

Approaches to VaR*

Hao Li Xiao Fan Yu Li Yue Zhou Ze Jin Zhao Liu

Stanford University

Abstract

Referring to related documents and papers, we implement several different approaches to compute the VaR of a delta-hedged portfolio constructed by 41 stocks and corresponding options. First we interpreted the concepts and techniques involved with our study. Then we discussed the details about both Historical Simulation and Monte Carlo Simulation, and pointed out their shortcomings through experiments. Resorting to more sophisticated theory, we applied both Unfiltered and Filtered Historical Simulation with different distribution assumptions (normal, t, GPD), sampling methods (bootstrap, empirical, MLE) and model options (GARCH, GJR, EGARCH). Comparing each method based on the results of backtesting, we drew various conclusions on the surface and analyzed their fundamental reasons for deeper understanding. Further we describe some future work to complement our study.

Key words: Delta Hedge; Historical Simulation; Monte Carlo Simulation; Filtered Historical Simulation; Extreme Value Theory

1 Introduction

These years have been characterized by significant instabilities in financial market worldwide. This has led to increasingly more concern about risk management. In financial mathematics and financial risk management, Value at Risk (VaR) is a widely used risk measure of the risk of loss on a specific portfolio of financial assets. For a given portfolio, probability and time horizon, VaR is defined as a threshold value such that the probability that the market-to-market loss on the portfolio over the given time horizon exceeds this value (assuming normal markets and no trading in the portfolio) is the given probability level.

VaR can be used to describe the event risk which is present in all areas of risk management such as market, credit, operational or insurance risk. There are an increasing

*MS&E 444 Investment Practice Project, Spring 2011-2102. Instructor: Kay Giesecke. CA: Gerald Teng.

number of approaches to estimate VaR. Several scholars have presented their classification of the methods. According to Perignon and Smith (2006) survey, 73% of banks among 60 US, Canadian and large international banks over 1996-2005 have reported that their VaR methodology used was historical simulation. The Monte Carlo (MC) simulation was the second most popular method.

Miura and Oue (2001) introduce a useful taxonomy of statistical methodologies for the VaR measurement along dual dimensions of distributional assumptions and dependence assumptions. It can be explained in a 3 by 2 matrix with normal, non-normal and non-parametric along one dimension and i.i.d. and time dependence along the other dimension. See Table 1 below for the classification under their work.

Table 1: VaR Classification I

	Normal	Non-Normal	Non-Parametric
i.i.d.	Variance-Covariance (VC) & Equally Weighted Moving Average (EQMA)		Historical Simulation
Time Dependence	Exponentially Weighted Moving Average (EXMA) & GARCH	Exponentially Weighted Moving Average (EXMA) & GARCH	

Manganelli and Engle (2001) conclude that three broad categories of VaR estimation approaches: Parametric, Nonparametric and Semiparametric. RiskMetrics and GARCH which can be used under both normal and non-normal assumption are parametric approaches. Historical Simulation is nonparametric method that uses the empirical distribution of past returns to generate a VaR. Semiparametric approaches are Extreme Value Theory, CAViaR and quasi-maximum likelihood. See Table 2 below for the classification under their work.

Variance-covariance is the simplest VaR method, sometimes called the delta-normal method. It assumes the portfolio exposures are linear and that the risk factors are jointly normally distributed. Since the portfolio return is a linear combination of normal variables, it is normally distributed itself. Thus, the portfolio volatility can be calculated by using covariance matrix and weight vector easily. The main benefit of variance-covariance method

Table 2: VaR Classification II

Parametric Normal & Non-Normal	Nonparametric	Semiparametric
RiskMetrics GARCH	Historical Simulation	Extreme Value Theory, CAVia quasi-maximum likelihood

is the simplicity. It provides a closed-form solution. However, its linearity assumption is also its drawback. It cannot measure the nonlinear effects embedded in options. Moreover, its normal distribution assumption also may underestimate the extreme outcomes.

Historical Simulation is the procedure for predicting value at risk by “simulating” or constructing the cumulative distribution function of asset returns over time. It does not require any statistical assumption beyond stationary of the distribution of returns or, in particular, their volatility. The limitation of the historical simulation lies in its i.i.d. assumption of returns. From empirical evidence, it is known that asset returns are clearly not independent as they exhibit certain patterns such as volatility clustering. Unfortunately Historical Simulation does not take into account such patterns.

Monte Carlo method was coined in the 1940s by John Von Neumann, Stanislaw Ulam and Nicholas Metropolis. Monte Carlo simulation uses random samples from known populations of simulated data to track a statistic’s behavior. With Monte Carlo VaR measures, an inference procedure typically characterizes the distribution of returns by assuming some standard joint distribution-such as the joint-normal distribution-and specifying a covariance matrix and mean vector.

Filtered historical simulation is a generalized historical simulation proposed by Barone-Adesi et al. (1999). It incorporates conditional volatility models like GARCH into the historical simulation model to overcome the shortcomings of traditional historical simulation. The results of FHS are sensitive to changing market conditions and can predict losses outside the historical range. From a computational standpoint, this method is very reasonable for large portfolios and empirical evidence supports its predictive ability.

Extreme Value Theory (EVT) is a theory dealing with the extreme deviations from the median of probability distributions. Extreme values are crucial for risk management because they are associated with catastrophic events such as market crash and extreme large losses. However, the occurrence of extreme events is rare. Unlike some research which assume a

certain distribution which may be not identical to the real distribution and give rise to error, EVT do not assume a specific distribution, but deal with extreme value specifically, which can describe the tail area of the distribution more exactly.

The Basel Accords refer to the banking supervision Accords-Basel I, Basel II and Basel III-issued by the Basel Committee on Banking Supervision (BCBS). The purpose of the accords is to ensure that financial institutions have enough capital on account to meet obligations and absorb unexpected losses. Our model is in compliance with Basel Accord. The Internal Models Approach (IMA) of Basel Accord specifies market risk charge is computed according to the set of rules

- **A horizon of 10 trading days**
- **A 99% confidence interval**
- **An observation period based on at least one year of historical data**

The market risk charge in day t is calculated by the following formula

$$MRC_t^{IMA} = \max \left(k \frac{1}{60} \sum_{i=1}^{60} VaR_{t-i}, VaR_{t-1} \right) + SRC_t \quad (1.1)$$

where SRC represents the specific risk charge, which is a buffer against idiosyncratic factors, including basis risk and event risk.

