UNIVERSITY OF VAASA

FACULTY OF BUSINESS STUDIES

DEPARTMENT OF ACCOUNTING AND FINANCE

Yanshuang Li

EVALUATION OF VAR CALCULATION METHODS

IN CHINESE STOCK MARKET

VAASA 2008

UNIVERSITY OF VAASA
Faculty of Business Studies
Author:	Yanshuang Li
Topic of the Thesis:	Evaluation of VaR calculation methods in Chinese stock market
Name of the Supervisor:	Professor Timo Rothovius
Degree:	Master of Science in Economics and Business Administration
Department:	Department of Accounting and Finance
Major Subject:	Accounting and Finance
Line:	Finance
Year of Entering the University:	2006
Year of Completing the Thesis:	2008 **Pages:** **70**

ABSTRACT

This paper evaluates different VaR calculation methods in measuring Chinese stock market in terms of the acceptability, variability, accuracy and measurement error of VaR models. Three VaR calculation methods based on 5 different models are evaluated, namely Variance-Covariance methods (VC) based on EARCH model (VCEA), RiskMetrics (VCRM) model, Monte Carlo Simulation (MC) modified with EARCH (MCEA) model and RiskMetrics model (MCRM) and historical simulation (HS).

The main findings of this paper are: First, HS method and VCRM method are unacceptable in calculating VaR in Chinese stock market based on the coverage test suggested by Christoffersen for 125-day evaluation window while only HS is unacceptable for a 50-day evaluation sample. Second, MCEA method has the lowest variability, HS has the highest variability and the variability of MCRM and VCEA are lower than MCEA but higher than VCRM for 125-day and 50-day evaluation windows based on RMSRB. Third, the accuracy of MCEA is the highest among all calculation method used in the paper for 125-day evaluation window while the accuracy of MCEA, MCRM and VCEA is high and similar for 50-day evaluation window. HS and VCRM model have relatively low accuracy for both evaluation windows. Finally, there is measurement error using HS method for 125-day evaluation window based on Hit_t test. It can be conclude that, MC method performs well in calculating VaR Chinese Stock market while HS is an inappropriate method based on the results of four aspects of evaluation test, however performance of each VaR calculation method is affected by the length of evaluation window.

KEYWORDS: VaR, Evaluation, Performance

1. INTRODUCTION

A critical step of financial risk management practice is to construct a proper measure of risk. Both literature researches and application of risk measurement methods are developing gradually over time. There are many techniques to measure financial risk, such as asset liability management (ALM) technique, mean-variance model introduced by Markowitz (1952), CAPM model introduced by William Sharpe and John Lintner's. All of these risk measurement techniques have their own limitations, researchers and risk measurement managers try great effort to improve and create new risk measurement methods. The most under focus and developing risk measure technique is Value At Risk. It is a technique used to estimate the probability of portfolio losses. It is easy understood and widely applied by financial institutions such as banks, security firms and companies that are involved in trading energy and other commodities for quantitative risk management for many types of risks. Moreover, it can calculate the portfolio risk of more than one financial asset. VaR technique is commonly used in the risk control fields. Since its introduction to China, more than 1000 banks, insurance companies, investment funds, and other kinds of non-financial companies use it as a main tool of measuring financial derivative risk. It helps participants to know more exactly how big risk of the transaction they are undertaking.

The key application of VaR is for assessing market risk. However, VaR is not a consistent method for measuring risk, as different VaR models will come up with different VaR results. The great availability of VaR technique has put researchers and risk measurement managers in difficult situation when using VaR since there are no single and standardized criteria to determine which method is the best. Hence evaluation towards performance of VaR methods and selection of appropriate VaR methods become very important. However, research on evaluation of VaR calculation methods is limited even there are endless papers studying about VaR since the day it's introduced. It is important and meaningful to evaluate the forecast ability of different kinds of VaR calculation methods both for literature and for practice. The existing papers on the topic of VaR are mainly about the VaR calculating method itself, or about the building and selecting of models under different methods. There are only few papers studied about the evaluation of performance of different VaR methods, especially for Chinese financial market.

1.1 Research problem

Research problem of this paper is to evaluate performance of different VaR methods in Chinese Stock market, in terms of acceptability, variability, accuracy and measurement error. Performance of five VaR methods will be evaluated, that is Variance-Covariance method based on EARCH model, Variance-Covariance method based on RiskMetrics model, Monte Carlo Simulation based on EARCH model, Monte Carlo Simulation based on RiskMetrics model and Historical Simulation method. For the rest of the paper, these methods are expressed as VCEA, VCRM, MCEA, MCRM and HS respectively.

1.2 Hypothesis

This section presents the hypothesis of this paper, four hypothesizes will be tested regarding the acceptability, variability, accuracy and measurement error of each VaR calculation method.

Hypothesis 1: VCEA, VCRM, MCEA, MCRA and HS are acceptable to calculate VaR in Chinese Stock Market. Although VCEA, VCRM, MCEA, MCRA and HS have their own limitation, all of them haven been applied by different financial institution based on the aim of risk management. Hence it can be hypothesized that VCEA, VCRM, MCEA, MCRA and HS are acceptable to calculate VaR of Chinese Stock Market.

Hypothesis 2: The variability of MCEA is the lowest among all the tested calculation methods and the variability of HS is the highest. Variability is used to measure the bias of VCEA, VCRM, MCEA, MCRA and HS respectively compared with the average. The bias of MCEA is supposed to be lowest due to the advantages of MC method and EARCH model discussed before. As discussed before, HS has more obvious disadvantages then VCEA, VCRM, MCEA, MCRA, hence the bias of HS is considered to be highest.

Hypothesis 3: The accuracy of MCEA is the highest, and the accuracy of HS is lowest in measuring Chinese Stock market risk. Among VCEA, VCRM, MCEA, MCRA and HS, MCEA is usually considered as the most accurate since the forecasted return is simulated based random innovation; it is more closed to real financial market. HS is considered as

least accurate since the forecasted return is simulated based on historical data, it can be very different from reality especially in long-term horizon.

Hypothesis 4: There is no measurement error from VCEA, VCRM, MCEA, MCRA and HS. It is hypothesized that underling models of VCEA, VCRM, MCEA, MCRA and HS are correctly established, then Hit value of each model is uncorrelated with its own lag , with the forecasted VaR and with a constant. There is correct fraction of loss for VCEA, VCRM, MCEA, MCRA and HS.

1.3 Contribution

This paper intends to evaluate performance of different VaR calculation methods in terms of their acceptability, variability, accuracy and measurement error; it contributes to the literature research on the evaluation of performance of VaR in Chinese stock market. As one of the fast developing emerging countries in the word, Chinese Stock market is different from other developed countries in risk measurement field. The rapid growing of financial tools and derivatives and the lacking of mature financial supervising system has enlarged market risk of stock market. Hence VaR of Chinese Stock market is supposed to be larger and more fluctuated. Meanwhile, when using RiskMetrics model to forecast variance of Chinese Stock Market, the decay factor provided by J.P Morgan may not be suitable for Chinese market since it was obtained based on western developed market. Thus it may affect the performance of those VCRM and MCRM, the evaluation may not correctly reflect the real situations.

So far, there is only one research studying the evaluation of performance of VaR in Chinese Stock market by Hua & Wu (2005). The research topic is similar with their study, but different data are used as representation of Chinese Stock Market index. Diversified findings are obtained from empirical results. Further more, this study is different and developed from paper of Hua & Wu (2005) in the following ways:

1. The models in each VaR calculation methods tested in the paper are different and have some modification compared with the ones tested in Hua & Wu's paper. When using Variance-Covariance method, the underlying models are RiskMetrics model and EARCH

model; they are modified by t distribution assumption of stock return instead of normal distribution. A fitting of distribution of stock index is also presented in the paper to discuss the distribution characteristic of stock index return, which is rarely done by other researchers. The regression of parameters of density function is based on maximum likelihood method. For Monte Carlo method, the standard deviation of stock return σ_t in the stock pricing mode is time verifying instead of stable, hence the conditional σ_t is obtained by RiskMetrics and EARCH instead of using stable σ from past history, it is a better reflection of the dynamic volatility of stock market.

2. In the research of Hua & Wu (2005), they used approaches from Hendricks (1995&1997) and Lopez (1999) to test the variability and accuracy of each VaR calculation method. In this paper, beside variability and accuracy test, the acceptability of each VaR method is also tested for Chinese stock market. The acceptability was examined using methods introduced by Kupiec (1995) and Christoffersen (1998). Moreover, measurement error of each method will be tested based on Dynamic Quantile Test introduced by Engle & Manganelli (2001).

1.4 Literature review on the evaluation of VaR

Kupiec (1995) introduced one correct unconditional coverage test to address the "acceptably accurate" of different VaR methods and discussed about the advantages and limitation of such unconditional coverage test. Christoffersen (1998) tried to avoid the limitation of Kupiec's method and developed this correct unconditional coverage test into a correct conditional coverage test. Both methods are widely used nowadays to test the "acceptably accurate" of VaR methods although they still present some limitation due to the limited nature of being hypothesis tests.

Hendricks (1996) examined performance of VaR models by applying value-at-risk models to 1,000 randomly chosen foreign exchange portfolios over the period 1983-94. Nine criteria were introduced and used to evaluate model performance. Results indicate that none of the twelve approaches tested was superior on every count. Moreover, it was found that the choice of confidence level—95 percent or 99 percent—could have a substantial effect on the performance of value-at-risk approaches.

Lopez (1999) discussed the limitation and application of evaluation methods by Kupiec (1995) and Christoffersen (1998) and introduced a loss function method based on three different functions. The empirical results based on simulated exercises showed that the loss function method could distinguish between VaR estimates from the actual and alternative VaR models and all these three functions should be useful in the regulatory evaluation of VaR estimates.

Engle & Gizycki (1999) compared the performance of specific implementations of four VaR model classes based on a range of measures that address the conservatism, accuracy and efficiency of each model. Four classes of VaR models are considered, that is: Variance-covariance models, historical simulation models, Monte Carlo models and extreme-value estimation models. Research portfolio data were from all Australian banks over the past ten years.

Bams, Lehnert and Wilff (2002) investigated the ability of different models to produce useful VaR-estimates for exchange rate positions. The authors divided the exam models into unsophisticated tail model and sophisticated models. And It is found that the uncertainty of VaR estimation is higher for more sophisticated tail-modeling approaches.

Mihailescu (2004) developed a technique for sequential assessment of the appropriateness of the VaR model by drawing on a control chart from statistical process control. The main finding was that an EWMA control chart is the most appropriate instrument for detecting changes in the process of the magnitude of interest in risk management.

Bredin & Hyde (2004) measured and evaluated the performance of a number of VaR methods using a portfolio based on the foreign exchange exposure of Ireland among its key trading partners. Both variability and accuracy of VaR methods were evaluated in this paper, as well as the internal forecast of different VaR models was presented. Results suggest that the EWMA is the more appropriate method.

Kuester, Mittnik and Paolella (2005) compare the out-of-sample performance of existing methods and some new models for predicting Value-at-Risk using more than 30 years of the daily return data on the NASDAQ Composite Index. The assessment of VaR methods is based on approaches introduced by Christoffersen (1998) and Engle and Manganelli

(2002). It was found that most approaches perform inadequately, although several models are acceptable under current regulatory assessment rules for model adequacy. A hybrid method, combining a heavy-tailed GARCH filter with an extreme value theory-based approach, performs best overall.

Liu, Lee and Wu (2005) evaluated empirical performance of various VaR models based on a range of measures that address the conservativeness, accuracy and efficiency. The main methodologies used in this paper were Mean Relative Bias, binary loss function and LR test introduced by Kupiec (1995) and later developed by Christoffersen (1998). The Backtesting results demonstrate that the power exponential distribution can properly capture the fat-tail characteristic of the asset return distributions thus most of the family of EWMA estimators that are based on the power exponential distribution outperform those VaR estimators that are based on the normal distribution, and offer an appropriate coverage of the extreme risk.

Lin, Chien and Hsieh (2005) compared three revised historical simulation methods, namely Richardson and Whitelaw's (1998) hybrid method, Filtered Historical Simulation method proposed by Barone-Adesi, Giannopoulos, and Vosper (1999), and Hull and White's (1998) method, for estimating Value-at-Risk. Using 11 years of 5 daily stock prices and 5 foreign exchange rates, the empirical results show that Hull & White's (1998) method is a substantial improvement for three confidence levels, based on analysis of conservative, accuracy and efficiency.

Angelidis, Benos and Degiannakis (2006) use a two-stage Backtesting procedure to choose one model among the various forecasting methods. The unconditional coverage test is used to examine the statistical "acceptably accurate" of the models in the first stage. In the second stage a loss function is applied to investigate whether the differences between the VaR calculation accuracy are statistically significant. And the results showed that combination of a parametric model with the historical simulation had reliable risk measurement ability.

Paper from Kanwer and Zaidi (2006) evaluated VaR Models in Pakistan using Binary Loss Function and interval forecasts proposed by Christoffersen (1998). And results from tests of the volatility of returns for the Index and Single Stock strongly favor using RiskMetrics with a λ of 0.85.

Kilic (2006) evaluated 13 VaR implementation based on a Turkish Market portfolio that contain foreign currency, stock and bonds. The author extended the methodology provided by Christoffersen and Pelletier (2004) to create duration based analogous of unconditional coverage, conditional coverage and independence tests and found that modified version of Weibull test can also detect coverage.

Pen, Rivera and Mata (2006) provide a discussion about the drawback of Basel Backtesting method and introduced a new statistical approach to assess the quality of risk measures (QCRM). But this paper didn't provide empirical test of any VaR models using Basel Backtesting or QCRM. It is just a method introduction paper.

Lamantia, Ortobelli and Rachev (2006) compared and investigated the forecasting power of different VaR models; how the performance of associated aggregation rules of different VaR models are also discussed. Research was based on several back test techniques on out-of-sample. Results show that stable Paretian models and the Student's t-copula have good future losses predicting ability and some stable parametric models present better performance for smaller percentiles and for large portfolios. The α-stable densities are reliable in the VaR calculation and are characterized by an approximating temporal aggregation rule but when the temporal horizon is too large the time rules cannot be applied.

Rivera, Lee and Yoldas (2007) investigate the implications of different loss functions in estimation and forecasting evaluation within RiskMetrics methodology using U.S. equity, exchange rates, and bond market data. The main finding was that results of estimation and forecasting evaluation could be different under alternative loss functions.

Smith (2007) studied the ability of conditional and unconditional tests to detect miss-specification of Value-at-Risk (VaR) models and develop a new conditional Lagrange Multiplier test based on a Probit model for situation that even when there are no exceptions. Some new conditioning variables to detect exception clustering are also proposed. Empirical results showed that all of the five actual bank VaR models tested are miss-specified and that much of the deficiency is due to their inability to adjust to changes in volatility.

In China, plenty amounts researches have been done on VaR, most of the researches are mainly focus on VaR calculation method itself or focus on building and selecting appropriate models for each method, but there is very few studies exam characteristic or forecast ability of these methods. So far, in China, when considering examine VaR calculation methods, most researches only use Basel criteria or Kupiec (1995) test directly as their last step of studies but not as a topic. There is few papers introduced some evaluation techniques of VaR methods and compared the evaluation ability of such techniques. For example, Li and Guo (2003) discussed about variety of feedback testing approaches and indicated that the mix Kupiec Testing and simplified CD testing can effectively evaluate VaR models.

While considering about researches on evaluation of performance of VaR methods, Chen and Yang (2003) proposed a conditional EVT method combining with APARCH model to estimate conditional quintiles (VaR). The model is compared with other three common VaR calculating methods and unconditional EVT method using Standard Deviation of Capital Employed and evaluation approaches introduced by Christoffersen (1998). Results showed that conditional EVT method yields statistically valid VaR measures and gives better one-day estimates than methods that ignore the fat tails of the innovations or the stochastic nature of the volatility so it is a robust tool for estimating risk of financial portfolios.

Liu & Zheng (2007) tested the forecast ability of VaR models and empirical results showed that current back-test tools including Basle test, Kupiec test and Christoffersen test used in the business banks' model risk management can be somehow misleading.

 Zhang & Zheng (2007) discussed above indices portfolio VaR models and used dynamic quintile test and failing rate method to compare accuracy of different models and find out ADCC model is better than RiskMetrics for portfolio and risk management with different portfolio weights.

Another paper that showed empirical evidences and conclusion about the accuracy and variability of VaR methods was from Hua & Wu (2005). MRB and RMSRB approaches introduced by Hendricks (1995 and 1997) were used as measurement of variability, while two loss functions started by Lopez (1999) were used to test the accuracy of different VaR

methods .The three main finding from Hua & Wu's paper are: First, parameter methods are most compatible to the movement of returns. Second the parameter model has least variability and no parameter method (Historical Simulation) has the highest variability. Third, the estimation accuracy by half- parameter methods (Monte Carlo simulation) and non- parameter methods is higher than by parameter methods.

1.5 Structure of the paper

The paper is set up as follows. Section 2 will present an introduction of theory and calculation methods of VaR. Methodologies used to assess performance of VaR models will be introduced in section 3. Section 4 presents the calculation and evaluation of different VaR methods. Empirical results and findings of assessment will be presented in section 5. Conclusions and propose for future research are offered in Section 6.

