) *Supervising Global Financial Markets May 1996: Submission to the G-7 Lyon Summit of June 1996* (Basle, Switzerland: BIS)

Bank of International Settlements (2001) *A Survey of Stress Tests and Current Practices at Major Financial Institutions* (Basle, Switzerland: BIS)

Basle on Committee Banking Supervision (1994) *Risk Management Guidelines for Derivatives* (Basle, Switzerland: BCBS)

Beder, T.S. (1995) 'Var: Seductive but dangerous.' *Financial Analysts Journal* pp. 12–23

Bessis, J. (2001) *Risk Management in Banking* (John Wiley & Sons)

Boudoukh, J., Richardson M., and R. Whitelaw (1998) 'The best of both worlds: A hybrid approach to calculating value at risk.' *International Journal of Theoretical and Applied Finance* pp. 64–7

Christoffersen, P., Hahn J., and A. Inoue (2001) 'Testing and comparing value at risk measures.' *Journal of Empirical Finance* 8, 325–42

Daníelsson, Jón (2002) 'The emperor has no clothes: Limits on risk modelling.' *Journal of Banking and Finance* pp. 1273–1296

Dowd, Kevin (1998) *Beyond Value at Risk: The New Science of Risk Management* (John Wiley and Sons Ltd.)

Engle, J., and M. Gizycki (1999) 'Conservatism, accuracy and efficiency: Comparing value at risk models.' *Working Paper: Australian Prudential Regulation Authority*

Engle, Robert, and Simone Manganelli (2001) 'Value at risk models in finance.' *European Central Bank Working Paper No.75*

Goorbergh, R., and Vlaar P. (1999) 'Value at risk analysis of stock returns, historical simulation, varianace techniques or tail index estimation?' *Working Papers Series: Netherlands Central Bank*

Gustafsson, M., and C. Lundberg (2009) 'An empirical evaluation of value at risk.' *Working Paper*

Heffernn, S. (1996) *Modern banking in theory and practice* (John Wiley & Sons)

Hull, John (2006) *Options, Futures and Other Derivatives* (Pearson Education Inc.)

Hull, John, and Alan White (1998a) 'Value at risk when daily changes in market variables are not normally distributed.' *Journal of Derivatives* 5(3), 9–19

Hull, John, and Allan White (1998b) 'Incoprating volatility into the historical simulation method for value at risk.' *Journal of Risk*

Jorion, Philippe (1996) 'Measuring the risk in value at risk.' *Financial Analysts Journal*

—— (2000) 'Risk mangement lessons from long-term capital management.' *European Financial Management* pp. 277–300

—— (2001) *Value at Risk: The New Benchmark for Managing Financial Risk* (New York:McGraw Hill)

—— (2002) 'How informative are value at risk disclosures?' *The Accounting Review*

Kupiec, P. (1995) 'Techniques for verifying the accuracy of risk measurement models.' *Finance and Economics Discussion Series, Federal Reserve Board, Washington DC* pp. 95–124

Longin, Francois (2000) 'From value at risk to stress testing: The extreme value approach.' *Journal of Banking and Finance* pp. 1097–1130

Milne, Frank (2008) 'Anatomy of the credit crisis: The role of faulty risk management systems.' *C.D. Howe Institure Commentary*

Motaman-Samadian, Sima (2005) *Risk Management in Emerging Markets* (Palgrave Macmillan)

Santomero, A.M. (1997) 'Commerical bank risk management: An analysis of the process.' *Journal of Financial Service Research* 12, 83–115

Stahl, G. (1997) 'Three cheers.' *Risk* pp. 67–69