One of the important aspects of the regulation is backtesting. Backtesting is the process of comparing losses predicted by the VaR model to those actually experienced over the testing period. The Basel Committee set up a framework based on the daily backtesting of VaR, see Table 3. Exceptions below a threshold of four is acceptable, which defines a green zone. If the number of exceptions is five or more, the bank falls into a yellow or red zone. The market capital charge is expressed as a multiplier of the 10-day VaR at the 99% level of confidence. The multiplier k is subject to a floor of 3. After a break into the yellow zone, the multiplicative factor is increased to 4, or by a plus factor described in the diagram below. Besides, our work can be applied to meet the regulatory requirements with only minor changes.

The traditional way of backtesting ignores the time pattern of losses. For an ideal and robust model, exceptions should occur uniformly over the backtesting period. Otherwise, for instance, despite the fact that only four exceptions occurred over the past year, which passes the Basel requirements, these exceptions occurred in the same month should cause extra attention, because it is more likely that the portfolio will suffer large losses in just a

Table 3: The Basel Penalty Zones

Zone	Number of Exceptions	Potential Increase in k
Green	0 to 4	0.00
Yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red	≥ 10	1.00

couple of days. In contrast, a pattern where exceptions tend to cluster indicates a weakness of the model. In response, our backtesting model takes time variation of losses into account

$$\begin{aligned}
 MRC_t^{IMA} = & \max \left(k \frac{1}{60} \sum_{i=1}^{60} VaR_{t-i}, VaR_{t-1} \right) \\
 & + \max \left(k_S \frac{1}{60} \sum_{i=1}^{60} SVaR_{t-i}, SVaR_{t-1} \right) + SRC_t \quad (1.2)
 \end{aligned}$$

The Basel Committee has added a Stressed VaR (SVaR) to the capital requirements against market risks in 2009 after the financial crisis. SVaR is the loss on the current portfolio that corresponds to a 99% confidence level over a 10-day period, calibrated over a continuous 12-month period such as 2007 & 2008. This period is covered by our analysis in the following sections.

As the title of our report tells, we implemented several different approaches to compute the VaR of a delta-hedged portfolio constructed by 41 stocks and corresponding options. In Section 2, we introduced the concepts and techniques involved with our study. In Section 3 we discussed the details about Historical Simulation and Monte Carlo Simulation. In Section 4, we applied both Unfiltered and Filtered Historical Simulation with different distribution assumptions, sampling options and method options to calculate the VaR, such as AR-GARCH model and Generalized Pareto Distribution. Further we conducted backtesting on each method, and compared the results for deeper understanding of these models in Section 5. The conclusions based on backtesting and the analysis of underlying reasons are listed in Section 6. Also, there is some future work we would like to describe in Section 7.

2 Concepts and Techniques

2.1 Historical Simulation

Historical Simulation is a good resampling method because of its simplicity and lack of distributional assumption about underlying process of returns. It is based on the assumption that history is repeating itself.

Suppose we observe data from day 1 to day t , and r_t is the return of portfolio on day t , then we get a series of return $\{r_{t+1-\tau}\}_{\tau=1}^m$. The value of risk with coverage rate p is calculated as the $(100 \cdot p)\%$ of the sequence of past portfolio returns

$$\widehat{VaR}_{t+1}^p = \text{percentile}\{\{r_{t+1-\tau}\}_{\tau=1}^m, (100 \cdot p)\%\} \quad (2.1)$$

Unlike other parametric methods, the historical simulation makes no specific distribution assumption about return distributions. However, the historical simulation implicitly assumes that the distribution of past returns is a good and complete representation of expected future returns. This method also relies on the specified short historical moving window.

Also, bootstrap historical simulation approach is an extension of traditional historical simulation. It is a simple and intuitive estimation procedure. The bootstrap technique draws a sample from the data set, records the VaR from that particular sample and “returns” the data. This procedure is repeated over and over and records multiple sample VaRs. Since the data is always returned to the data set, this procedure is like sampling with replacement. The best VaR estimate from the full data set is the average of all sample VaRs.

2.2 Monte Carlo Simulation

Monte Carlo simulation method is similar to historical simulation method, except that the movements in risk factors are generated from estimated distribution

$$\Delta f^k \sim g(\theta), \quad k = 1, \dots, K \quad (2.2)$$

where g is the joint distribution and θ its parameters. When we need to specify the dependency structure of the risk factors, we can determine the marginal distributions as well as their copula.

Monte Carlo simulation requires users to make assumptions about the stochastic process. It is subject to model risk. It also creates inherent sampling variability because of the randomization. Different random numbers will lead to different results. It may take a large

number of iterations to converge to a stable VaR measure. Unlike other methods, Monte Carlo simulation makes explicit the sampling variability in the risk numbers. Although, Monte Carlo Simulation can be time-consuming according to the properties of problem, the main benefit of running it is that it can model instruments with non-linear and path-dependent payoff functions, especially complex derivatives.

2.3 Filtered Historical Simulation

The filtered simulation is the most complicated and comprehensive example of the non-parametric approaches. The process combines the traditional simulation model with conditional volatility models (like GARCH or EGARCH), which makes it attractive in dealing with changing volatility. The filtered simulation model is flexible to capture conditional volatility, volatility clustering and factors that can have an asymmetric effect on volatility. These models are very reasonable for large portfolios and have good predictive ability for changing market conditions.

2.3.1 ARMA-GARCH Model

Accounting for the two most significant characteristic of financial asset returns, namely strong time-varying volatility and excess kurtosis relative to the normal distribution, we fit an ARMA time series model to our data and then use the parameters of this model for VaR prediction and GARCH time series for estimating the time-varying variance of our model. In general, the ARMA(p, q)-GARCH(r, s) model is given by

$$r_t = a_0 + \sum_{i=1}^p a_i r_{t-i} + \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}, \quad \varepsilon_t = z_t \sigma_t \quad (2.3)$$

$$\sigma_t^2 = c_0 + \sum_{i=1}^r c_i \varepsilon_{t-i}^2 + \sum_{j=1}^s d_j \sigma_{t-j}^2 \quad (2.4)$$

where the equation (2.3) is ARMA(p, q) modeling of returns r_t with AR(p) terms $\{a_i\}_{i=1}^p$ and MA(q) terms $\{b_j\}_{j=1}^q$, and equation (2.4) is GARCH(r, s) modeling of random residuals ε_t (noises), which defines volatility of the random residuals ε_t as a function of past residuals $\{\varepsilon_k\}_{k=1}^{t-1}$ and volatilities $\{\sigma_k^2\}_{k=1}^{t-1}$. Random residuals are assumed to be unpredictable and conditionally heteroskedastic as follows

$$\mathbf{E}[\varepsilon_t | F_{t-1}] = 0 \quad (2.5)$$

$$\text{Var}[\varepsilon_t | F_{t-1}] = \sigma_t^2 \quad (2.6)$$

then the standardized residual returns $z_t = \frac{\varepsilon_t}{\sigma_t}$ are i.i.d with mean 0 and variance 1.