2. THEORETICAL BACKGROUND

The main purpose of this chapter is to present essential theoretical background of Value-at-risk and the most commonly used calculation methods. The first part of this chapter briefly introduces the definition of VaR for a single financial asset, while the second part provides some introduction of three most commonly VaR calculation methods

2.1 Definition of VaR

The definition of VAR provided by Philippe Jorion is that Value at Risk (VaR) is the maximum loss not exceeded with a given probability defined as the confidence level, over a given period of time. The mathematic expression can be:

(1) $\mathrm{Pr}ob(\Delta P \leq -VaR_t) = a$

Where: ΔP is the change of asset price,

   t is the asset holding period

And   a is the given probability.

Based on Jorion (1996), the VaR of a single asset within a one-day of holding period at time period t can be calculated as the difference between the expected value (mean) of financial asset with the minimum close price under given confidence level α, which is:

(2) $VaR_t = E(P_t) - P_t^* = P_{t-1}(1+\mu) - P_{t-1}(1+r^*) = P_{t-1}(\mu - r^*)$

Where: μ is the expected value of financial asset,

   r* is the minimum close price under given confidence level a

And P_{t-1} is the asset price at time period t-1.

Suppose return of financial asset follows one type of particular distribution, the critical value of such return distribution under given confidence level is Z_a, σ is the standard deviation of return, then the minimum rerun will be r*=μ- Z_aσ, thus:

(3) $VaR_t = P_{t-1}(\mu - r^*) = P_{t-1}(\mu - (\mu - Z_a\sigma)) = -Z_a\sigma_t P_{t-1}$

To calculate the VaR of asset return instead, it can be supposed that $P_{t-1}=1$, formula (2) will become:

(4) $VaR_t=\mu- r^*$

And formula (3) will become:

(5) $VaR_t= -Z_a\sigma_t$

2.2 Calculation methods

There are three common VaR calculation methods based on above calculation formulas, namely variance-covariance method (VC), Historical Simulation method (HS) and Monte Carlo Simulation method (MC).

2.2.1 Variance-covariance method (VC):

The variance-covariance, or delta-normal model was popularized by J.P Morgan (now J.P. Morgan Chase) in the early 1990s when they published the RiskMetrics Technical Document. It is a parametric, analytic technique where the distributional assumption made is that the daily geometric returns of the market variables are multivariate normally distributed with zero mean return. Historical data is used to measure the major parameters: means, standard deviations, correlations. When the market value of the portfolio is a linear function of the underlying parameters, the distribution of the profits is normal as well.

From the formula (5) $VaR_t=-Z_a\sigma_t$, it is noticeable that the determinants of VaR using VC method are the value of Z_a and σ_t, thus the calculation processes of a VC method can be divided as following:

1) The value of Z_a is determined by the asset return distribution assumption. When using standard VC method, it is usually assumed to be normal distribution. However, The distribution of daily returns of any risk factor in reality would typically show significant amount of positive kurtosis (See for example, Fama (1965). This leads to fatter tails and extreme outcomes occurring much more frequently than would be predicted by the normal

distribution assumption, which would lead to an underestimation of VaR (since VaR is concerned with the tails of the distribution). Hence Z_a based on normal distribution will not well reflect the real situation. Discussion of return distribution situation is necessary, some other popular distribution assumptions such as student t distribution or GED can be considered to fit the real distribution situation of financial asset. In this paper, student t distribution will be discussed and applied to have a fitting using maximum likelihood method towards Chinese stock market index.

2) Forecast σ_t. Volatility of financial market is found to be time verifying and conditional (Engle (1982)). ARCH family models can be used to forecast the volatility of stock market index using historical data. An autoregressive conditional heteroskedasticity (ARCH, Engle (1982)) model considers the variance of the current error term to be a function of the variances of the previous time period's error terms. It relates the error variance to the square of a previous period's error. It is employed commonly in modeling financial time series that exhibit time-varying volatility clustering, i.e. periods of swings followed by periods of relative calm. There are many forms of ARCH family models. From basic ARCH to GARCH, GARCH-M, EARCH, TARCH and many other developed forms. Consider the significance of regressed parameters and the minimum AIC criteria, EARCH model will be chosen as the forecast model of σ_t. The form of EARCH (1,1) is expressed as formula (6)

$$(6) \qquad \ln(\sigma_t^2) = \alpha_0 + \theta_1 \ln(\sigma_{t-1}^2) + a_1 \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

Another model used to forecast σ_t is RiskMetrics developed by J.P Mogen (1996). The form of RiskMetrics model is:

$$(7) \qquad \sigma_{t+1}^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) r_t^2$$

Where: λ is the decay factor that used to smiplify the set of weight factors.
The value of λ is between 0 and 1. For calculating daily VAR, λ is set to be 0.94 or 0.97 for VAR in RiskMetrics method. From the model it is easily noticed that the weighting for each older data point decreases exponentially, giving much more importance to recent observations while still not discarding older observations entirely.

The advantages of VC method include its speed and simplicity, and the fact that distribution of returns need not be assumed to be stationary through time, since volatility updating is incorporated into the parameter estimation. While the disadvantages lie in its distribution assumption and the fact that it inadequately measures the risk of nonlinear instruments, such as options or mortgages.

2.2.2 Historical Simulation (HS)

The key assumption in historical simulation (HS) is that the set of possible future scenarios is fully represented by what happened over a specific historical window. HS involves collecting the set of risk factor changes over a historical window: for example, daily changes over the last two years. The set of scenarios obtained is assumed to be a good representation of all possibilities that could happen between today and tomorrow. The instruments in the portfolio are then repeatedly re-valued against each of the scenarios. This produces a distribution of portfolio values, or equivalently, a distribution of changes in portfolio value from today's value. Usually, some of these changes will involve profits and some will involve losses. Ordering the changes in portfolio value from worst to best, the 95% VaR, for example, is computed as the loss such that 5% of the profits or losses are below it, and 95% are above it.

For a single asset, the calculation process for VaR of asset return using HS method is relatively simple. Based on the formula (4) $VaR_t = \mu - r^*$, the crucial steps in HS method is to calculated the expected return and the minimum return at given confidence level. To calculate VaR at time period t, returns data from estimation window T period ahead will be used as a representation of the possible returns for period t. Then expected return at period t μ and minimum return under given confidence level r^* can be obtained using those historical data.

The main advantage of historical simulation is that it makes no assumptions about risk factor changes being from a particular distribution. Therefore, this methodology is consistent with the risk factor changes being from any distribution. Another important advantage is that HS does not involve estimation of any statistical parameters, such as variances or co-variances, and is consequently exempt from inevitable estimation errors. It

is also a methodology that is easy to explain and defend to a non-technical and important audience, such as a corporate board of directors.

However, HS also has some disadvantages. The most obvious disadvantage is that historical simulation, in its purest form, can be difficult to accomplish because it requires data on all risk factors to be available over a reasonably long historical period in order to give a good representation of what might happen in the future. Another disadvantage is that historical simulation does not involve any distributional assumptions; the scenarios that are used in computing VaR are limited to those that have happened in the historical sample.

2.2.3 Monte Carlo Simulation (MC)

The calculation steps of MC is similar with HS method, the key difference between HS and MC is that the HS carries out the simulation using the real observed changes in the market place over the last T periods to generate hypothetical portfolio profits or losses, whereas in MC simulation a random number generator is used to produce tens of thousands of hypothetical changes in the market. These are then used to construct thousands of hypothetical profits and losses on the current portfolio, and the subsequent distribution of possible portfolio profit or loss. Finally, the VaR is determined from this distribution according to the parameters set (e.g. 95 % confidence level) using the formula μ- r*. To simulate stock price movement, Geometric Brownian Motion is generally used to describe the movement of stock price in short horizon, then form of Geometric Brownian Motion is:

(8) $dS_t = \mu_t S_t dt + \sigma_t S_t dw_t$

Where: dS_t is the changing amount of financial asset,

μ_t is asset return,

σ_t is standard deviation of return

And $dw_t \sim N(0, dt)$ is Brownian motion.

The changing process of asset price in particular period (0,T) can be simplified as:

$$(9) \qquad \frac{\Delta S_t}{S_t} = \mu \Delta t + \sigma \varepsilon_t \sqrt{\Delta t} \quad (t = 1, 2, \cdots, N; \; N\Delta t = T).$$

It is noticeable from the asset pricing model that the keys of an asset simulation process is the stochastic event ε_t, usually it is assumed to be a normal distribution process with zero mean and a standard deviation of 1 $\varepsilon_t \sim (0,1)$. This simulation of stochastic events will be modified based on the results of distribution fitting to get a better and more accurate simulation of stochastic process. Other key factors of the asset simulation are the parameters in the pricing model, namely μ and σ. As discussed before, the volatility of stock price is dynamic and time verifying and conditional, hence the stable parameters σ has to be modified as dynamic σ_t, σ_t will be forecasted using both EARCH (1,1) model and RiskMetrics models as discussed in VC method. Value of μ will also be obtained from historical data. After simulation of asset movement, the calculation of VaR then proceeds as for the historical simulation method.

The advantages of MC simulation are obvious; it is by far the most flexible and powerful method, since MC method is able to take into account all non-linearity of the portfolio value with respect to its underlying risk factor, and to incorporate all desirable distributional properties, such as fat tails and time varying volatilities. Also, MC simulations can be extended to apply over longer holding periods, making it possible to use these techniques for measuring credit risk. However, these techniques are also by far the most expensive computationally

3. METHODOLOGY

The evaluation of VaR forecasts is not straightforward. A direct comparison between the forecast VaR and the realized VaR cannot be made since the latter is unobservable. Plenty amount of methods have been proposed (see, for instance, Kupiec (1995); Christofferson (1998); Lopez (1998)) to evaluate performance of VaR. Up to now, there is no single definition of VaR model performance that has been developed. To evaluate the performance of this family of models, a range of statistics that address four aspects of the usefulness of VaR models to risk managers and the supervisory authorities are proposed in this paper.

Firstly, interval forecasts proposed by Kupiec (1995) and Christoffersen (1998) are adopted to test the acceptability of those VaR calculation methods. The evaluation frameworks introduced by Kupiec (1995) and Christoffersen (1998) were generally used by financial regulators to evaluate and determine whether the underlying VaR methods were "acceptably accurate" (Lopez (1999)). In this paper, they will be applied ahead of accuracy test (Lopez (1999)) to exam whether each VaR method is "acceptably accurate"; this aspect of test is defined as "acceptability test" of VaR methods in this paper. These two evaluations are independent of VaR calculating process and they can capture whether a particular model exhibits correct coverage (both unconditional and conditional). If the VaR calculated using particular method exhibits correct coverage, then it is an acceptable method in measuring Chinese Stock market risk.

Secondly, two measures of relative size and variability developed by Hendricks (1996) will be applied to test the variability of each calculating methods. The variability of each method is the volatility of VaR calculated using such method compared with the mean of VaR obtained from all of the calculated methods. The variability test of VaR enables us to assess whether a particular model produces higher risk estimates relative to the other models.

Thirdly loss function approach of Lopez (1999) will be used to test the accuracy of each method. In this study, accuracy of VaR model is defined as the rate of failure (or exception) associated with how close each specific model came to the preset level of significance. The functions are defined to produce higher values when exceptions occur. In this paper we

adopt two functions, a basic binary loss function which in a sense equivalent to the Christoffersen test of correct conditional coverage, and a quadratic loss function which takes into account the magnitude of the exception. Compared the loss function approaches by Lopez (1999) with the correct coverage approaches by Kupiec (1995) and Christoffersen (1998), the latter can test whether a VaR method is acceptable (Acceptably accurate) while the former can't, Lopez's loss function approaches can be used to provide relative comparisons of model accuracy over different time periods and in relation to other VaR models.

Finally, A Dynamic Quantile Test introduced by Engle & Manganelli (2001) will be implemented to test whether there is measurement error from each VaR calculation method. This test is applied by testing whether there is correlation between the Hit value and its lag and current VaR. It there is autocorrelation in the hits, the fraction of loss occurs in each VaR calculation method will not be correct and there will be some measurement error.

3.1 Acceptability test

3.1.1. Kupiec (1995)

Kupiec (1995) proposed a likelihood ratio test based on the binomial process that can be applied to determine weather the rate of failure is statistically compatible with the expected level of confidence. Given the sample size T and the frequency of failure N governed by a binomial probability. Ideally, the failure rate N/T, should be equal to the left tail probability p. Thus, the relevant null and alternative hypotheses are:

H_0: N/T = p
H_1: N/T ≠ p

And the appropriate likelihood ratio statistic is

$$(10) \qquad LR_{uc} = 2[\log((\frac{N}{T})^N (1-\frac{N}{T})^{T-N}) - \log(p^N (1-p)^{T-N})] \sim \chi^2_{1,a}$$

Under the null correct hypothesis of correct unconditional coverage, the LR_{uc} has a chi-squared distribution with one degree of freedom.

The problem of regarding to the finite sample and evaluation power of this unconditional coverage test has been discussed. For example, Lopez (1999) pointed out that the finite sample distribution of LR_{uc} for the specified parameters may be sufficiently different from a $\chi^2_{(1)}$ distribution that the asymptotic critical values may be inappropriate. Kupiec (1995) describes how this test generally has a limited ability to distinguish among alternative hypotheses and thus has low power, even in moderately large samples. Despite the natural appeal and simplicity of this unconditional coverage test it lacks power to detect violations (see, e.g., Jorion (2006)). For example, at a 95% confidence level, the expected number of failure for a 125 days sample is 125*(1-95%)=6. If the actual failure happed is 7, less than 5% of significant level, the LR_{uc} is less then 3.84 then it can be said that it has correct unconditional coverage, so the model cannot be rejected. However in this case if more than 5 of the failure among these 7 happens within the nearest two weeks (the failure is violate), then the underlying model can not be considered as valid since it doesn't have correct conditional coverage. Due to such weakness of LR_{uc} test much effort has been devoted to develop conditional tests with better power, the correct conditional coverage test introduced by Christoffersen (1998) is one of good examples.

3.1.2. Christoffersen

Generally, the VaR forecasts should be small in periods of low volatility and larger in more volatile periods. The failures should therefore be spread across the sample and should not appear in clusters. As discussed by Christoffersen (1998), The LR_{uc} test is an unconditional test since it simply counts exceptions over the entire period. A VaR model that inadequately captures volatility clustering will tend to have too many exceptions during periods of market turbulence. Christoffersen (1998) shows that such inadequate volatility modeling will result in serial correlation in exceptions; interval forecasts that ignore such variance dynamics may have correct unconditional coverage but, at any given time, will have incorrect conditional coverage. Hence he suggested a conditional coverage test which tests for independence in the exceptions. The interval forecast proposed by Christoffersen (1998) is a framework that is independent of the process of generating the VaR forecasts and captures whether a particular model exhibits correct conditional coverage.

Christoffersen (1998) approach includes a three-step procedure for the evaluation of interval forecasts, which is: A test for "Correct Unconditional Coverage", a test for "Independence" and a test for "Correct Conditional Coverage". Interval forecasts can be evaluated conditionally or unconditionally, that is, with or without reference to the information available at each point in time.

1) A test for "Correct Unconditional Coverage"

It is the same as the test for "Correct Unconditional Coverage" introduced by Kupiec (1995), which is:

$$(11) \qquad LR_{uc} = 2[\log((\frac{N}{T})^N (1-\frac{N}{T})^{T-N}) - \log(p^N (1-p)^{T-N})]$$

Though a poor interval forecast may still produce correct unconditional coverage it fails to capture the higher order dynamics of the series. The test, however, for correct unconditional coverage can be utilized to penalize firms it does not capture asymmetries or leverage effects which will affect the accuracy and efficiency of any forecasts. The test for independence tests the hypothesis that the failure process is independently distributed against an alternative that the process follows a first order Markov process.

2) A test for "Independence"

If a VaR model accurately captures the conditional distribution of returns, as well as its dynamic properties such as time-varying volatility, then exceptions should be unpredictable, and hence independently distributed over time. To test the independence of the exceptions of a VaR model, Christoffersen (1998) has derived an LR statistic, which is the likelihood ratio statistic for the null hypothesis of serial independence against the alternative of first-order Markov dependence, the null hypothesis is:

H_0: $\pi_{01} = \pi_{11} = \pi$

While the likelihood function under this alternative hypothesis is:

(12) $L_u = (1-\pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1-\pi_{11})^{T_{10}} \pi_{11}^{T_{11}}$

Where the T_{ij} notation denotes the number of observations in state j after having been in state i the period before, $\pi_{01} = \dfrac{T_{01}}{T_{00}+T_{01}}$, $\pi_{11} = \dfrac{T_{11}}{T_{10}+T_{11}}$

Under the null hypothesis of independence, the relevant likelihood function is

(13) $L_R = (1-\pi)^{T_{00}+T_{10}} \pi^{T_{01}+T_{11}}$

Where $\pi = \dfrac{T_{01}+T_{11}}{T}$, T is the total number of observing sample.

The test statistic for "independence" is:

(14) $LR_{ind}=2(\ln L_u - \ln L_R)$

Which has an asymptotic $\chi^2_{1,a}$ distribution.

3) A test for "Correct Conditional Coverage"

To provide Correct Conditional Coverage is an important requirement of a VaR model. If a VaR model has the ability to capture the conditional distribution of returns and its dynamic properties such as time varying volatility accurately, then exceptions should be unpredictable. The importance of testing this aspect stems from the financial time series characteristic of volatility clustering. The LR_{cc} test is a joint test of these two properties, the relevant test statistic is

(15) $LR_{cc}=LR_{uc}+LR_{ind} \sim \chi^2_{2,a}$

Which is asymptotically distributed $\chi^2_{2,a}$

3.2 Variability test

3.2.1 MRB

To assess the relative size of the VaR estimates produced by the various models, mean relative bias statistic developed by Hendricks (1996) will be applied. This statistic captures the extent to which different models produce estimates of similar average size. Given T time periods, and N VaR models, the mean relative bias of any model i is calculated as:

$$(16) \qquad MRB_i = \frac{1}{T}\sum_{t=1}^{T}\frac{VaR_{it} - \overline{VaR_t}}{\overline{VaR_t}}$$

Where: $\quad \overline{VaR_t} = \frac{1}{N}\sum_{i=1}^{N}VaR_{it}$

In the study of Engel & Gizycki (1999), the MRB method was applied to measure the conservativeness of VaR models; the conservativeness is measured in terms of the relative size of the VaR in relation to the risk assessment. The larger the VaR value was, the more conservative the model became. Those models that systematically produce higher estimates of risk are consider as conservative models relative to other others. The mean relative bias statistic captures the degree of the average bias of the VaR of the specific model from the all-model average.

However, the MRB measure is in terms of the relative but not absolute concept. If the evaluated models included are different, then we might obtain different results regarding the relative conservativeness of the models.

3.2.2 RMSRB

To better reflect the variability of different VaR estimation methods, Hendricks (1997) introduced simplified average relative model, this model is a better reflection of the bias of means of estimation towards the means of all estimation methods. The form of model is as following:

$$(17) \qquad RMSRB_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\frac{VaR_{it} - \overline{VaR_t}}{\overline{VaR_t}})^2} \;, \text{ Where } \overline{VaR_t} = \frac{1}{N}\sum_{i=1}^{N}VaR_{it}$$

3.3 Accuracy test

Different users of the VaR model will focus on different types of model inaccuracies. Supervisors may be expected to pay more attention to the underestimation of losses while financial institutions will be more concerned about the over-prediction of losses due to capital adequacy requirements. Lopez (1999) proposes a regulatory loss function in order to assess the accuracy of the VaR estimates. The general loss function of financial institution i an time t is:

$$(18) \qquad L_{i,t+1} = \{ \begin{array}{ll} f(\Delta P_{i,t+1}, VaR_{i,t}); & \Delta P_{i,t+1} < VaR_{i,t} \\ g(\Delta P_{i,t+1}, VaR_{i,t}); & \Delta P_{i,t+1} \geq VaR_{i,t} \end{array}$$

Where f () and g () are functions that satisfy f () $\geq$ g () and ΔP represents the realized profit or loss. In this paper two specific loss functions are considered-a binary loss function which takes account of whether any given days loss is greater or smaller than the VaR estimate and a quadratic loss function which also takes account of the magnitude of the losses that exceed the VaR estimate.

3.3.1 Binary loss function (BLF)

The binary loss function is based on whether the actual loss is larger or smaller than the VaR estimate. Here we are simply concerned with the number of failures rather than the magnitude of the exception. If the actual loss $\Delta P_{i, t+1}$ is larger than the VaR then, it is termed an "exception"(or failure) and has a value equal to 1, with all others having a value of 0. That is

$$(19) \qquad L_{i,t+1} = \{ \begin{array}{ll} 1 & ; \Delta P_{i,t+1} < VaR_{i,t} \\ 0 & ; \Delta P_{i,t+1} \geq VaR_{i,t} \end{array}$$

The aggregate number of failures across all dates is divided by the sample size. The BLF is obtained as the rate of failure. The BLF provides a point estimate of the probability of failure. In other words, the accuracy of the VaR model requires that the BLF, on average, is equal to one minus the prescribed confidence level of the VaR model. The closer the BLF

value is to the confidence level of the model, the more accurate the model is. If the VaR model is truly providing the level of coverage defined by its confidence level, then the average binary loss function over the full sample will be equal to 0.05 for the 95% VaR estimate. An important feature of the failure distribution is that failures should be independently distributed.

3.3.2 Quadratic loss function

The binary loss function has considered only the number of exceptions; no additional information beyond that is contained in the binomial method such as the magnitude of the exception happened. As noted by the Basle Committee on Banking Supervision (1996), the magnitude is also a matter of concern to regulators. As discussed by Hendricks (1996), the magnitude of the observed exceptions can be quite large. Thus Lopez introduced a loss function by incorporating a magnitude term into the binomial loss function. The magnitude is measured using a quadratic term. Lopez (1999) pointed out that a quadratic loss function provides more information than Binary loss function about the measurement accuracy of the VaR estimation methods. For Binary loss function a score of 1 is imposed when an exception occurs, but for a Quadratic loss function an additional term based on the magnitude of the exception is included. The numerical score increases with the magnitude of the exception and can provide additional information on how the underlying VaR model forecasts the lower tail of the f () distribution. The form of a quadratic loss function is:

$$(20) \qquad L_{i,t+1} = \{ \begin{array}{ll} 1 + \ (\Delta P_{i,t+1} - VaR_{i,t})^2 & ; \Delta P_{i,t+1} < VaR_{i,t} \\ 0 \ ; & \Delta P_{i,t+1} \geq VaR_{i,t} \end{array}$$

Sarma et al (2000) suggest that a loss function of the form in formula (20) captures the goals of financial regulators, referring to it as a regulatory loss function.

3.4 Measurement error test

The Dynamic Quantile Test introduced by Engle & Manganelli (2002) is an F test of the hypothesis that all coefficients as well as the intercept are zero in a regression of this variable on its past, on current VaR, and any other variables.

$$(21) \quad Hit_t = I - a(\Delta P_{i,t+1} < VaR_{i,t}) = \{ \begin{matrix} 1-a; \Delta P_{i,t+1} < VaR_{i,t} \\ -a; \Delta P_{i,t+1} \geq VaR_{i,t} \end{matrix}$$

Where α is the confidential level, the function $Hit_t = I - a \ t$ is assumed to take a value 1-α every time when $\Delta P_{i,t+1} < VaR_{i,t}$ and it takes the value -α in all other cases. The equation (19) implies that the expectation of Hit_t is zero. Furthermore, from the definition of the Quantile function, the conditional expectation of *Hit* given any information known at *t-1* must also be zero. The Dynamic Quantile (DQ) test is as following:

$$(22) \quad Hit_t = \delta_0 + \sum_{i=1}^{r} \delta_i Hit_{t-i} + \delta_{r+1} VaR_t$$

The DQ test is computed using the regression of the variable Hit_t on its past, on current VaR, and any other variables. In particular, Hit_t must be uncorrelated with any lagged Hit_{t-k}, with the forecasted VaR_t and with a constant. If Hit_t satisfies these conditions, then it is sure that there will be no autocorrelation in the hits, there will be no measurement error as in (22), and there will be the correct fraction of loss.

4. EMPIRICAL PART

In this section, various VaR calculation methods will be applied to the Chinese stock market index HS300. Throughout the analysis, a holding period of one day will be used. A 5% of the left tail probability level will be considered. The various VaR models will be estimated using the data proceeding the last 500 days of the sample and will be evaluated in the 125 days and 50 days of evaluation sample by means of four test approaches.

4.1 Market risk situation in Chinese Stock market

As one of the fast developing emerging financial markets in the world, Chinese Stock market is undergoing great development in both underling assets and derivatives. In China nowadays, there are about 1300 listed companies in stock spot market, the total market value reaches about 5000 billions RMB and it occupies about 30% of total GDP. Compared this ratio\ with developed international financial markets when they introduced stock index future, which was 44% for USA in 1982, 21% for Germany in 1990 and 29% for South Korea in 1996, the degree and scale of Chinese stock spot market are enough for practicing of stock index future. Both superiority institution and participants of Chinese financial market try great effort for the carrying out of the stock index future. From 25th September 2007 the simulation transaction system of HS 300 index future was stated in China Financial Futures Exchange. Aim of this simulation system is to test and improve of mechanism and technique of stock index futures; these are all related to the success of future running of the product. After the list of HS 300 index futures, more derivatives will be created based on index such as stock index futures and options, meanwhile, the success of transaction of HS 300 index futures will be the basic of developing of other kinds of derivatives based on interest, foreign exchange rate and so forth.

Stock index futures market is the product of innovation of financial and is the important form of creativity of financial transaction tools of futures market. It is also a financial risk control technique towards the uncertainty of stock spot market. The creating of stock index futures will play an important role in the development of Chinese financial market. It does not only provide more investment tools in Chinese financial market but also help to develop and large institutional investors. Stock index futures also increase the efficiency

and liquidity of stock market, as well as reduce the system risk through hedging transaction and protect benefit of investors. However, like other kind of financial derivatives, stock index futures also has the characteristic of high leverage, sensitive to price change and completive of transaction rule. Compared to stock spot market, the risk of stock index futures market is much higher and advanced risk measurement and control techniques need to be created. For stock index futures market, VaR is a commonly used risk measurement technique; also it is one method of calculating the margin level of futures m in reality. Hence, research on VaR in Chinese stock index futures market has both literature and practical meaning, testing and selecting appropriate model to calculate the VaR of HS 300 index can help to exam the market risk of index futures, as well as providing important tools to calculate and set the margin level of futures contract later.

As discussed above, Chinese stock market nowadays is gradually becoming one of volatile financial markets, both market participants and market regulators need models for measuring, managing and containing risks. Market participants need risk management models to manage the risks involved in their open positions. Market regulators on the other hand must ensure the financial integrity of the stock exchanges and the clearinghouses by appropriate margining and risk containment systems. However, there is no single optimal tool used to measure market risk, thus it is important for both market participants and regulators to understand the strengths and weaknesses of different risk measurement approaches. VaR is one of the most useful techniques that being discussed by both researcher and stock market experts nowadays in risk measurement filed since its creation.   It is a popular and simple method to compute finance risk because it takes the loss of investors as the risk.

4.2 Data description

HS 300 index was officially released on 8th April 2005; it is a componential index constructed by 300 large-scare A stocks with high degree of liquidity selected from both Shanghai Exchange and Shenzhen Exchange, 179 from Shanghai Exchange and 121 from Shenzhen Exchange. The basic period of the index is 31st Dec 2004.The sample of the index covers about 70% of total market value while 60% of liquidation value of Shanghai and Shenzhen stock exchange and it is a good representation of market. It is the first jointly public index that reflects the trend of the whole A stock market. Introduce of HS 300 index

enriches the existing market index system and increases one indicator of market trend. It helps investors to analyze the running of financial market as a whole, as well as provides fundamental condition for the innovation and development of derivative of index investment product. Due to its high market coverage rate and identification, it is the most suitable one for developing into stock index future in Chinese stock market. The HS300 is an equity basket consisting of a 300 Chinese listed stocks with high liquidity and good performance in different weights. Throughout the analysis, it will be used as a representative stock. A time series of 625 daily data running from 11/05/2005 to 02/11/2007 will be analyzed. During that time span of about 2.5 years, the index rose from 1003.45 to 5472.93, about 18% a year. The return rate of the index is calculate using natural log difference of the price using formula $R_t = lyI_t - lyI_{t-1}$

Some descriptive statistics of the data are shown in Table 1

Table 1. Descriptive Statistics.

Mean	Median	Maximum	Minimum
0.002681	0.003034	0.078627	-0.096952
Std.Dev.	Skewness	Kurtosis	Jarque-Bera
0.017643	-0.742895	6.767033	427.0341

Kurtosis 6.767033, the kurtosis exceeds 3, the distribution of HS300 is peaked (leptokurtic) relative to the normal. Skewness is-0.742895, a negative Skewness implies that the distribution has a long left tail. A Jarque-Bera value of 427.0341 also leads to the rejection of the null hypothesis of a normal distribution. The descriptive statistics show that the distribution of HS 300 index return does not fulfill the null hypothesis of normal distribution. It has the obvious characteristics of sharp peak and flat tail of financial data, normal distribution cannot describe the characteristic of flat tail financial data and thus the accuracy of models base on normal distribution assumption will be relatively low.

Figure 1 also indicates that the distribution of the series is not normal. The QQ-plot does not lie on a straight line; the distributions of the return series differ along some dimension. So it can be concluded that there is flat tail and sharp peak exist among the return series of HS 300 index.
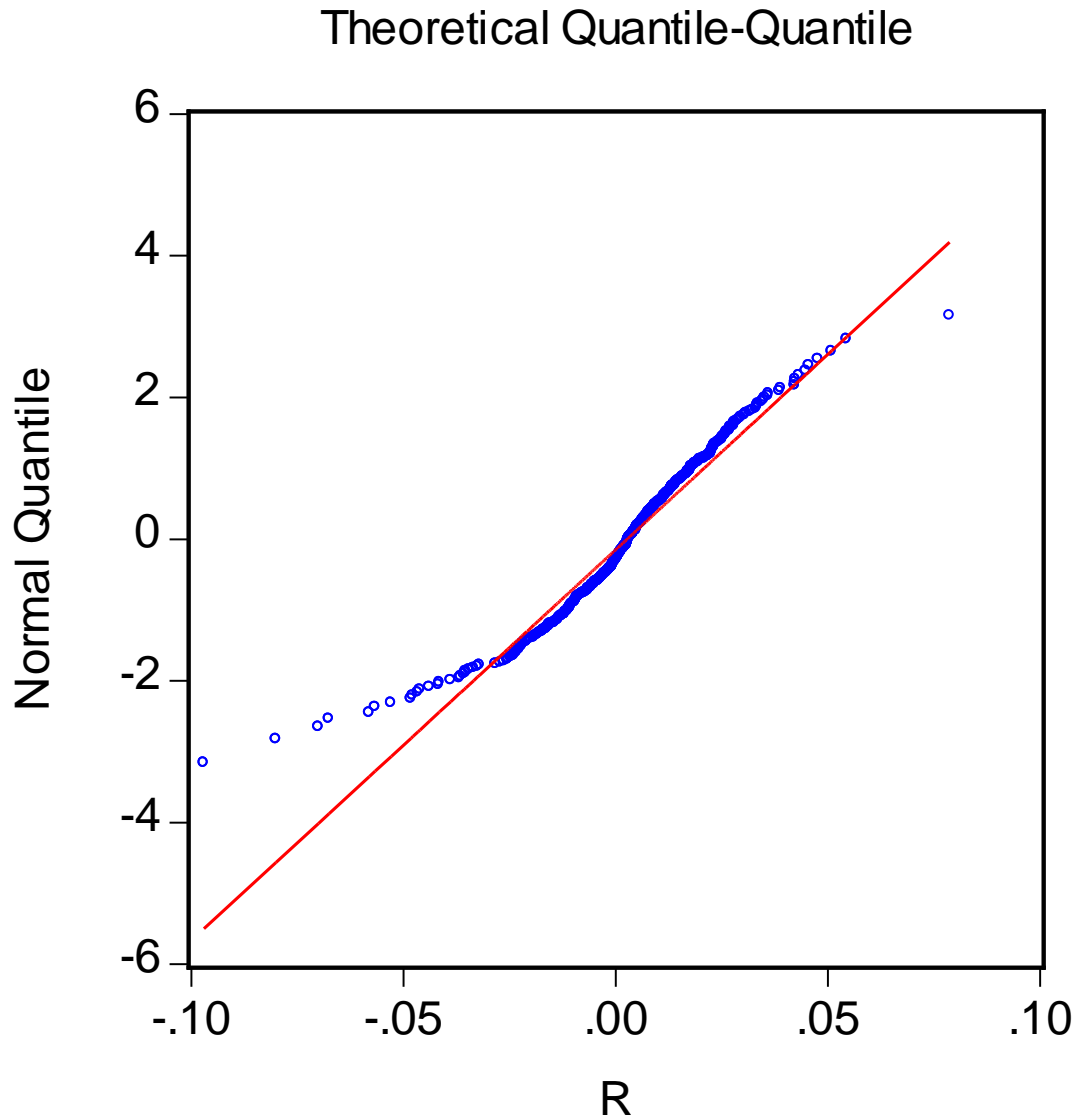
Figure 1. QQ-plot

4.3 Statistical tests

4.3.1 Distribution fitting

As discussed before, normal distribution assumption cannot describe the fat tail and sharp peak phenomenon of stock return of Chinese market. t distribution is another popular distribution assumption that used to describe distribution of financial asset. It is proved by

many researchers that t distribution is a better distribution in describing the fat tail and sharp pear characteristic of financial return. To have a comparison with normal distribution, density function of both assumptions will be regressed using maximum likelihood; a distribution fitting towards return of Chinese stock index is based on normal distribution and student t distribution will be realized. The normal density function is expressed by formula (23):

(23) $f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\dfrac{1}{2\sigma^2}(x-\mu)^2 \right]$

The form of density function of t distribution is expressed by formula (24):

(24) $f(x) = \dfrac{\Gamma[(v+1)/2]}{\Gamma(v/2)} \cdot \dfrac{1}{\sqrt{v\pi}} \cdot \dfrac{1}{(1+x^2/v)^{(v+1)/2}}$

Where Γ is the gamma function. As v increases, this function converges to the normal distribution