We use the model with the most common formulation with $p = 1$, $q = 0$ and $r = s = 1$, it has the advantage of being very simple to estimate, and its theoretical properties of interest, such as moments and stationarity conditions, are tractable. The normal assumption of z_t is found to be inadequate for sample fitting and forecasting not long after its inception, so t-distribution is adopted with its additional shape parameter and performs better than a normal model, particularly for more extreme (1% or less) VaR thresholds.

Evidence shows that returns are negatively correlated with changes in returns volatility, that is, volatility tends to rise in response to excess returns lower than expected and to fall in response to excess returns higher than expected. However, GARCH models are symmetric in the sense that they only consider the magnitude and not the positivity or negativity of unanticipated excess returns.

This suggests that models in which σ_t^2 responds asymmetrically to positive and negative residuals might be preferable and many nonlinear extensions of GARCH model such as EGARCH and GJR-GARCH models have been proposed.

2.3.2 EGARCH Model

To accommodate for the asymmetric relation between returns and volatility changes, Nelson proposes the asymmetric EGARCH model which includes an adjusting function $g(z)$ in the conditional variance equation, it is expressed by

$$\ln \sigma_t^2 = \alpha_0 + \ln \sum_{i=1}^q \alpha_i g(z_{t-i}) + \sum_{j=1}^p \beta_j \ln(\sigma_{t-j}^2) \quad (2.7)$$

where $z_t = \frac{\varepsilon_t}{\sigma_t}$ is the normalized residual series. The value of $g(z_t)$ is a function of both the magnitude and sign of z_t

$$g(z_t) = \theta_1 z_t + \theta_2 [|z_t| - \mathbf{E}|Z_t|] \quad (2.8)$$

where $\mathbf{E}|z_t|$ depends on the assumption of the unconditional density. Compared to the standard GARCH model, the EGARCH model allows positive and negative shocks to have a different impact on volatility and also allows big shocks to have a greater impact on volatility.

2.3.3 GJR-GARCH Model

The GJR-GARCH model by Glosten, Jagannathan and Runkle (1993) also models asymmetry in the ARCH process. Its generalized form is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q (\alpha_i \varepsilon_{t-1}^2 + \omega_i S_{t-i}^- \varepsilon_{t-1}^2) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (2.9)$$

where $S_t^- = I_{\{\varepsilon_t < 0\}}$. In this model, it is assumed that the impact of ε_t^2 on the conditional variance σ_t^2 is different when ε_t is positive or negative.

We compare the performance of the GARCH, EGARCH and GJR-GARCH models and also introduce different density assumptions (Normal, Student t).

2.4 Extreme Value Theory

This section introduces how to use the extreme value theory to calculate VaR. Instead of placing emphasis on the whole distribution, Extreme Value Theory (EVT) pays attention just to the tail area of the distribution and uses GPD (Generalized Pareto Distribution) to describe the tail area. As it focus on the extreme value, this method can describe the tail area of the distribution more exactly. The GPD is a two-parameter distribution with distribution function

$$G_{\xi, \beta} = \begin{cases} 1 - (1 + \frac{\xi x}{\beta})^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp(-\frac{x}{\beta}), & \xi = 0 \end{cases} \quad (2.10)$$

where $\beta > 0$, and where $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

The distribution of excesses losses over a high threshold u is defined to be

$$F_u(y) = P\{X - u \leq y | X > u\}, \quad \text{for } 0 \leq y \leq x_0 - u \quad (2.11)$$

where $x_0 \leq \infty$ is the right endpoint of F .

There is a very important theorem in EVT: for a large class of underlying distributions we can find a function β_u such that $\lim_{u \rightarrow x_0} \sup_{0 \leq y < x_0 - u} |F_u(y) - G_{\xi, \beta(u)}(y)| = 0$. That is to say, GPD is the natural model for the unknown excess distribution above sufficiently high threshold. Setting appropriate u and using maximum likelihood estimation, we can obtain the estimator of the two parameters.

Using the outcome of the estimation of the distribution of excesses losses, we can estimate the tails of the distribution

$$F(x) = (1 - F(u))G_{\xi, \beta}(x - u) + F(u), \quad \text{for } x > u \quad (2.12)$$

We can utilize the empirical estimator $(n - N_u)/n$ to estimate $F(u)$. Thus, the cumulative distribution function for the tail area of the distribution ($x > u$) is

$$\hat{F}(x) = 1 - \frac{1}{N_u} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}} \quad (2.13)$$

Therefore, for a given probability $q > F(u)$, we can estimate VaR

$$\widehat{VaR}_{t+1}^q = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - q) \right)^{-\xi} - 1 \right) \quad (2.14)$$

What has been mentioned above is the theory background for static EVT. However, in the practical market, we need to take into account volatility of market instruments. It is more likely to get large value in a period of high volatility. We use Stochastic Volatility (SV) model to describe this phenomenon

$$X_t = \mu_t + \sigma_t Z_t \quad (2.15)$$

The randomness in the model comes through the random variables Z_t , which are the noise variables or the innovations of the process. Expected value and the volatility can be obtained through GARCH model and we assume i.i.d. variables Z_t .

In dynamic risk management we are interested in the conditional return distribution: $F_{X_{t+1} + \dots + X_{t+k} | \mathcal{F}_t}(x)$, in contrast with the static EVT which calculates the unconditional distribution. According to SV model

$$VaR_q^t = \mu_{t+1} + \sigma_{t+1} VaR(Z)_q \quad (2.16)$$

here the $VaR(X)$ is not calculated directly but through the calculation of $VaR(Z)$. And $VaR(Z)$ can be obtained through the static EVT method discussed before.

2.5 Copula

Most VaR methods assume that the risk factors have a multivariate normal distribution. Some scholars have used correlation to define the dependence structure between different risk factors. Embrechts, McNeil and Straumann (1999) demonstrated that the concept of correlation causes several pitfalls in risk management. Jörn and Thomas (2002) have introduced copula techniques into computation of VaR to quantify the dependency.