For both distribution density, the mean is E [X]=μ and variance V [X]=σ^2

The regression results of density function of t distribution are: μ=0.00281162, σ=0.0112316 and v=3.85319, from the figure 2, it is obvious that student t distribution is a better assumption than normal distribution. Both figure and regression results are realized using Matlab 7.0
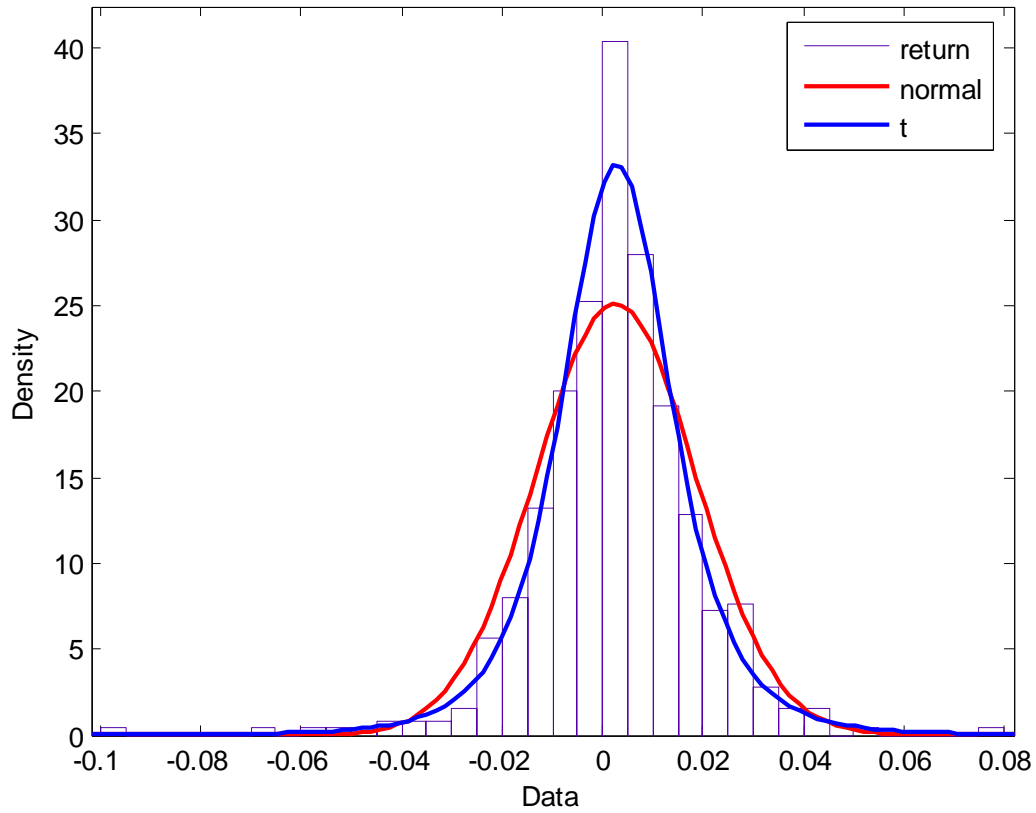
Figure 2. Distributions fitting of return series.

4.3.2 Stationary test (ADF)

Table 2. Stationary test.

Null Hypothesis: R has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic based on SIC, MAXLAG=18)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-24.53799	0.0000
Test critical values:	1% level	-3.440600	
	5% level	-2.865954	
	10% level	-2.569179	

The ADF statistic value is –24.53799 and the associated one-sided p-value (for a test with 623 observations) is 0.0000. Notice here that the statistic value is smaller than the critical values at the 1%, 5% and 10% levels, so that we reject the null at conventional test sizes. The series does not have a unit root; it can be considered as stationary.

4.3.3 Autocorrelation test

See Correlograms Q-statistics test, the autocorrelations and partial autocorrelations are nearly zero up to lag 32, and all Q-statistics are insignificant with large p-values, which indicates that there is no serial correlation in the residuals.

Included observations: 625

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.015	0.015	0.1428	0.706
		2	-0.020	-0.021	0.4028	0.818
		3	0.050	0.051	1.9875	0.575
		4	0.035	0.033	2.7511	0.600
		5	0.033	0.034	3.4470	0.631
		6	-0.053	-0.056	5.2388	0.514
		7	-0.023	-0.024	5.5882	0.589
		8	-0.046	-0.053	6.9520	0.542
		9	0.031	0.035	7.5597	0.579
		10	0.050	0.053	9.1729	0.516
		11	0.072	0.085	12.521	0.326
		12	0.026	0.026	12.959	0.372
		13	0.049	0.046	14.519	0.338
		14	0.044	0.023	15.746	0.329
		15	0.084	0.077	20.298	0.161
		16	0.035	0.029	21.102	0.175
		17	-0.080	-0.072	25.179	0.091
		18	0.019	0.018	25.414	0.114
		19	-0.005	-0.010	25.432	0.147
		20	0.007	0.010	25.459	0.184
		21	-0.040	-0.035	26.517	0.187
		22	0.017	0.024	26.714	0.222
		23	-0.026	-0.041	27.152	0.250
		24	0.105	0.100	34.284	0.080
		25	-0.033	-0.065	35.004	0.088
		26	0.005	0.006	35.018	0.111
		27	0.078	0.055	38.983	0.064
		28	-0.002	-0.001	38.987	0.081
		29	0.060	0.048	41.363	0.064
		30	-0.044	-0.040	42.637	0.063
		31	0.044	0.043	43.938	0.062
		32	0.006	0.016	43.962	0.077

4.3.4 ARCH test

A plot of the log returns displays the volatility-clustering phenomenon, large and small swings tend to cluster, see Figure 3. it is clear that this will turn out to be important for the measurement of risk; the Value-at-Risk will be higher in tumultuous times than when the financial markets are smooth. Furthermore, the maximum and minimum statistics are quite large in absolute value, indicating the presence of extreme returns, which is supported by the occasional extreme spikes in Figure 3 and the high sample kurtosis, which is indicative of the fatness of the tails of the distribution.
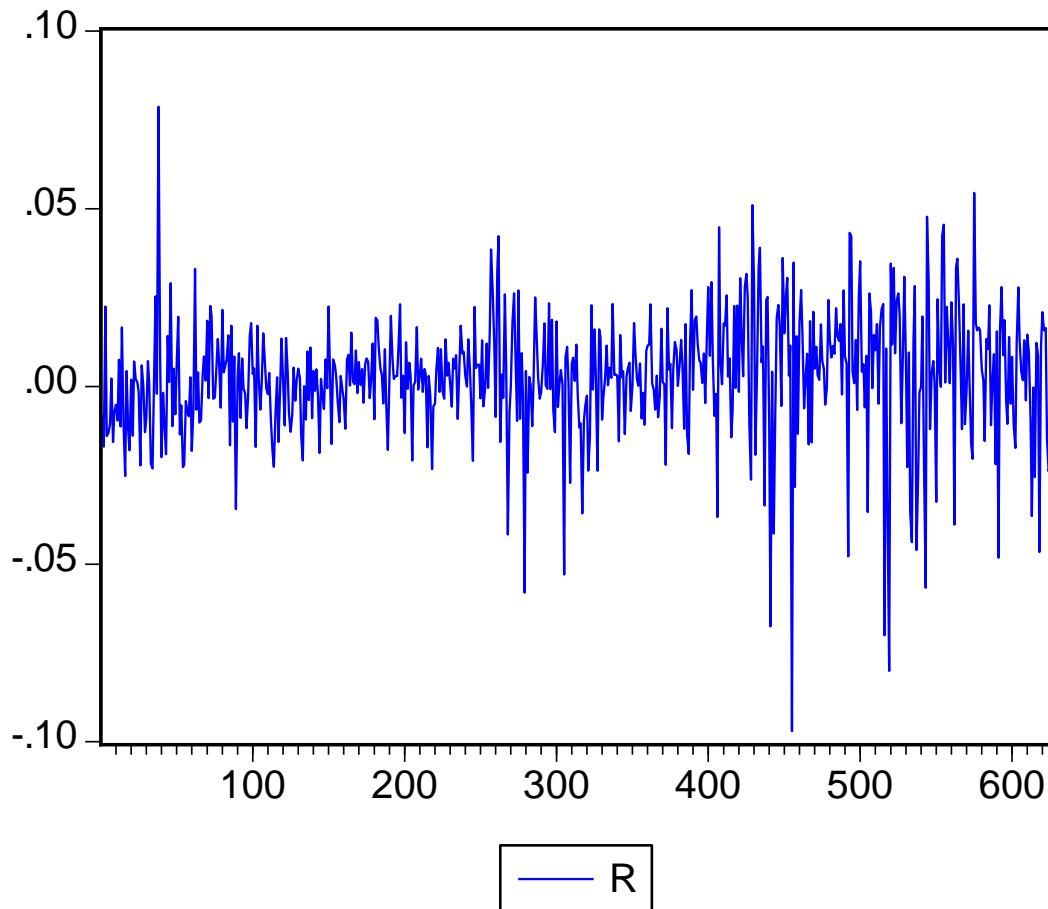
Figure 3. Plot of log return.

From the descriptive statistic analysis and the observation of the daily return graph, there should be autoregressive conditional heteroskedasticity (ARCH) in the residuals. We observe the correlograms of the squared residuals first, the autocorrelations and partial autocorrelations are not zero at all lags and the Q-statistics is significant up for almost all lags. The Histogram and Normality Test showed that the histogram is not bell-shaped and the Jarque-Bera statistic is significant and lead to the rejection of the null hypothesis of a normal distribution. The result of ARCH LM test up to order q=8 is smaller than 0.05 and it lead to the rejection of the null hypothesis that there is no ARCH in the residual up to order 8. It shows that there is high order ARCH in residual series.

Table 3. ARCH LM test.

ARCH Test:			
F-statistic	2.44774	Prob. F (8,597)	0.01303
Obs*R-squared	19.2458	Prob. Chi-Square (8)	0.01359

4.3.5 Model forecasting volatility of financial asset return

1. EARCH model

Use AIC criteria; EARCR (1,1) model is selected as the model to forecast volatility of Chinese stock market, form of EARCR (1,1) is as equation (6)

$$(6) \quad \ln(\sigma_t^2) = \alpha_0 + \theta_1 \ln(\sigma_{t-1}^2) + a_1 \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

The regression model based on 500 data is as follow:

$R_t = 0.002747 + 0.089256 R_{t-11}$

   (4.416660) (2.191681)

$$\ln(\sigma_t^2) = -0.197428 + 0.983921 \ln(\sigma_{t-1}^2) + 0.086715 * \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + 0.058967 * \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}$$

       (-1.870508) (90.72243)   (2.233092)   (2.878253)

ν=5.311335

  (4.400924)

All the regressed parameters are statically significant except the constant of the GARCH equation, which means that it shows that the leverage effect is not obvious in Chinese Stock market; there is no asymmetry information effect. Market participants do not react more strongly to negative information than positive information.

2. RiskMetrics model

The form of RiskMetrics model is presented in equation (7), which is:

(7) $h_t = \lambda h_{t-1} + (1 - \lambda)\varepsilon_{t-1}^{2}$

The parameter λ, also called the decay factor, determines the relative weights placed on previous observations. In this paper 0.94 of a λ value is used. This valued was calculated by JP Morgan using minimum RMSRB criteria based on financial data from western financial markets. It is generally used for calculating dairy VaR.

4.4 Calculation of VaR

4.4.1 Variance-covariance method

Based on the formula $VaR_t = -Z_a\sigma_t$, the calculation of VaR using VC method can be divided into several steps: first, the value of Za, which is determined by the asset distribution assumption and confidence level. In this paper, it is assumed that the return of HS300 follows a t distribution with a degree of freedom ν=5, this assumption is based on the regression results of past 500-day observations. Hence the critical value Z_a is 2.57 accordingly. Second, the value of standard deviation σ_t, from previous discussion, the dynamic fluctuation of HS300 will be captured using EARCH model and RiskMetrics model to forecast σ_t. The regression results of EARCH model were presented in section 3. Based the regression results, predicting σ_t for the next following 125 and 50 days is available and thus the 125 and 50 daily VaR using formula $VaR_t = -Z_a\sigma_t$. The calculation

process using RiskMetrics is similar with the one using EARCH model, the difference is that decay factor 0.94 is given and is stable. The predicting VaR used VCEA and VARM (with evaluation sample of 125 days and 50days) and the realized daily returns are presented from figure 4 to figure 8 respectively. Both regression of parameters and graphs are realized by Eviews 5.0.
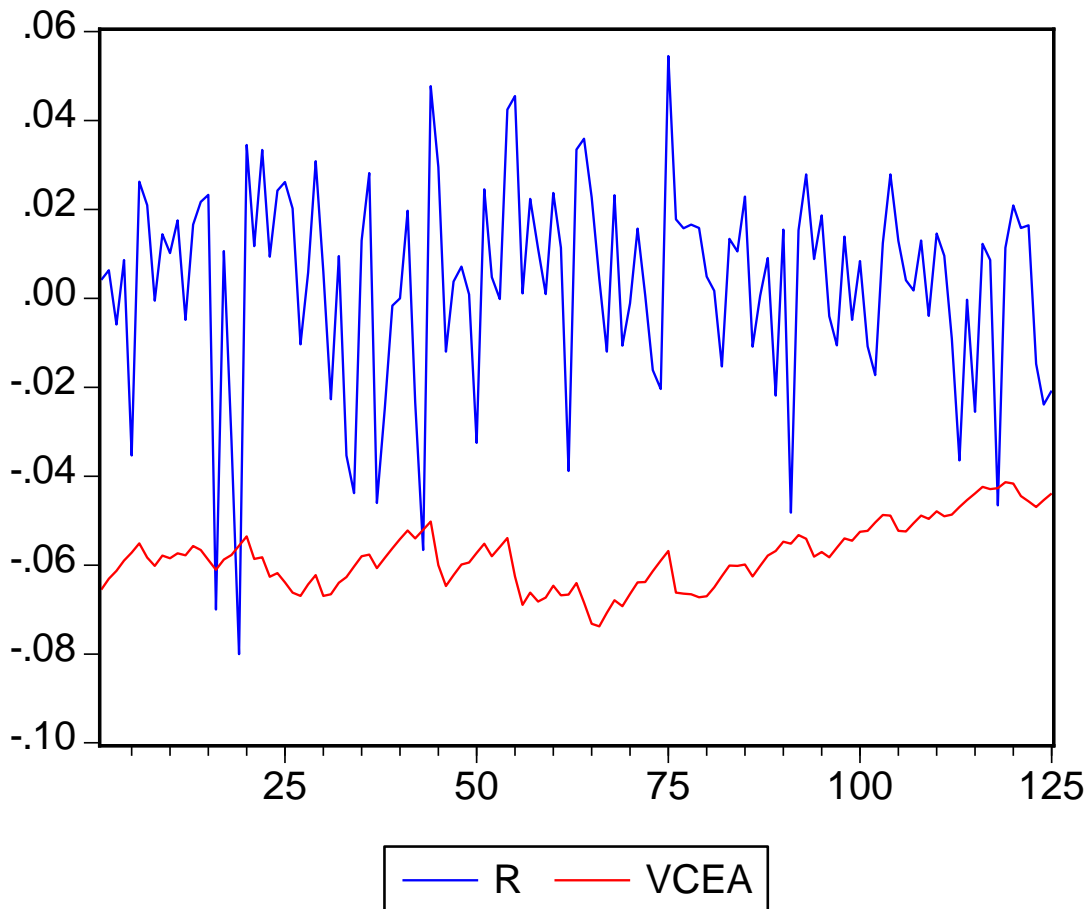

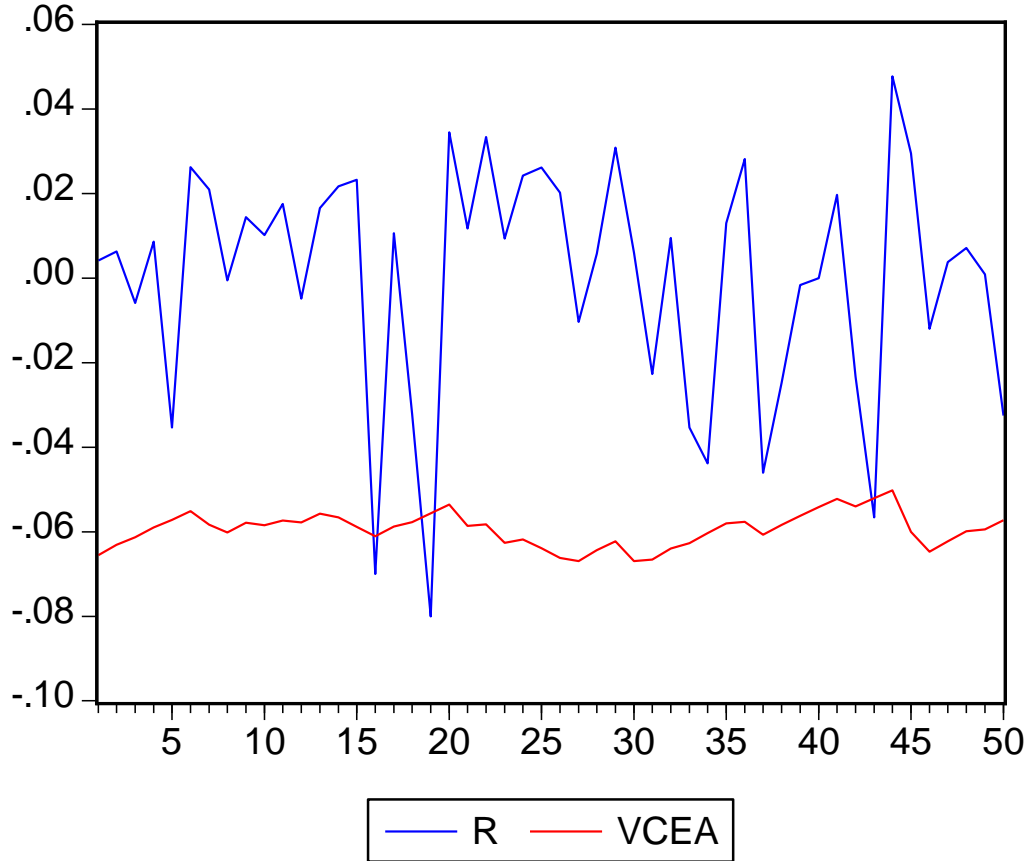**Figure 4.** VCEA with a 125 days evaluation sample.

Figure 5. VCEA with a 50 days evaluation sample.

VaR calculated by VCEA and the realized daily returns are compared in figure 4. It is showed that there are four VaR that was "passed though" by real return. The phenomenon that a calculated VaR is smaller than a real return is marked as "failure" or "exception". With a 5% of confidence level, the number of failure for a 125 days sample should be 6. The first impression from figure 4 maybe that VCEA overestimated market risk of Chinese stock market since the "failure" happened is less than expected. However, more evaluation towards each VaR calculation method will be presented in section 5. Considering the comparison of VaR with realized dairy return for a 50-day evaluation sample, the number of "exception" is 3, which is quite closed to the expected number of 2.5 under a 5% confidence level. It is indicated from the results that the performance of VaR calculation methods is affected by the length of evaluation period. It also shows that for VCEA method,

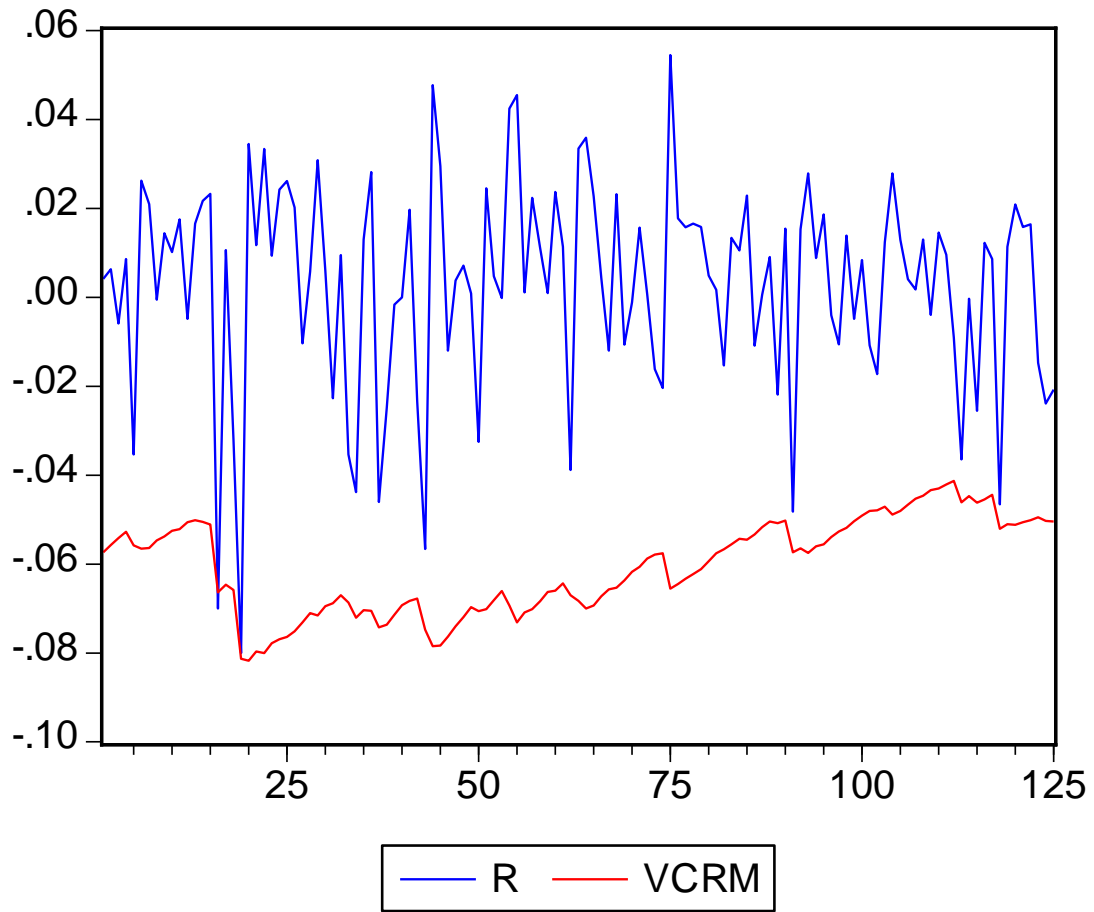the "exception" mainly in short run evaluation window, which indicates a cluster of "exception".

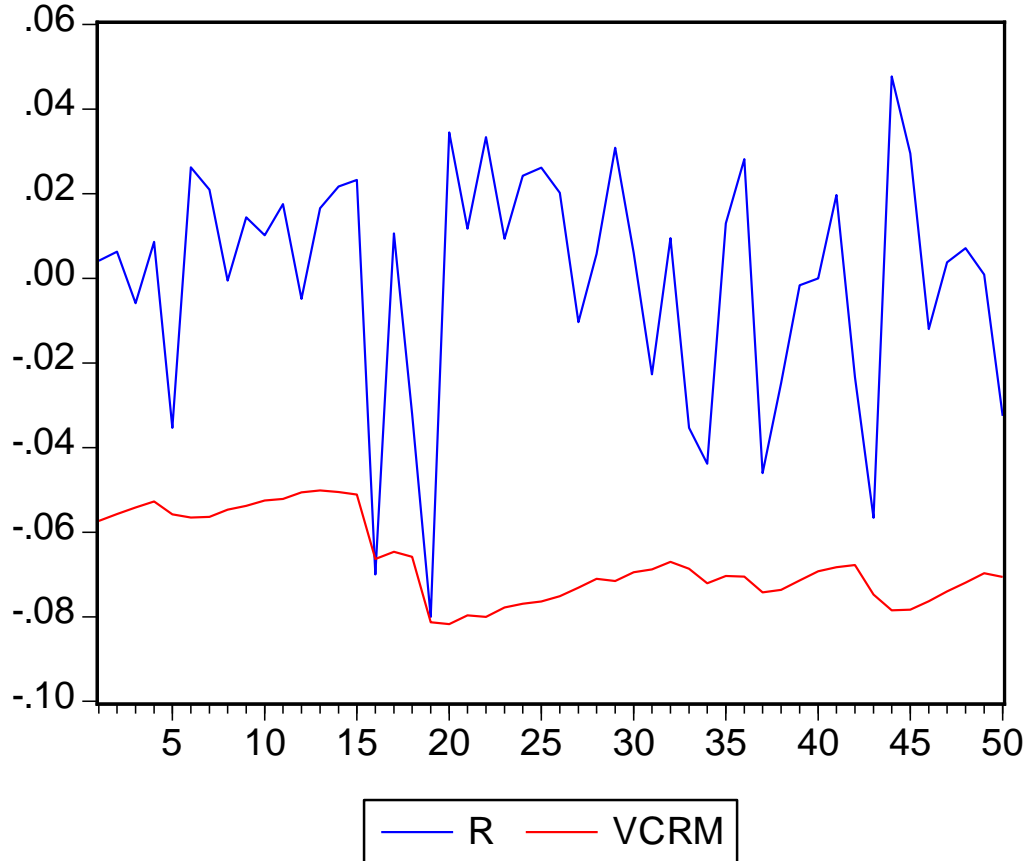**Figure 6.** VCRM with a 125 days evaluation sample

Figure 7. VCRM with a 50 days evaluation sample.

Figure 6 and figure 7 are the results of VaR calculated by VCRM compared with real daily return. For the 125-day evaluation sample, it is showed from figure 6 that number of exception of VCRM is only 1, far less than the expected number of 6, which indicates that VCRM overestimated financial market risk. When comparing the VaR and realized dairy return for a 50-day evaluation sample, the "exception" happened is 1 while the expected number of failure under a 5% confidence level should be 2.5, VCRM overestimated the market risk of Chinese stock market. It is also noticeable from figure 7 that the "exception" happened between days 15-20, for evaluation period outside these days, there is no exception.

4.4.2 Monte Carlo method

The process of MC method are as following:

(1) Use Geometric Brownian Motion to reflect the changing of stock market index:
$$S_{t+1} = S_t + S_t(\mu\Delta t + \sigma\varepsilon_t\sqrt{\Delta t})$$

(2) Estimate the parameters of Geometric Brownian Motion. For the purposes of our example, it is assumed that the stock market index returns are drawn from a t-distribution with 5 degrees of freedom. μ is the mean value of returns from the 500 historical data while σ_t is obtained from both EARCH (1,1) model and RiskMetrics model.

(3) Simulate 1000 random draws following t distribution with a 5 degree of freedom to reflect random market factors.

(4) Get 1000 simulated stock prices and stock returns for day t+1, sort from least value to largest value. And follow the same process with HS to get the VaR for day t+1.

(5) Repeat steps 3) and 4) 125 times to get the next following 125 days VaR for day t+1 to day t+125.

Figure 8 and figure 9 provide a view of VaR calculated by MCEA and the realized daily return for an evaluation window of 125 days and 50 days. It is visible from figure 8 that there are 5 "exception" VaR by MCEA for a 125-day evaluation sample, the expected numbers of "exception" under a 5% of confidence level is 6. As far as a 50-day evaluation window is concerned, the expected number of failure is 2.5 while the realized "exception" is 3. "Exception" from both samples is very close the expected numbers, which indicate a relative well performance of MCEA method. Meanwhile, "exception" not only happened at the beginning of the evaluation sample as other calculation methods but also happened during the whole sample.
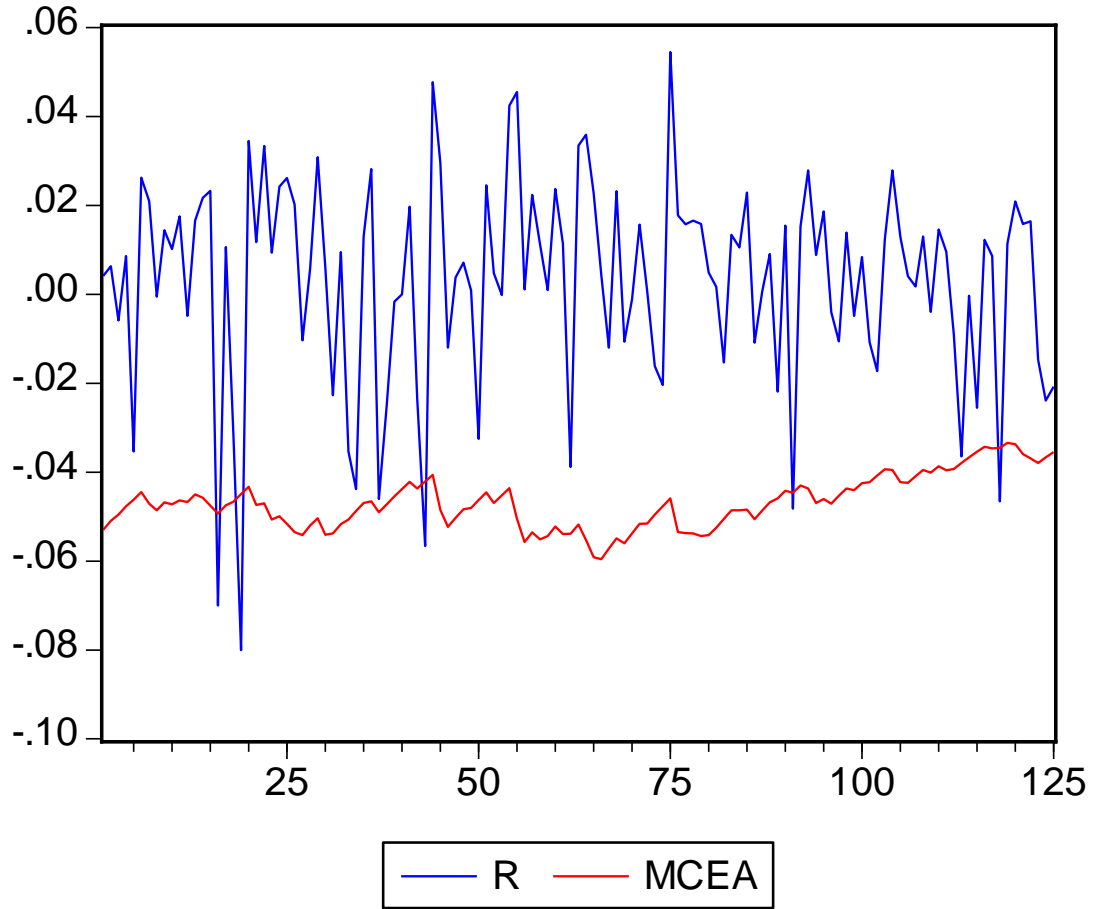
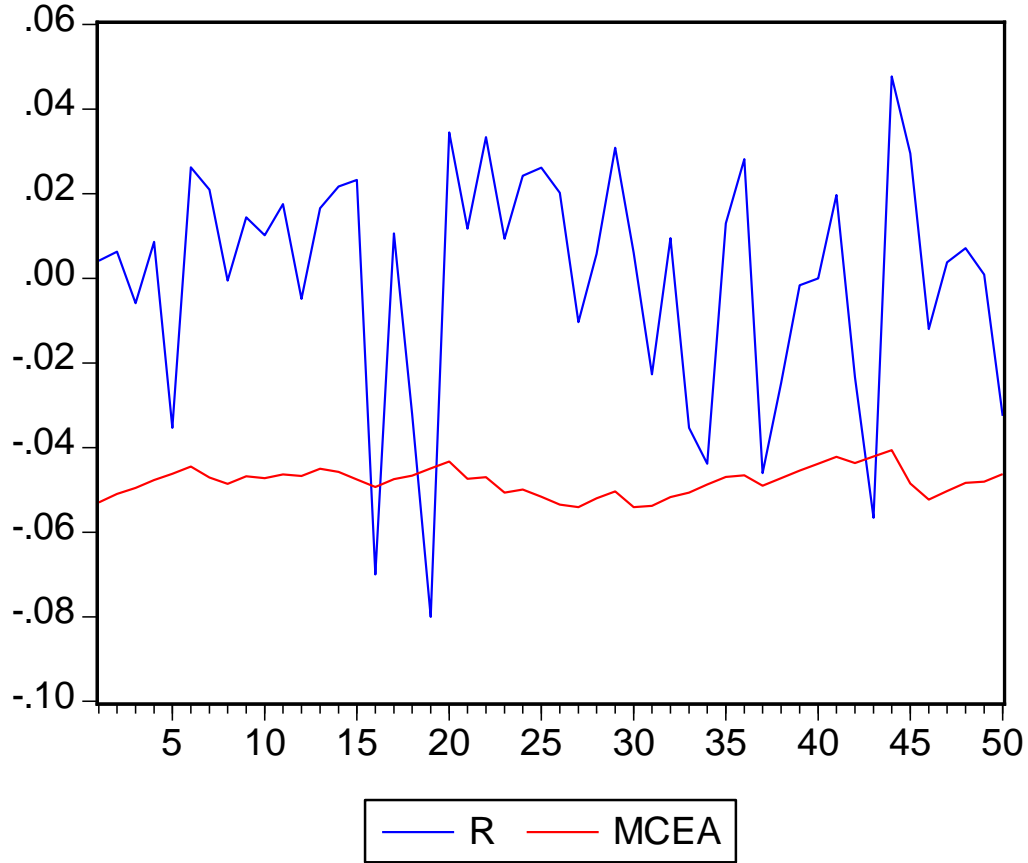**Figure 8.** MCEA with a 125 days evaluation sample.

Figure 9. MCEA with a 50 days evaluation sample.