A copula is a function which relates a multivariate distributional function to a lower dimensional marginal distributional function, generally a one-dimensional function. The concept of copulas was introduced by Sklar in 1959. Copulas can be used to describe the dependence between two or more random variables with arbitrary marginal distributions.

Referring the notations to Nelsen (1999), a copula is a function $C : [0, 1]^n \rightarrow [0, 1]$ with special properties. The joint multidimensional cumulative distribution is

$$\begin{aligned}
& P(X_1 \leq x_1, \dots, X_n \leq x_n) \\
&= C(P(X_1 \leq x_1) \cdots P(X_n \leq x_n)) \\
&= C(F_1(x_1), \dots, F_n(x_n))
\end{aligned} \tag{2.17}$$

where F_1, \dots, F_n denote the cumulative functions of n random variables X_1, \dots, X_n .

Sklar's Theorem Let H be a joint distribution function with margins F_1 and F_2 . Then there exists a copula C with

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \tag{2.18}$$

for every $x_1, x_2 \in R$. If F_1 and F_2 are continuous, then C is unique. Otherwise, C is uniquely determined on $\text{Range } F_1 \times F_2$. On the other hand, if C is a copula and F_1 and F_2 are distribution functions, then the function H defined by the equation above is a joint distribution function with margins F_1 and F_2 .

There are common copulas like Multivariate Normal Copula (MVNC), Multivariate Student's Copula (MVTC), Archimedean Copula, Extreme Value Copula (EVC) etc. Compared with multivariate normal copula, Student's t copula function is also symmetric, which can only capture the symmetrical information in the financial market. While, Student's t copula has fatter tails and is more sensitive to the variation in tails. Thus, we used t-copula in our method to better monitor the variables' dependency structure in tails. Multivariate Student's t copula, MVTC, distribution function and density function are

$$\begin{aligned}
& C(u_1, u_2, \dots, u_N; \rho, \nu) \\
&= T_{\rho, \nu}(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2), \dots, T_\nu^{-1}(u_N)) \\
&= \int_{-\infty}^{T_\nu^{-1}(u_1)} \int_{-\infty}^{T_\nu^{-1}(u_2)} \cdots \int_{-\infty}^{T_\nu^{-1}(u_N)} \frac{\Gamma(\frac{\nu+N}{2}) |\rho|^{-\frac{1}{2}}}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{N}{2}}} \left(1 + \frac{1}{\nu} x' \rho^{-1} x\right)^{-\frac{\nu+N}{2}} dx_1 dx_2 \cdots dx_N \\
& \quad c(u_1, u_2, \dots, u_N; \rho, \nu) \\
&= |\rho|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+N}{2}) [\Gamma(\frac{\nu}{2})]^{N-1}}{[\Gamma(\frac{\nu+1}{2})]^N} \frac{(1 + \frac{1}{\nu} \zeta' \rho^{-1} \zeta)^{-\frac{\nu+N}{2}}}{\prod_{n=1}^N \left(1 + \frac{\zeta_n^2}{\nu}\right)^{-\frac{\nu+1}{2}}}
\end{aligned} \tag{2.19}$$

2.6 Backtesting

As is mentioned above, the backtesting methods are used to assess the accuracy and performance of a VaR model. The design of VaR models includes the use of confidence levels. We expect to have a frequency of exceptions that corresponds to the confidence level. For example, if we use a 99% confidence interval, we expect to find exceptions in 1% of the instances. By determining a range of the number of exceptions that we would accept, we must strike a balance between rejecting an accurate model (Type I error) and accepting an inaccurate model (Type II error). We define the “exception sequence” of VaR exceptions as $I_t = I_{X_t \leq VaR_t}$. For a perfect VaR model, the exception sequence should be independently distributed over time as a Bernoulli variable.

2.6.1 Unconditional Coverage Testing (UCT)

Kupiec (1995) determined a measure to accept or reject models using the tail points of a log-likelihood ratio. Let π be the fraction of exceptions for a model, the likelihood function of the Bernoulli sequence is

$$L(\pi) = \prod_{i=1}^T (1 - \pi)^{1 - I_t} \pi^{I_t} = (1 - \pi)^{T_0} \pi^{T_1} \quad (2.20)$$

where T_0 and T_1 are the number of 0s and 1s in the sample. The MLE of π is $\hat{\pi} = T_1/T$ and $L(\hat{\pi}) = (T_0/T)^{T_0} (T_1/T)^{T_1}$.

Let p be the known VaR coverage rate under the unconditional coverage null hypothesis that $\pi = p$. We have the log-likelihood ratio test statistic $LR_{uc} = -2 \log[L(p)/L(\hat{\pi})] = -2 \ln[(1 - p)^{T - T_1} p^{T_1}] + 2 \ln\{[1 - (T_1/T)]^{T - T_1} (T_1/T)^{T_1}\}$, which has asymptotical χ_1^2 distribution.

The p-value is $1 - F_{\chi_1^2}(LR_{uc})$ and we reject the null hypothesis if the p-value is below the desired significance level.

2.6.2 Independence Testing (IND)

In addition to having a predictable number of exceptions, the exceptions should also be fairly equally distributed across time. A bunching of exceptions may indicate that market correlations have changed or that the trading positions have been altered.

The independence testing is to detect the clustering of VaR exceptions which usually happen in the historical simulation method. To test the independence, we use the first order

Markov model with transition probability matrix

$$\Pi_1 = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix} \quad (2.21)$$

with π_{ij} the transition probability $\pi_{ij} = P[I_t = i \text{ and } I_{t+1} = j]$.

The likelihood function is $L(\Pi_1) = \pi_{00}^{T_{00}} \pi_{01}^{T_{01}} \pi_{10}^{T_{10}} \pi_{11}^{T_{11}}$ where T_{ij} is the number of observations with a j following an i . The MLE of Π_1 is

$$\hat{\Pi}_1 = \begin{pmatrix} T_{00}/T_0 & T_{01}/T_0 \\ T_{10}/T_1 & T_{11}/T_1 \end{pmatrix} \quad (2.22)$$

Under the independence hypothesis, we have $\pi_{01} = \pi_{11} = \pi$, and the MLE of transition matrix is

$$\hat{\Pi} = \begin{pmatrix} 1 - \hat{\pi} & \hat{\pi} \\ 1 - \hat{\pi} & \hat{\pi} \end{pmatrix} \quad (2.23)$$

with $\hat{\pi} = T_{11}/T$. And the log-likelihood ratio test statistic is $LR_{ind} = -2 \log[L(\hat{\pi})/L(\hat{\Pi}_1)] \sim \chi_1^2$, where $L(\hat{\pi})$ is the likelihood under the null hypothesis.

2.6.3 Conditional Coverage Testing (CCT)

Christofferson proposed extending the unconditional overage test statistic to allow for potential time variance of the data and developed the overall log-likelihood test statistic for conditional coverage.

The conditional coverage test is used for testing simultaneously if the average number of exceptions is right and the VaR exceptions are independent. The null hypothesis is $\pi_{01} = \pi_{11} = p$ and we use the test statistic $LR_{cc} = LR_{uc} + LR_{ind} = -2 \log[L(p)/L(\hat{\Pi}_1)] \sim \chi_2^2$.

In our paper, we will apply all of these three back-testing approaches to compare the accuracy, independence and the joint performance of each VaR estimation methods.

2.7 Delta Hedge

The delta (Δ) of an option is the rate of change of the option price with regards to change of the price of the underlying asset. In general, $\Delta = \partial C / \partial S$, where C is the price of the call option and S is the stock price. A position with a delta of zero is referred to as being delta neutral. Delta hedging is an example of dynamic hedging. In contrast, static hedging is referred to as the hedge where the position is never adjusted after setting up. To create a perfect hedge, the portfolio must be adjusted continuously, because the delta

changes when the price of the underlying changes and when time passes. Therefore, dynamic hedging has to be adjusted periodically, which is known as rebalancing. In our model, we adjust the positions at the end of each trading day to maintain delta neutral such that the portfolio consists of delta shares of the underlying asset and a short call option. Specifically, our resulting portfolio is a combination of 41 such stock-and-option portfolios with equal weight.

2.7.1 Continuous-Time Delta Hedges

We first consider the continuous-time hedging. Consider a portfolio that is short one call option and long δ shares of the underlying asset and that has a cash position equal to $C - \delta S$. The change in the value of the portfolio in a short period dt is

$$-dC + \delta dS + q\delta S dt + (C - \delta S)r dt \quad (2.24)$$

The change in the value of the portfolio results from four factors, change in the option price, capital gain or loss on δ shares of stocks, dividend received and interest expense on the cash position. Applying Ito's formula, we can easily show

$$-\Theta dt - \frac{1}{2}\Gamma\sigma^2 S^2 dt + q\delta S dt + (C - \delta S)r dt \quad (2.25)$$

There are some interesting points noticeable. No dS term is found in (2.25) since the delta hedging fully eliminates the change of stock price. Θ captures the time decay. The portfolio is short gamma, the second derivative capturing the convex shape of option price. Together with the volatility σ , they have a profound impact on the evolution of the portfolio. In practice, any hedge will therefore be imperfect as the rebalancing frequency varies.

2.7.2 Discretely-Rebalanced Delta Hedges

To do discretely-rebalanced Delta hedges, we either simulate the changes in S over time or use the historical prices, and then sum the gains and losses over discrete rebalancing periods. The dynamics are as follows. We begin with the portfolio that is short a call, long δ shares of the underlying, and short $\delta S - C$ in cash. After the stock price changes, say from S to S' , we compute the new delta δ' . The cash flow from adjusting the hedge is $(\delta - \delta')S'$. Accumulation (or payment) of interest on the cash position is captured by the factor $e^{r\Delta t}$. Although continuous payment of dividends can be modelled similarly, we ignore dividends payment here to simplify the model without much loss of accuracy. The cash

position is adjusted due to interest, and the cash flow from adjusting the hedge. At date T , the value of the portfolio is the cash position less the intrinsic value of the option.

2.7.3 Portfolio Construction

We start from a single stock and option pair and then combine the 41 pairs into a portfolio, see Table 4 for stock tickers. Stocks are selected by the following criteria. First, they must have been actively traded in U.S. equity markets over the time periods 2005-2012. Second, the three-month call options implied volatility must be available throughout the periods. Third, portfolio must be well-diversified containing a variety of stocks across a range of industries. Bloomberg terminal is our major data source.

Table 4: Tickers

AA	AAPL	ALU	AMAT	ATVI	BAC
BPOP	C	CHK	CMCSA	CSCO	DELL
EMC	F	FCX	FTR	GE	HBAN
HPQ	INTC	JPM	MMM	MRVL	MSFT
MU	NOK	NTAP	NVDA	ORCL	PBR
PFE	PHM	QCOM	RF	RIMM	SBUX
SVU	T	WFC	WMT	YHOO	

2.7.4 Profit and Loss Analysis

We did a full valuation on our option positions using Black-Scholes option pricing model and the discrete model mentioned above using the following model inputs

- S_0, \dots, S_9 = stock prices in the ten-day period (historical or simulated)
- K = strike price of option, we choose at-the-money option on each of the underlying stock so the strike price is S_0 , stock price at the beginning of the period
- r = risk-free rate, which is three-month treasury bill rate
- σ = volatility, which is the implied volatility of the 3-month option traded in the market
- T = time to maturity, which is three months

In the individual stock level, we assess the profit & loss by dividing the absolute profit & loss by $\delta_0 \cdot S_0$, which is the initial position in the stock. Since the initial value of the portfolio is zero, the classical measurement of return is not applicable in this case. The reasoning behind such treatment is that we want to measure the risk per unit of stock position. This measurement is beneficial because the return can be intuitively comparable to the absolute return of the stock itself over the same period. See Figure 1 for our portfolio's return over a period of five years from 2007 to 2012. It is apparent to us that the hedged portfolio consistently drifts around zero except for the periods during the financial crisis in 2008 & 2009 and European debt crisis in mid 2011. During economic downturns, the return changes drastically reflecting dependency on the changing market volatility.