Figure 10 is the VaR calculated by MCRM and the real daily return for 125-day sample. The realized "exception" is less then the expected number (4<6), which means that MCRM in a 125-day evaluation sample overestimates the market risk of Chinese stock market. Figure 11 shows the comparison of VaR with realized returns in a 50-day evaluation sample, the number of exception is 2, which is also quite close to the expected number of 2.5.
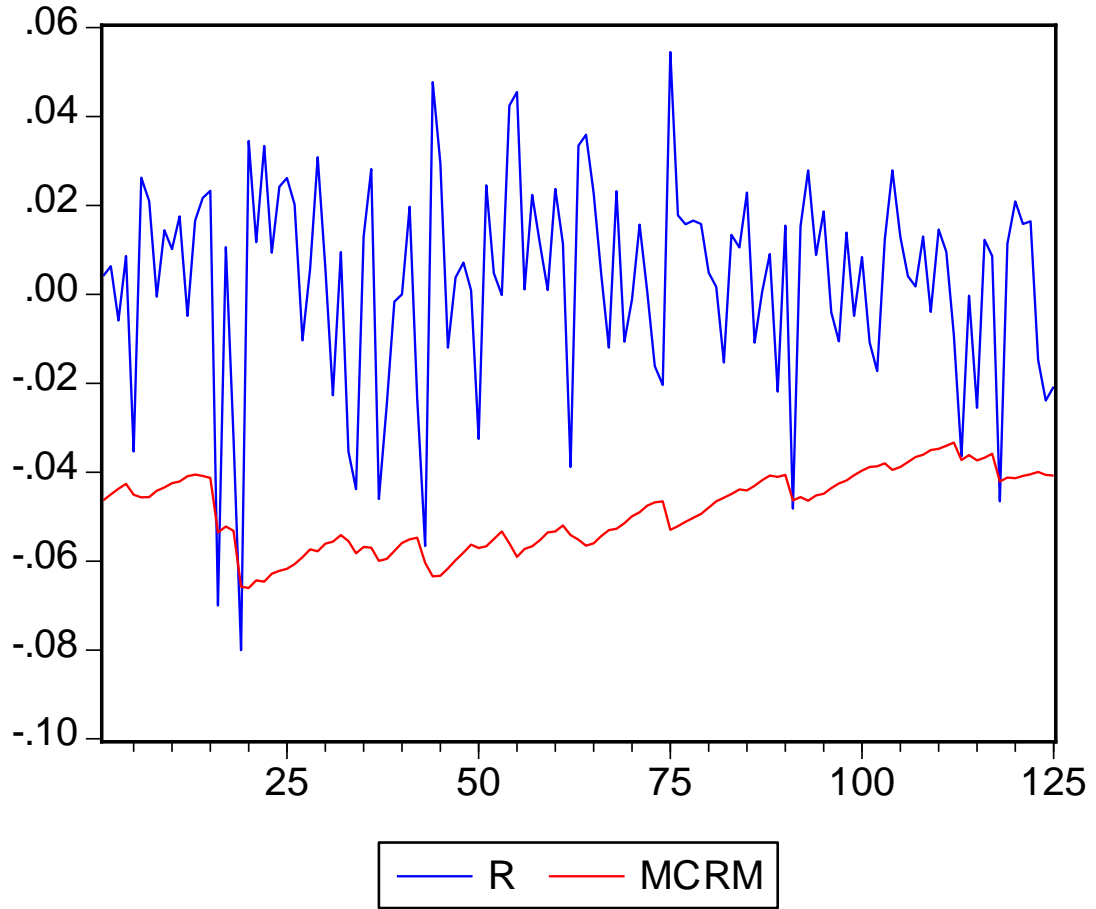
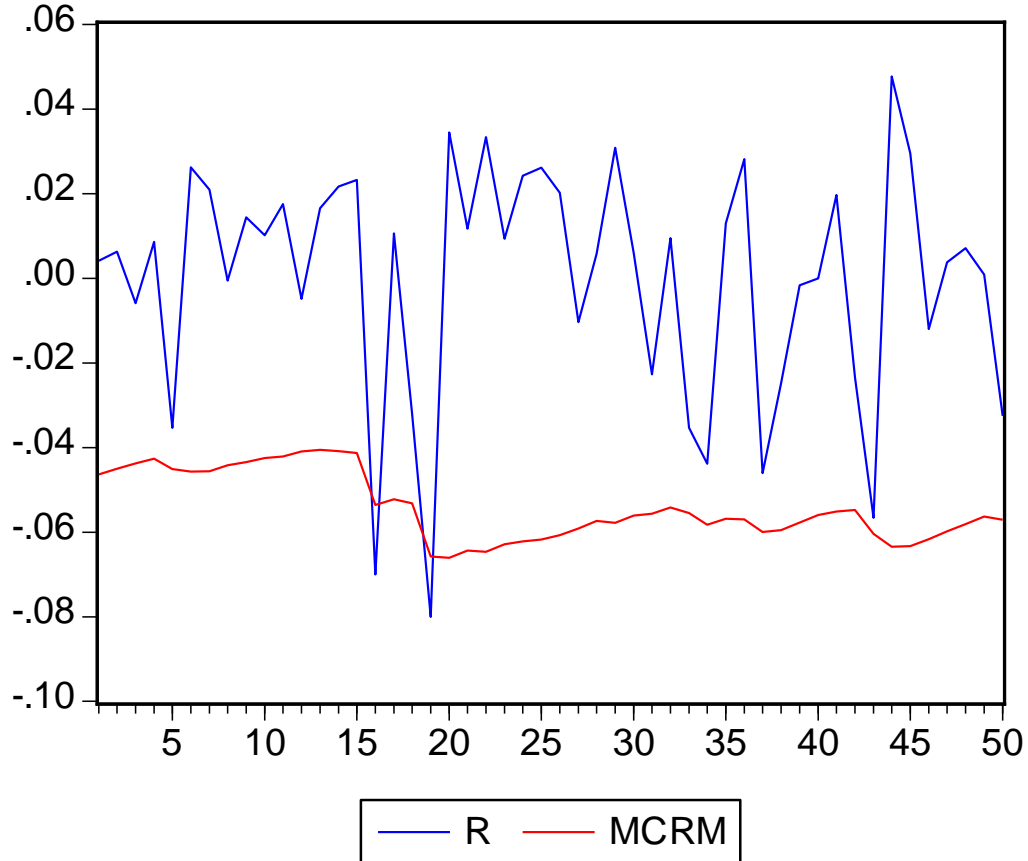**Figure 10.** MCRM with a 125 days evaluation sample.

Figure 11. MCRM with a 50 days evaluation sample.

It is indicates from figure 10 and figure 11 that MC method (MCEA and MCRM) performs very well for both long evaluation sample and short evaluation sample. The performance in 50-day sample is especially well because for both models, the numbers of exception are very close to excepted number. However, the reason may come from the limit number of the sample.

From the calculation process and results, it can be concluded that MC method is conceptually simple but is generally computationally intensive. Huge amount of future possible return data are simulated and VaR is calculated based on those simulated return. Because Monte Carlo simulation method can simulate hundre of thousands of return possibility and it reflects uncertaity of financial assets price, so it is the most sientific and

predictable calculation method. However, the assumption behind MC method is that that financial asset return follows some distribution, the distribution of return need to be assumed when we use computer program to generate random data. In this paper, out assumption is based on the results regressed from historical data.

4.4.3 Historical Simulation

Formula VaRt=μ-r* is used to calculate VaR for period 08/05/2007 to 02/11/2007. The estimation window is 500 return data before the calculation date. For example, when we calculate VaR of 08/05/2007, average of returns of 500 transaction days before is considered as the expected return of this day, which is μ. While the minimum return under 5% of confidence level of this day is one that ranks the least 25 (500*5%) from these 500 returns. Then follow formula VaRt=μ-r* VaR of 08/05/2007 is obtainable. VaR for the next following 124 and 50 days are calculated using the same way. The results of VaR by HS are realized by Micro Excel and the graphs are obtained by Eviews 5.0.
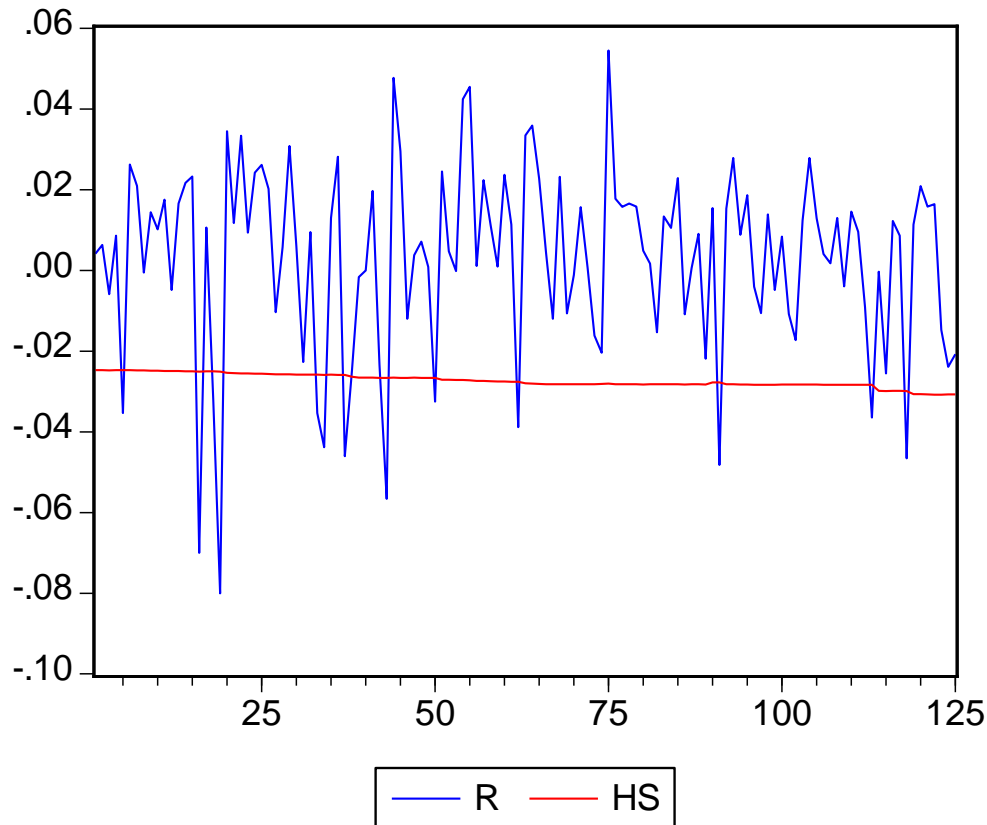

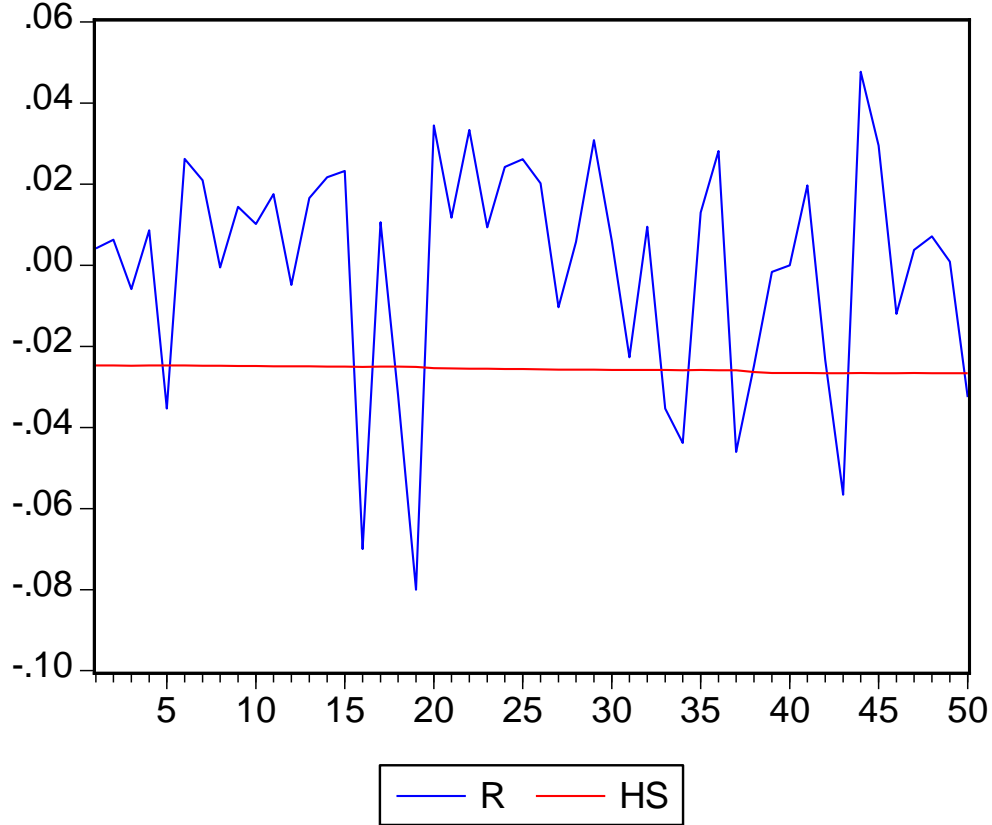**Figure 12.** HS with a 125 days evaluation sample.

Figure 13. HS with a 50days evaluation sample.

Figure 12 presents the VaR calculated by HS and the real return for 125-day samples. It is showed from the figure that the number of "exception" for HS is 13, which is far more then the expected number of 6. Result indicates that HS has underestimated the market risk of Chinese Stock market. The number of "exception" for 50-day sample is 9, which is also much higher than the expected number of 2.5. It shows from the results that HS method underestimate Chinese Stock market risk for both 125-day and 50-evaluation window. It also shows that most of the exception happened at the beginning of evaluation window.

From the calculation processed and the results, implication about the advantages and disadvantages of HS method may be obtained. HS is the simplest and most transparent method of calculation; we just need to re-organizes actual historical returns and put them in order from worst to best. It doesn't require any restriction on the return distribution and just

get random sample from historical return series. The benefits of this method are its simplicity to implement, and the fact that it does not assume a normal distribution of asset returns and thus avoids the fat tail problem of the financial data. The assumption behind this method is that history will repeat itself from a risk perspective. However the past value can only reflect the changing of very near future. And as appointed by Engle, the fluctuation of financial asset is frequent and can only be forecast in short period, when there is large fluctuation of asset return in short run, historical simulation will not be an accurate method. Further more, this method gives same weight value to all observation, which is not with fit to the reality. Usually the nearer the observation value, the more effect it should be on the future value. But even some researcher set different weighting value on the sample to avoid this problem, the assigned value is subjective and too much depends on experience. Finally, the accuracy of this method is relying on the length of the estimation sample, if the sample volume is too small, the estimation of VaR will not be accurate. Large market database are required and computationally intensive calculation is needed especially for porfolio with complicative construction.

4.5 Evaluation of each method

Evaluation of the above three VaR calculation methods based on 5 models will be processed in terms of test of acceptability, variability, accuracy and measurement error. VaR calculated using Variance-Covariance method based on EARCH model will be simplified as VCEA, Variance-Covariance method based on RiskMetrics will be VCRM, Monte Carlo simulation modified by EARCH will be MCEA, Monte Carlo simulation modified by RiskMetrics will be MCRM and historical simulation will be simplified as HS in the following contents. All the results obtained for the evaluation test are realized using Microsoft Excel.

4.5.1 Acceptability

The first of acceptability test is to calculate the maximum likelihood of the unconditional correct coverage of each model introduced by Kupiec (1995). P is the loss failure rate defined by given probability, in this paper we use p=0.05, T is the total number of evaluation sample period, that is 125 and 50 days in this paper, while N is the number of

"failure" that occurs for each VaR calculation model. N for all the methods are reported in the Table 4

Table 4. Number of failure.

Methods	VCRM	VCEA	MCRM	MCEA	HS
N (T=125)	1	4	4	5	13
N (T=50)	1	3	2	3	9

The next step is to follow the formula from Kupiec and calculate the correct unconditional coverage LR_{uc} of each method, after that, compare these LR_{uc} to the interval value of a LR_{uc} which has a chi-squared distribution with one degree of freedom ($\chi^2_{1,a}$), in this paper, a=0.05 hence the interval value for LR_{uc} is 3.841455. If a calculated LR_{uc} is greater then 3.841455, that will leads to a rejection of acceptance of the VaR method.

The steps of calculating correct conditional coverage LR_{cc} is similar with LR_{uc}, however, the LR_{cc} calculated will be compared with the interval value of LR_{cc} which has a chi-squared distribution with two degree of freedom ($\chi^2_{2,a}$). The interval value for LR_{cc} that under a 0.05 of confidence level is 5.991476. If an LR_{cc} calculated is smaller than this value, then it will lead to an acceptance of the underlying VaR method.

The results of LR_{uc} and LR_{cc} with 125 days and 50 days evaluation sample are presented in Table 5 and Table 6.

Table 5. Results of LR_{uc} with 125 days and 50 days evaluation sample.

Method	T=125	T=50
VCRM	7.063595	1.256379
VCEA	0.972068	0.099211
MCRM	0.972086	0.112671
MCEA	0.281676	0.099211
HS	5.932733	10.98988

Table 6. Results of LR_{cc} with 125 days and 50 days evaluation sample.

Method	T=125	T=50
VCRM	7.079921	1.256379
VCEA	1.239050	0.492088
MCRM	1.239050	0.28378
MCEA	0.702303	0.492088
HS	14.26835	18.14761

4.5.2 Variability

To implement variability test, the first procedure is to calculate the MRB of each method based on approaches introduced by Engel & Gizycki (1999), the total evaluation sample periods in this paper are 125days and 50 days. $\overline{VaR_t}$ is the mean value of VaR calculated using 5 different models at horizon t, and then the difference between VaR of each model with $\overline{VaR_t}$ will be divided by this mean value. Finally all these difference percentage should be summarized and divided by total sample 125 and 50 and we will get the results of MRB of each model. The mean relative bias statistic captures the degree of the average bias of the VaR of the specific model from the all-model average, hence the larger the MRB value, the more variability a VaR model will have. However, the MRB measure is in terms of the relative but not absolute concept. The bias of the model evaluated by MRB is not absolutely. To calculate the RMSRB of each model, which is introduced by Hendricks (1997), the process is similar; the only difference is that we use absolute difference instead of the above relative one. RMSRB is a better reflection of the bias of means of estimation towards the means of all estimation methods. Hence when analyze the results; the absolute value of MRB and RMSRB will be analyzed respectively. The results of MRB and RMSRB with both 125-day and 50-day evaluation sample are showed in table 7 and table 8.