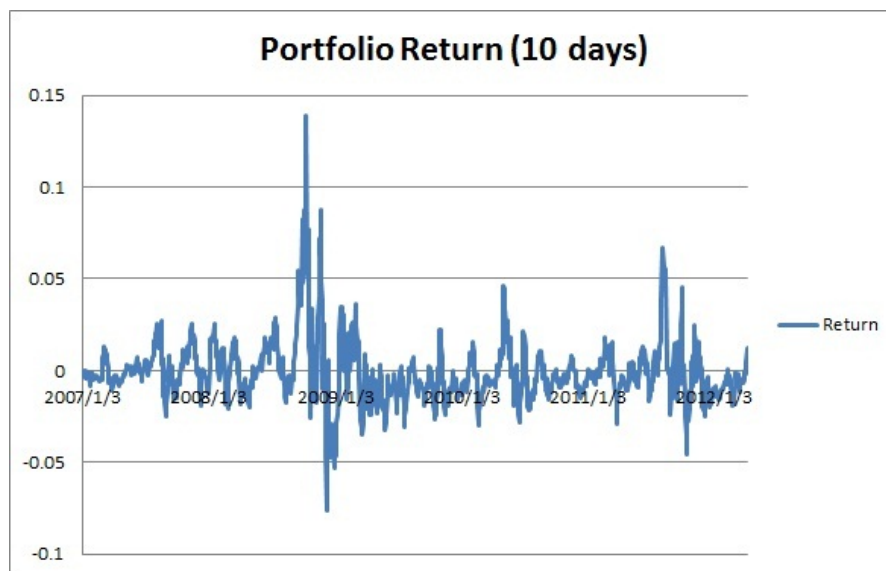


Figure 1: Portfolio Returns (10 days)

2.7.5 Implications for the Volatility Premium

The fact that the profit and loss of such strategy depends on the market volatility give rise to our interests in find a relationship between these two. In addition to implied volatility, we introduce the concept of realized volatility, sometimes referred to as the historical volatility. It measures what actually happened in the past while the implied volatility refers to the market's assessment of future volatility. It is an observable phenomenon that, the implied volatility is generally higher than the typical volatility that the asset eventually displays. We believe the gap or premium between implied volatility and realized volatility is that sellers of naked options bear an unlimited risk and therefore receive a premium from the

buyers.

We define the volatility premium as Implied Volatility - Realized Volatility in percentage terms. We can compute the realized volatility using the below equation

$$\sigma_{realized} = 100 \times \sqrt{\frac{252}{n} \sum_{t=1}^n R_t^2} \quad (2.26)$$

Where

$\sigma_{realized}$ = Realized Volatility,

n = number of trading days in the period,

R_t = continuously compounded daily return

If you short an option and dynamically delta hedge (you are actually short volatility), the actual volatility needs to be less than the implied volatility in order to make money. The drop in the difference of implied volatility and realized volatility corresponds to the drop in profits or rise in losses, see Figure 2.

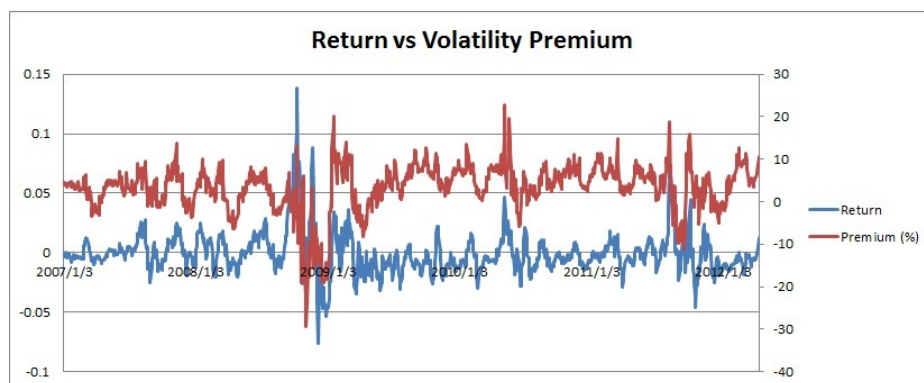


Figure 2: Returns vs Volatility Premium

2.8 Q-Q plot

A natural question in the analysis of VaR is, “From what distribution the data is drawn?” The truth is that you will never really know since you only observe the realizations from random draws of an unknown distribution. However, visual inspection can be a very simple but powerful technique.

In particular, the quantile-quantile (Q-Q) plot is a straightforward way to visually examine if empirical data fits the reference or hypothesized theoretical distribution. The process graphs the quantiles at regular confidence intervals for the empirical distribution against the theoretical distribution. As an example, if both the empirical and theoretical data are

drawn from the same distribution, then the median (confidence level = 50%) of the empirical distribution would plot very close to zero, while the median of the theoretical distribution would plot exactly at zero.

Continuing in this fashion for other quantiles (40%, 60%, and so on) will map out a function. If the two distributions are very similar, the resulting Q-Q plot will be linear.

let us compare a theoretical standard normal distribution relative to an empirical t-distribution (assume that the degree of the freedom from the t-distribution are sufficiently small and that there are noticeable differences from the normal distribution). We know that both distributions are symmetric, but the t-distribution will have fatter tails. Hence, the quantiles near zero (confidence level = 50%) will match up quite closely. As we move further into the tails, the quantiles between the t-distribution and the normal will diverge, see Figure 3.

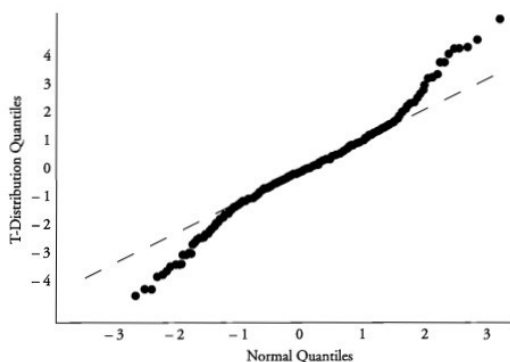


Figure 3: Q-Q plot: t distribution vs normal distribution

For example, at a confidence level of 95%, the critical z-value is -1.65, but for the t-distribution, it is close to -1.68 (degree of freedom of approximately 40). At 99% confidence, the difference is even larger, as the z-value is equal to -1.96 and the t-stat is equal to -2.02. More generally, if the middles of the Q-Q plot match up, but the tails do not, then the empirical distribution can be interpreted as symmetric with tails that differ from a normal distribution (either fatter or thinner).

3 Historical Simulation and Monte Carlo Simulation Methods

For our baseline models, we implement two models: Lognormal Monte-Carlo Model (Baseline-MC) and Historical Simulation (Baseline MC). These two models capture the basic features of VaR methodologies, and reveals potential issues that encourage us to seek more complicated solutions.

3.1 Model Specification

To evaluate these models, we construct a delta hedged equally weighted equity portfolio. This portfolio longs equity, short call options and risk-free bonds. It has initial net value of zero (self-financed) and rebalanced daily to keep delta neutral. Rebalancing gain and loss is invested into risk-neutral assets.

3.2 Baseline-HS

Baseline-HS does not assume any parametric underlying model for stock price change. Instead, it directly uses the non-parametric empirical distribution function of historical data. Therefore, it is largely free from the first problem, that is, the fat tails of empirical distribution of returns, but still suffers from the second one heavily, that is, the limitation and lag effect of volatility distribution of returns. When implementing Baseline-HS, we directly draw samples from historical data to simulate future price change.

3.3 Baseline-MC

Baseline-MC model assume that the return is log-normally distributed, as specified on most finance textbooks

$$dX_t = \mu dt + \sigma dB(t) \tag{3.1}$$

$$X_t = \log\left(\frac{P_t}{P_{t-1}}\right) \tag{3.2}$$

Then after estimating μ and σ , we can simulate the price change and then estimate the VaR. When implementing this method, we assume strong correlation between stocks. It will overestimate potential loss, but will provide a more solid estimation on the lower bound of loss.

However, the underlying assumption is highly problematic leading to two issues. The first one is the fat tail. As is shown in Figure 4 below, the empirical distribution of stock log returns has much fatter tails than normal distributions.

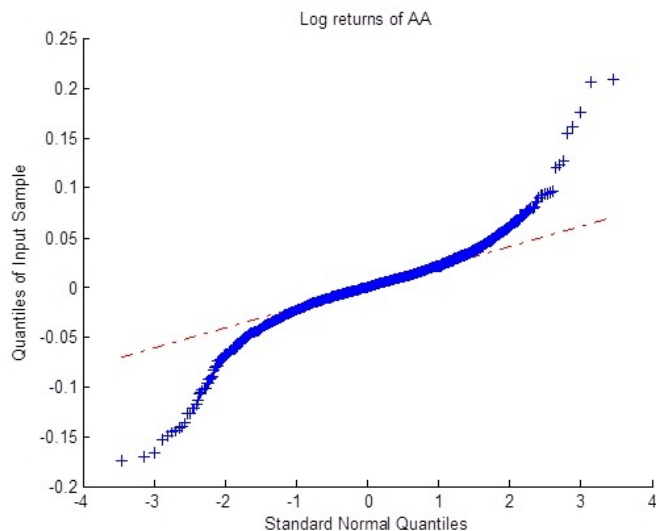


Figure 4: Log returns of AA

The second problem is the lag effect of volatility distribution, which is shown in Figure 5 below. Stock price movement is not stationary, so historical data is unable to capture the unprecedentedly large price volatilities, as have been shown in the financial crisis. After the crisis, when volatility fall back into normal level, the memory effect will cause these methods continue to report unnecessarily high risk, which causes a waste in resource allocation. These drawbacks will be reflected in the results.

3.4 Results Analysis

According to Figure 6 below, there are totally more than 10% violations during the 922 trading days evaluated. Moreover, it is apparent that all the tests failed.

The two models suffered heavily from the two problems we mentioned above considering Figure 7 below. When the financial crisis broke out, the portfolio suffered heavily in a long sequence of trading days. Due to the benign environment before the financial crisis and the slow response of our model, it still predicted very low potential loss. It was not until the crisis calm down that the model successfully incorporated the crisis information. As a result, the model seems too pessimistic, predicting too aggressive VaR. Such model, if used, will cause the investors to weight too much in safety asset and lead to a waste of resource.

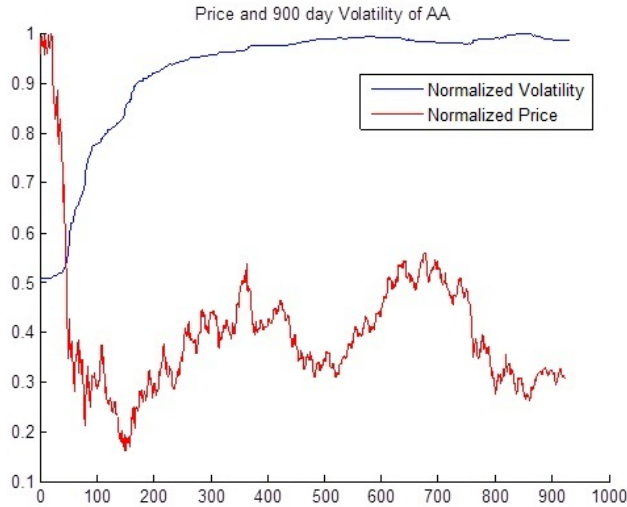


Figure 5: Price and 900 day Volatility of AA

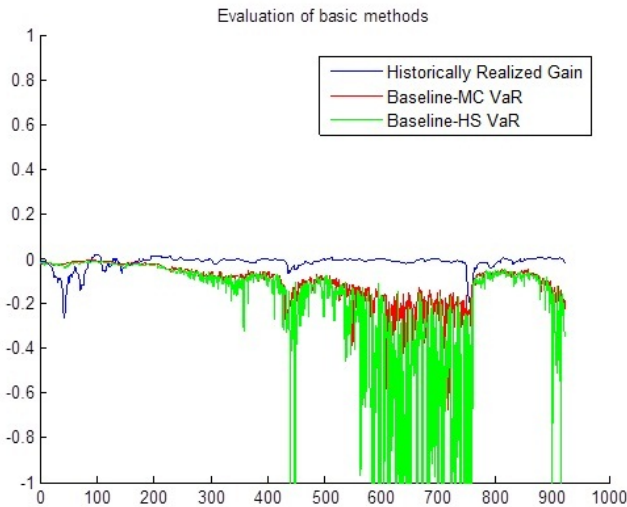


Figure 6: Evaluation of Basic Methods

In the following sections, we work to improve our models to address these two problems. For the first issue, we use EVT and t-distribution to model the heavy tails. For the second problem, we use heteroskedasticity models to capture time variant volatility. As shown below, these improved models work out quite decently.