Table 7. Results of MRB with a 125-day and 50-day evaluation sample.

Method	T=125	T=50
VCRM	0.242362	0.311691
VCEA	0.200751	0.173495
MCRM	0.007265	0.060061
MCEA	-0.0296	-0.05162

| HS | -0.42478 | -0.49362 |

Table 8. Results of RMSRB with a 125-day and 50-day evaluation sample.

Method	T=125	T=50
VCRM	0.265421	0.32882
VCEA	0.27118	0.197226
MCRM	0.080147	0.103793
MCEA	0.073096	0.091711
HS	0.434964	0.49479

Due to the relative nature of MRB, we will compare and range the absolute value of MRB and RMSRB of each method to observe the variability of the 5 tested VaR calculation methods. The results of absolute MRB and RMSRB will be showed from figure 14 and figure 17.
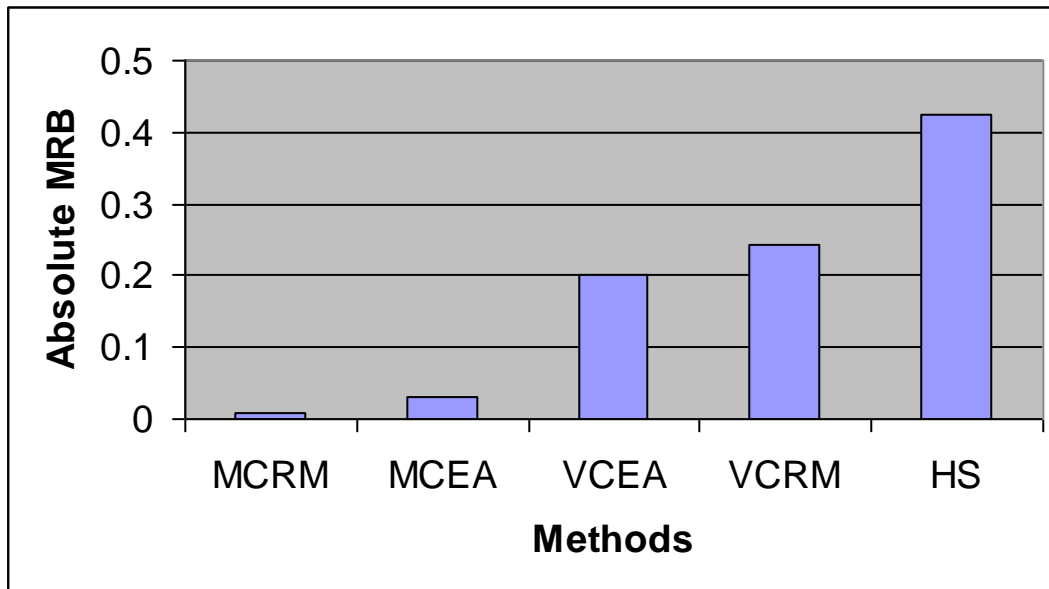

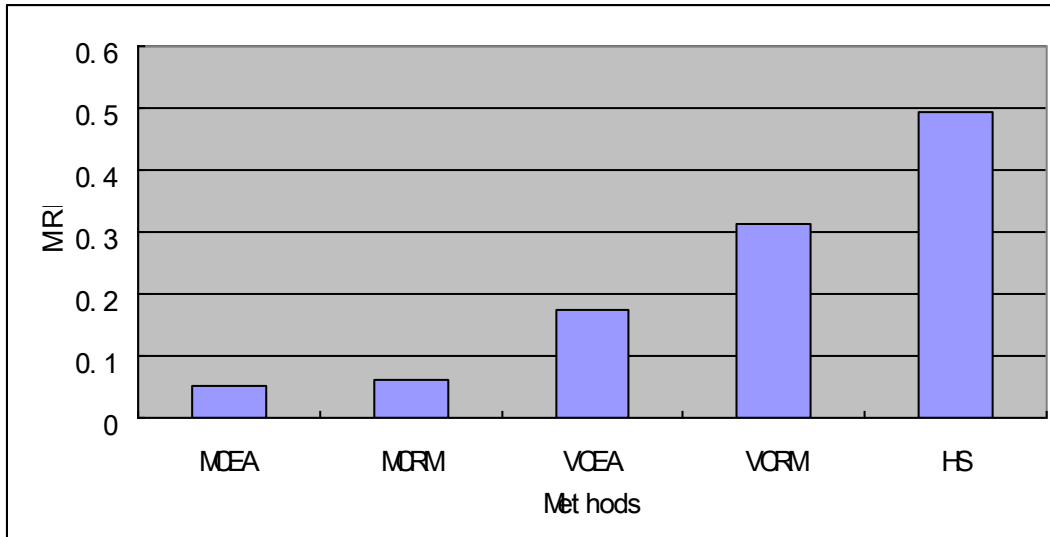**Figure 14.** Range of absolute MRB with 125-day evaluation sample.

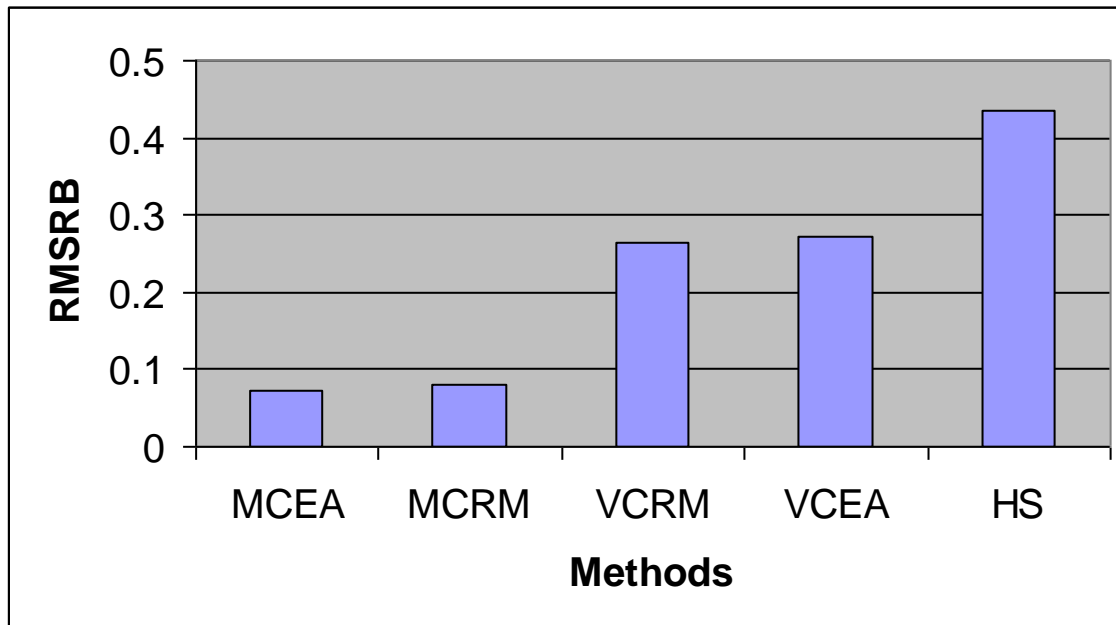**Figure 15.** Range of absolute MRB with 50-day evaluation sample.

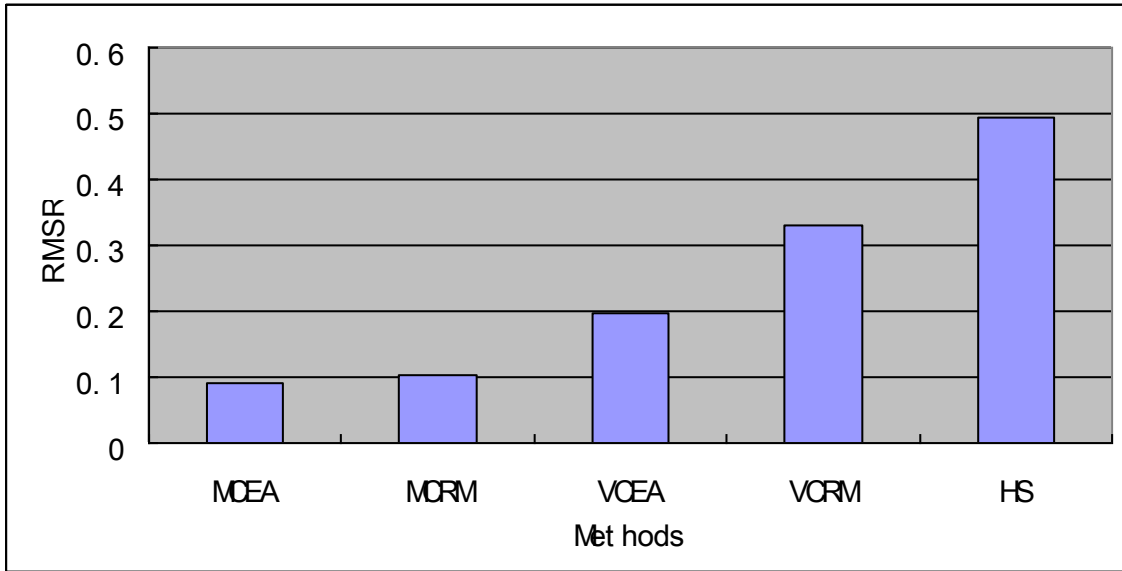**Figure 16.** Range of RMSRB with 125-day evaluation sample.

Figure 17. Range of RMSRB with 50-day evaluation sample.

4.5.3 Accuracy

According to Binary loss function, first the forecast daily VaR with the realized return of index for a 125-day and 50-day evaluation window should be compared. If the realized daily return exceeds the value of forecast VaR, it is considered as failure or exception and gives a number of 1; a number of 0 will be given for other cases. Then summarize the number of "failure" and divide this number by 125 and 50 of total sample and we will get the value from Binary loss function, which is marked as BLF in the table 9.

The key of a binary loss function is the summarize number of failure or exception, it only considers the number that an exception happens but doesn't consider about the serious that exception happens, a equal weight is given to all the forecast VaR that is small then the realized return. If the VaR calculated using particular model truly reflects the risk situation under given probability, for example, with a 5% of confidential level, the value of BLF should be 0.05 if the model is exactly reflects the real situation. The closer this value to 0.05, the more accurate the model will be.

While a Quadratic loss function also considers about the magnitude that one "exception" occurs. As with Binary loss function, first we give a number of 1 to the VaR that is smaller than the realized return and 0 to others, meanwhile the magnitude will be examined by

calculating the square of difference of a VaR and a realized return when it is an exception, then summarize 1 and this calculated square difference together. Finally compute all these numbers and divided by 125 and 50 of total sample days and we will get the result of a quadratic loss, which is expressed as QLF and showed in the three column of the following table. A VaR model that has a QLF number equals or very close to 0.05 will be a good model in this case. Results of QLF with 125-day and 50-day evaluation sample will be reported in table 10.

Table 9. Results of BLF with 125-day and 50-day evaluation sample.

Methods	T=125	T=50
VCRM	0.008	0.02
VCEA	0.032	0.06
MCRM	0.032	0.04
MCEA	0.04	0.06
HS	0.104	0.18

Table 10. Results of QLF with 125-day and 50-day evaluation sample.

Methods	T=125	T=50
VCRM	0.008	0.02
VCEA	0.032006	0.060014
MCRM	0.032004	0.040009
MCEA	0.040016	0.060037
HS	0.104063	0.180139

4.5.4 Measurement error test

By regressing hit of each VaR calculation method with its own lags, it is noticeable that there is no autocorrelation for all the lags of hit. While when examine the correlation between the hit and a constant, as well as the correlation between the hit and the current VaR, a QLS regression method was used. The results of Hit test with 125-day evaluation sample and 50-day evaluation sample are showed in table 11 and table 12 respectively.

Table 11. Results of Hit test with 125-day evaluation sample.

	C	VaR
Methods	P-value	P-value
VCEA	0.2071	0.1576
VCRM	0.1540	0.5834
MCEA	0.1745	0.1501
MCRM	0.3677	0.4724
HS	0.0445	0.0576

Table 12. Results of Hit test with 50-day evaluation sample.

	C	VaR
Methods	P-value	P-value
VCEA	0.1518	0.1564
VCRM	0.8893	0.9426
MCEA	0.1518	0.1564
MCRM	0.3004	0.3194
HS	0.9571	0.9073

5. RESULTS AND FINDINGS

5.1 Main findings from the acceptability test

Considering the acceptability of ach method in 125-day evaluation sample, the results of both unconditional correct coverage test (LR_{uc}) and conditional correct coverage test (LR_{cc}) show that the LR of both VCRM and HS are greater than two interval values, which means that both VCRM and HS are unacceptable in calculating the market risk of Chinese stock market. The LR values of other methods are all smaller than the interval value of LR, statistically; they are falling within the acceptance interval in measuring market risk for Chinese Stock market. Among all these values, the LR_{uc} and LR_{cc} of MCEA are the smallest ones, which are 0.281676 and 0.702303 respectively. From the above discussion about the correct coverage test, it is obvious that the more this values close to 0, the more exact coverage of VaR the methods perform. It is noticeable from the results that the VaR using MCEA has covered most of the loss happened. The LR_{uc} and LR_{cc} of VCEA and MCRM show that these two methods are acceptable VaR calculating methods because both of their LR values are smaller than the interval value. As far as VARM is concerned, the results indicate that its LR values, both conditional and unconditional is greater than the interval LR, which leads to a rejection of using this method to calculate VaR for Chinese stock market index. It is showed from the results that the "exception" from this method is far less than the number that expected under the 5% of given confidential level, which means that VARM has overestimated the risk of Chinese stock market. The possible reason of overestimate from this method may due to the way that the standard deviation generated. The standard deviation is calculated using RiskMetrics model modified with a t distribution assumption about the return series. Results indicted that t distribution assumption is not the main reason that a great error happened here, the main reason is that a RiskMetrics model applied in this paper can not accurately capture the volatility of Chinese stock market. It is understandable that a 0.94 of decay factor (λ) may not be suitable for Chinese market since this λ value is calculated based on western developed financial market so when applied to developing countries it will generate some bias in predicting the standard deviation. Meanwhile, the decay factor should be dynamic, but in the RiskMetrics method, it is set to be a constant, which cannot reflect the dynamic change of innovation of financial market. Considering about the large value of LR_{uc} and LR_{cc} of HS. The "exception" noticed from

this method is far more than the number expected, which means an underestimate of market risk by this method. This underestimate leads to low correct coverage and thus a high value of LR. HS is unacceptable to be applied in calculating VaR in this case. Possible reason of such great bias comes from the limitation of historical simulation itself. A 500 days of estimation window were used to predict the expect return and minimum return of the coming day, this window may be too long because whenever innovation happens to the market, the historical information is not able to adjust to the new innovation and hence can't not reflect the current situation of the market.

The results of LR_{uc} and LR_{cc} for a 50-day evaluation sample are somehow diversified. Results of both unconditional correct coverage test (LR_{uc}) and conditional correct coverage test ($LR_{cc)}$ show that the LR of HS is greater than two interval values, which means that both HS is an unacceptable method in calculating the market risk of Chinese stock market. The LR values of other methods are all smaller than the interval value of LR, statistically; they are falling within the acceptance interval in measuring market risk. Among all these values, the LR_{uc} and LR_{cc} of VCEA and MCEA are the same, which is 0.099211 for LR_{uc} and 0.492088 for $LR_{cc.}$ Results of these two values indicate that in a 50-day evaluation sample, VCEA and MCEA perform same level of exact coverage of VaR. It is noticeable from the results that the VaR using VCEA has covered most of the loss happened. The LR_{uc} and LR_{cc} of VCEA MCEA and MCRM show that these three methods are acceptable VaR calculating methods because both of their LR values are smaller than the interval value. As far as VCRM is concerned, the results indicate that its LR values, both conditional and unconditional are smaller than the interval LR, which leads to an acceptance of using this method to calculate VaR for Chinese stock market index. It is showed from the results that the "exception" from this method in 50-day sample is less than the number that expected under the 5% of given confidential level but still under an acceptable interval. Compare with the rejection of this method in 125-day sample, the acceptance of this method is because the only 1 exception happen is within this first 50-evaluation period, it indicates that the forecast ability of RiskMetrics model modified with a t distribution assumption about the return series is stronger in short run then in long run. The disadvantage of a stable decay factor is weaker in short-term, the difference of between the forecast volatility by this model and the real dynamic innovation of financial market is smaller in a 50-day evaluation sample. Considering about the large value of LR_{uc} and LR_{cc} of HS. The "exception" noticed from this method is far more than the number expected, which means an underestimate of

market risk by this method. This underestimate leads to low correct coverage and thus a high value of LR. HS is unacceptable to be applied in calculating VaR in 50-day evaluation as well.

From the discussion we know that HS is unacceptable for both 125-day and 50-day evaluation sample while VCRM is unacceptable for a 125-day evaluation sample, hypothesis 1 is rejected based on the empirical results.

5.2 Main findings from the variability test

From results of MRB and RMSRB for a 125-day evaluation sample, the variability of each VaR calculation method can be judged by absolute MRB and RMSRB. The absolute MRB (0.42478) and RMSRB (0.434964) of HS are the highest among all the tested methods; it can be concluded that the variability of HS is the highest among all these VaR calculation methods for Chinese stock market. As discussed above, RMSRB is a better predictor of variability then MRB, when results of these two criteria are conflicted, RMSRB is preferred and will be used as a main criterion. It is notice from the results that RMSRB of MCEA is only 0.073096, which indicates that MCEA has the smallest variability and hence again proves that MCEA is quite good a method in calculating VaR in Chinese stock market. While RMSRB of MCRM is 0.080147, which is the second smallest among these evaluated methods, it shows that variability of MCRM is also relatively low compared with VC methods and HS. It is easy to notice that variability of MC, no matter MCEA or MCRM are quite low, but MCEA has better predicting ability then MCRM. As far as VC method is concerned, RMSRB of VCRM is 0.265421, which is higher than MCRM but lower than VCEA, this is a little bit out of expectation because we expected that variability of VCRM should be higher than MCRM and also higher than VCEA, but the results show that VCRM has lower variability (0.265421) then VCEA (0.27118) although the difference of variability between these two models are very small.

The results of RMSRB for 50-day evaluation sample are similar with the results for 125-day sample, RMSRB (0.49479) of HS are the highest among all the tested methods; the variability of HS is the highest among all these VaR calculation methods for Chinese stock market. RMSRB of MCEA is only 0.091711, which indicates that MCEA has the smallest

variability and hence again proves that MCEA is quite good a method in calculating VaR in Chinese stock market. While RMSRB of MCRM is 0.103793, the ranking of variability among these three methods is consistent with the ranking for 125-day evaluation sample. Considering about VC method, RMSRB of VCEA is 0.197226, which is higher than MCRM but lower than VCRM (0.32882), this result is quite much expected since we expected that variability of VCRM should be higher than MCRM and also higher than VCEA.

Hypothesis 2 is supported by empirical results of RMSRB but not supported by results of MRB since the method that has the lowest variability will be MCRM based on ranking of absolute MRB.

5.3 Main findings from the accuracy test

Concerning results of accuracy tests for 125-day evaluation sample, it is suggested that for both BLF and QLF, the one that has most close to 0.05 BLF and QLF values is MCEA, which are 0.04 and 0.040016 respectively. It indicates that the accuracy of MCEA is the highest among all the methods used in this paper. This results is quite much expected because Monte Carlo method can simulate the possible path of stock price movement, combines with the modification from EARCH model, which can capture the volatility of stock prices, it is a very good method in calculating VaR of financial market. However, it is also noticeable that, even as accurate as MCEA, the BLF and QLF are not 0.05, which means that MC method itself still has some model risk. The accuracy of both the pricing model as well as the volatility should be improved to obtain a better prediction of VaR. for example, in the volatility model, t distribution assumption of return is a better assumption then normal distribution, but still it has some difference with reality. The accuracy of MCRM and VCEA are at the same level with BLF and QLF criteria. They are just right after MCEA. The possible reason that a VaR calculate by MCRM has lower accuracy than MCEA may come from the limitation of RiskMetrics model that discussed above. That is also why VaR calculated from VCEA has higher accuracy then the one from VCRM. The one that has most bias of BLF and QLF is HS. The accuracy of HS is the lowest accuracy among all the methods tested. Hence hypothesis 4 is confirmed by results of both BLF and QLF.

From results of accuracy tests for 50-day evaluation sample, it is suggested that for both BLF and QLF, those methods that have most close to 0.05 BLF and QLF values are MCEA and MCRM. For MCEA, the BLF and QLF are 0.06 and 0.060037 respectively. While for MCRM, the BLF and QLF are 0.04 and 0.040009 respectively. It is difficult to conclude from the results that which of these two methods has better accuracy; the high accuracy of Monte Carlo is confirmed again in this 50-day evaluation sample. Since MCRM performs well in short run, the accuracy of both MCEA and MCRM are quite high. The accuracy of VCEA is also very high in 50-day evaluation sample; BLF (0.06) and QLF (0.0014) are close to the expected value of 0.05. It is indicated from the results that there is no great difference among the accuracy of MCEA, MCRM and VCEA for short run evaluation window. The accuracy of VCRM is still lower than the above three methods, which is 0.02 for both BLF and QLF. The accuracy of HS in 50-day sample ranks the lowest, which is consistent with the results from 125-day evaluation window. It is noticeable from the analysis that the accuracy of HS is very low both for short run and long run horizon. Hence hypothesis 4 is confirmed by results of both BLF and QLF for 125-day evaluation window.

5.4 Main findings from the measurement error test

The results of measurement error test, or in another word, Hit_t test show that the regression of the Hit_t variables for all the calculation methods are not correlated with its past for all lags because all the regression coefficients are not statistically significant. Considering about the regression of constants on Hit_t variables, it is noticeable from the results with a 125-day evaluation sample that the regression a constant on Hit_t variable of HS method is statically significant at 5% confidence level (p=0.0445), which means that Hit_t is correlated with a constant at a 5% level. Moreover, the p value of a regression of VaR on Hit_t is 0.0576, which is statistically significant at 10% level. This indicates that Hit_t of HS also correlated with the its current VaR for a 125-day evaluation sample. It is showed from the results that Hit_t does not satisfy the condition of being uncorrelated, which means that there is measurement error for the predicting VaR of HS and its own Hit_t. The fraction of loss generated by HS is incorrect under given confidence level. However, The p value of regressions of both constants and VaR of on Hit_t of all other calculation methods are grater than 0.1, all the regression values are statistically insignificant. It shows Hit_t of all other VaR calculation satisfies the conditions of being uncorrelated, there is no autocorrelation in

the hits, there is no measurement error is measurement error for the predicting VaR of all the other calculation methods and their own Hit_t and there will be the correct fraction of loss. While for a examination of 50-day evaluation sample, The p value of regressions of both constants and VaR of on Hit_t of all calculation methods are grater than 0.1, all the regression values are statistically insignificant, which means that there is no measurement error for all VaR calculation methods in 50-day evaluation sample. Since there is measurement error for the predicting VaR of HS and its own Hit_t for a 125-day evaluation sample; the fraction of Hypothesis 5 is rejected by the empirical results.

6. CONCLUSION, LIMITATION AND SUGGESTION FOR FUTURE STUDY

6.1Conclusion

This paper evaluates performance of three VaR calculation methods based on 5 models in terms of acceptability, variability, accuracy and measurement error. The performance of each VaR calculation method is different for different length of evaluation sample. It was found that both HS and VCRM are not acceptable in calculating VaR for Chinese stock market for 125-day evaluation sample based on the interval test of Kupiec (1995) and Christofferson (1998) because the LR value of these two methods are greater than given interval value. Only HS is unacceptable in calculating VaR of Chinese Stock market for a 50-day evaluation window. The variability of MCEA ranks the lowest while the accuracy of HS ranks the highest among all the evaluated models for both evaluation windows based on the criteria of RMSRB. Accuracy of HS is the lowest and accuracy of MCEA is the highest for 125-day evaluation window. While for 50-day evaluation window the accuracy of MCEA, MCRM and VCEA is similar and all are very high. For measurement error test, HS has some measurement error in calculating risk of Chinese stock market for 125-day evaluation window but not for 50-day evaluation window. It can be conclude that due to the results of the above four tests, Monte Carlo method modified with EARCH model is the best in measuring VaR of Chinese stock market in long run, but the performance of this method can be improved by both improving asset pricing process and improving the model that capture volatility of stock return. HS with a 500 days estimation window is not suitable to apply in Chinese stock market since it cannot pass all the tests of this paper. Modification should be processed to improve the performance of HS method in Chinese stock market. It is also showed from the results that the forecast ability of RiskMetrics model is higher in short-run then in long run. But optimal decay factor for Chinese Stock market may be different from the one for western financial market.

6.2 Limitation and suggestion for future studies

1.The available data is limited. Due to the short history of HS300, there are not as many data available as other data series in Chinese stock market both for estimation window and evaluation window. The limited number has limited the accuracy of VaR calculated. If the sample period is long enough, accuracy of particular calculation methods will be improved since more historical data can be used to estimate the parameters of the models. While if evaluation window is longer, different length of sample can be compared to evaluate performance of each methods, which can provide more evidence and proof about the performance of different VaR calculation methods relative to different time length.

2.Although the limitation of normal distribution for stock market index is discussed and t distribution was used instead in this paper, still t distribution can not for sure to be the best distribution assumption for financial asset, more discussion and empirical test about distribution assumption is necessary and meaningful, such as using maximum likelihood to regression the parameters of different kinds of distribution assumption like Generalized Error Distribution, Logistic Distribution and so forth. And choose one distribution assumption to replace normal distribution assumption in terms of their significance. Another research direction is to set up the unconditional distribution as a mixture of a normal distribution and another kind of distribution such as a normal-Poisson (Jorion (1988)), a normal-lognormal (Hsieh (1989)) or a Bernoulli-normal distribution (Vlaar & Palm (1993)). A more close to reality distribution assumption will not only improve the accuracy of VaR calculation in Variance-Covariance method, but also help to have a better understanding towards the volatility of financial market and provide help for future research in many other aspects. In risk measurement field, to avoid the shortcoming of an inexact distribution assumption, some other methods such as Expected Shortfall and Press Test are developed from traditional VaR and new techniques to measure financial risk.

3. When using RiskMetrics model provided by J. P Morgan to forecast the volatility of stock market, the decay factor 0.94 is calculated based on the western financial market, it may not accord to other market like Chinese Stock market, if a decay factor calculated based on Chinese Market specifically can be used in this paper, the accuracy of VaR calculation methods based on RiskMetrics may be improved and variability may be reduced. Because decay factor is obtained by the least RMSRB criteria, it makes sense to

study RMSRB and relative approaches. For using GARCH family to capture the volatility clustering of financial market, even EARCH consider about the information level and different reaction of good news and bad news. More advance models like multivariable models can be test to have an exacter describe on financial series or for more than one asset.

4. Due to the development of computer technique, Monte Carlo Simulation is easier than before; it is a very good method in the pricing of financial assets especially in divertive assets like option. The studies and improvement in MC methods will no only have literature meaning but also important in practice, for example, the pricing of financial assets, the setting of marginal level of futures contracts. If the model risk in MC method can be minimized, we can even use VaR based on MC methods to calculate credit risk.

5. Considered about limitation of Historical Simulation, the low accuracy comes from using historical price totally to forecast future; the length of estimation sample is a problem. Meanwhile, if cannot reflect new innovation or big change. To improve the accuracy of HS method, some modified methods like bootstrap and kernel density function can be used.

6. There is no single optimal approach to evaluate the forecast ability of VaR methods; each approach applied in this paper just exam one aspect of ability of VaR from different angle, hence more approach can be discussed to improve the evaluation ability. Meanwhile, for the evaluation methods used in this paper, when there are no exceptions in a given sample period, except for variability test, all the other tests can not be perform to evaluate performances of VaR calculation methods because they use exception-based conditioning variables. More important, because most of these four evaluation methods belong to hypothesis test, their power or in another word their ability to reject the null hypothesis when it is incorrect is an important issue. If a hypothesis test exhibits poor power properties, then the probability of misclassifying an inaccurate VaR model as acceptably accurate will be high. Besides for the evaluation methods that used in this paper to assess the performance of VaR calculation methods, another back-casting method is also commonly used, namely "stress test" but it was not used in this paper.

REFERENCES

Angelidis, T. & Degiannakis, S. A (2006). Backtesting VaR Models: An Expected Shortfall Approach, Available at SSRN: http://ssrn.com/abstract=898473

Bams, D., T. Lehnert & C.P.Wolff (2002). An Evaluation Framework for Alternative VaR Models, EFA Berlin Meetings Discussion Paper.

Chen, X.H. & Yang, H.Y (2003). Value-at-Risk: Measures with Fat Tails and Relative Performance Evaluation, *Management Sciences in China*, 17:1, 39-46

Christoffersen, P (1998). Evaluating Interval Forecasts, *International Economic Review*, 39:841-62.

Christoffersen, P. & D. Pelletier, (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, 2:1,84-108, Oxford University Press.

Don Bredin. & H. Stuart (2004). FOREX Risk: Measurement and Evaluation Using Value-at-Risk, *Journal of Business Finance & Accounting,* 31: 1389–1417.

Ekrem, K (2006). Violation duration as a better way of VaR model evaluation: evidence from Turkish market portfolio, Finecus Financial Software and Consultancy, MPRA Paper.

Engel, J. &M, Gizycki (1999). Conservatism, Accuracy & Efficiency: Comparing Value-at-Risk Models, Working Paper 2, And Australian Prudential Regulation Authority.

Engle, R. & S, Manganelli (2002). CAViaR: Conditional Value At Risk By Regression Quartiles, NBER Working Paper.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50,987-1008.

Fama, E. F. (1965). The Behavior of Stock Market Prices, *Journal of Business*, 38:1,34-105.

Gloria, G. R.,T. H. Lee & Emre, Y (2007). Optimality of the RiskMetrics VaR Model, *Finance Research Letters,* 4:3, 137-145.

Hua, J.Z. & Wu, C.F. (2005). Research on The Variability Of VaR Calculation Models And Accuracy Of VaR Estimation, Aetna management school of Shanghai Jiaotong University.

Hendricks, D (1996). Evaluation of Value-at-Risk Models Using Historical Data [J], *Economic Policy review*, Federal Bank of New York, April 39-69.

Hsieh, D (1989). Modeling Heteroskedasticity in Daily Foreign Exchange Rates, J*ournal of Business and Economic Statistics*, 7, 307–317.

J. P. Morgan, (1996), RiskMetrics-Technical Document, Fourth Edition.

Jorion, P (1988). On Jump Processes in the Foreign Exchange and Stock Markets, *Review of Financial Studies*, 68, 165–176.

Jorion, P (1996). Risk: Measuring the risk in Value at Risk, *Financial Analysts Journal* 52:11, 47-56

Jorion, P (2000). Value at Risk: The New Benchmark for Controlling Market Risk, McGraw-Hill, New York.

Jorion, P (2003). Financial risk manager handbook, second edition, John Wiley & Sons, Inc.

Kanwer, A. & A. Zaidi (2006). Evaluation of Alternative Value at Risk model in Pakistan: Analyzing the risks faced by the capital market participants in Pakistan, Working Paper.

Keith, K. M. Stefan, &M. S. Paolella (2005). Value-at-Risk Prediction: A Comparison of Alternative Strategies, *Journal of Financial Econometrics*, Available from Internet: 10.1093/jjfinec/nbj002.

Kupiec (1995). Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivatives* 3: 73-84.

Lamantia, F., S.Ortobelli & S. T. Rachev (2006). An Empirical Comparison among VaR Models and Time Rules with Elliptical and Stable Distributed Returns, *Investment Management and Financial Innovations* 3:8-29.

Laurentiu & Mihailescu (2004). A Sequential Method for the Evaluation of the VaR Model Based on the Run Between Exceedances, *Allgemeines Statistisches Archiv,* 51-72.

Li, J. X. & Guo, D. Z (2003). The feedback testing o f VaR model, *Journal of Guizhou College of Finance and Economics* 7: 17-19.

Lin, C.H., Chang, C.C.C & C.H. Ching (2005). A Comparison of Three Revised Historical Simulation Methods for Estimating VaR, *Journal of risk measurement* 7:2, 183-201.

Lintner, J (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 47, 13-37.

Liu, M. Y., H. F. Lee & C. Y. Wu (2005). Evaluation of Power EWMA Models for Bank Portfolio – Conservativeness, Accuracy and Efficiency, Annual Meeting of European Financial Management Association, June 29-July 2, 2005 Milan, Italy.

Liu, X. S. & Z. L. Zheng (2007). Testing the Forecasting Ability of the VaR Model of the Commercial Banks, *Contemporary Finance & Economics,* 8:39-43.

Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates, *Economic Review*, Federal Reserve Bank of San Francisco, 3-17.

Markowitz, M (1952). Portfolio selection, *Journal of Finance*, 7:1,78-91.

Ogryczak, W. &A. Ruszczynski (1997). From Stochastic Dominance to Mean-Risk Models: Semi-deviations as Risk Measures, *Interim Report IR-97-027*, IIASA, Laxenbury.

Sarma, M., S. Thomas & A. Shah (2000) Performance Evaluation of Alternative VaR Models", mimeo, *Indira Gandhi Institute of Development Research.*

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance*, 19:3, 425-442.

Smith, D. R. (2007). Conditional Backtesting of Value-at-Risk Models, Simon Fraser University, Working Paper.

Victor H. D.L.Pen, R. Rivera & R.M. Jesus (2006). Quality Control of Risk Measures: Backtesting VaR Models, Journal of Risk, 9:2, 39-54.

Vlaar, P. & F. Palm (1993) The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps, *Journal of Business and Economic Statistics*, 11, 351–360.

Zhang, L. & Z. L Zheng (2007). Indices Portfolio VaR Model And Model Tests, *Journal of Shanxi Finance and Economics University.*