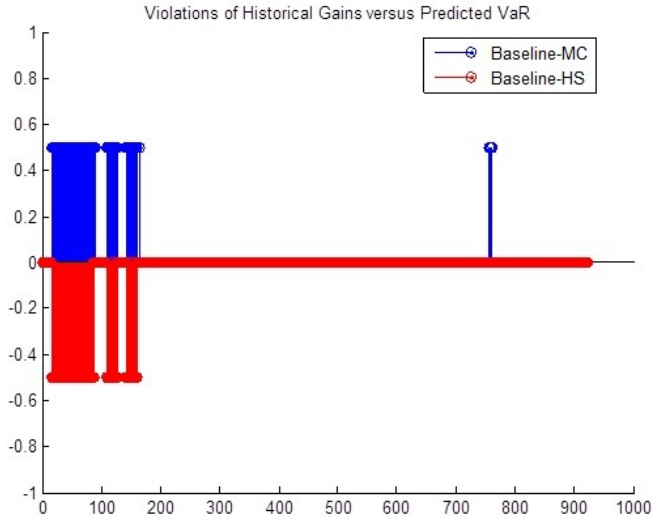


Figure 7: Violations of Historical Gains versus Predicted VaR

3.5 CVaR

The Conditional VaR (CVaR) is the expected value of the loss when it exceeds VaR. This measures the average of the loss conditional on the fact that it is greater than VaR. Define the VaR number as $-q$. Formally, the CVaR is the negative of

$$\mathbf{E}[X|X < q] = \frac{\int_{-\infty}^q x f(x) dx}{\int_{-\infty}^q f(x) dx} \quad (3.3)$$

Note that the denominator represents that probability of a loss exceeding VaR, which is also $p = 1 - c$. This ratio is also called the expected shortfall, tail conditional expectation, conditional loss, or expected tail loss. CVaR indicates the potential loss if the portfolio is “hit” beyond VaR. Because CVaR is an average of the tail loss, one can show that it qualifies as a subadditive measure.

We also make use of our models to study CVaR in Figure 8 below. It can be shown on the model that CVaR produce more aggressive predictions on the potential loss, although this is far not sufficient to address the severe issues we engaged in this case.

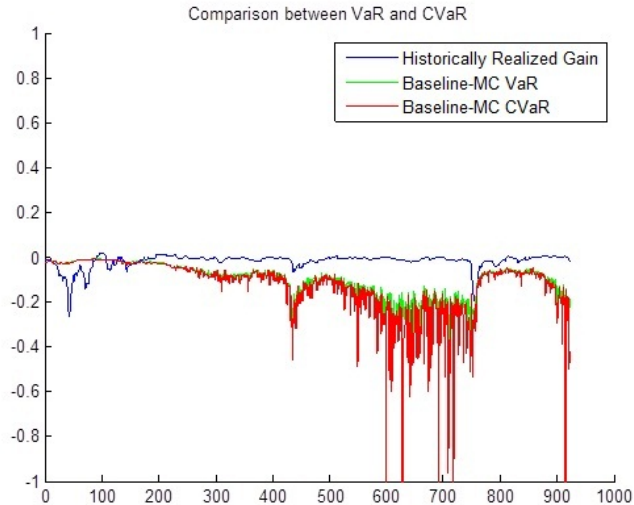


Figure 8: Comparison between VaR and CVaR

4 Unfiltered and Filtered Historical Simulation Methods

4.1 Normal Distribution Assumption

Input Data

- The daily historical returns of the delta hedging portfolio composed of 41 stocks and corresponding options based on Baseline-HS in Section 3

4.1.1 Unfiltered Historical Simulation and Normal Distribution of Returns

Procedure

- 1 Examine the historical portfolio returns**
 - 1.1** Plot the daily returns of the portfolio to observe general trend
- 2 Simulate the historical portfolio returns with normal distribution**
 - 2.1** Treat the the historical portfolio returns as a normal distribution sample, and estimate its mean and variance using MLE (Maximum Likelihood Estimate)
 - 2.2** Sample portfolio returns from the estimated normal distribution
- 3 Summarize the statistical results of portfolio returns**
 - 3.1** Interpret the results from the simulating returns, like maximal value, minimal value, VaR at various confidence levels, distribution function (CDF) and the probability density function (PDF) of the simulating portfolio returns

And we name it unfiltered MLE method under normal distribution assumption in the rest of the paper for identification and comparison.

4.1.2 Filtered Historical Simulation (FHS) and Normal Distribution of Standardized Residuals

Procedure Using Bootstrapping

1 Examine the historical portfolio returns

1.1 Plot the daily returns of the portfolio to observe general trend

2 Filter the historical portfolio returns

2.1 Plot the ACF of portfolio returns to observe autocorrelation and the ACF of squared portfolio returns to observe heteroskedasticity

2.2 Construct a first order autoregressive (AR(1)) model for autocorrelation and an asymmetric exponential GARCH (EGARCH) model for heteroskedasticity to filter portfolio returns, and apply the Augmented Dickey-Fuller test to justify the AR(1) model

2.3 Fit the AR(1) and EGARCH models based on historical portfolio returns, plot the residuals and volatilities for comparison

2.4 Get the standardized residuals = residuals / sigmas, plot the ACF of standardized residuals and squared standardized residuals to justify that standardized residuals are approximately i.i.d., and apply the Kolmogorov-Smirnov test to judge whether that standardized residuals come from a standard normal distribution

3 Simulate the historical portfolio returns with FHS

3.1 Bootstrap from the standardized residuals with replacement to obtain the simulating residuals

3.2 Use the simulating residuals in the simulation of the AR(1) and EGARCH models starting from the latest information (residual, sigma and return of the latest day) to predict the simulating returns

4 Summarize the statistical results of portfolio returns

4.1 Interpret the results from the simulating returns, like maximal value, minimal value, VaR at various confidence levels, distribution function (CDF) and the probability density function (PDF) of the simulating portfolio returns

And we name it filtered bootstrap method under normal distribution assumption in the rest of the paper for identification and comparison.

Procedure Using Empirical Distribution

3[◇] Simulate the historical portfolio returns with FHS

3.1[◇] Sample from a uniform distribution at random as the quantile in the empirical distribution to obtain the simulating residuals

With all else the same as **Procedure Using Bootstrapping**

And we name it filtered empirical method under normal distribution assumption in the rest of the paper for identification and comparison.

Procedure Using MLE of Distribution

3[◇] Simulate the historical portfolio returns with FHS

3.1[◇] Treat the standardized residuals as a normal distribution sample and estimate its mean and variance using MLE (Maximum Likelihood Estimate), and sample from the estimated normal distribution to obtain the simulating residuals

With all else the same as **Procedure Using Bootstrapping**

And we name it filtered MLE method under normal distribution assumption in the rest of the paper for identification and comparison.

4.2 t Distribution Assumption

Input Data

- The daily historical returns of the delta hedging portfolio composed of 41 stocks and corresponding options based on Baseline-HS in Section 3

4.2.1 Unfiltered Historical Simulation and t Distribution of Returns

Procedure

1 Examine the historical portfolio returns

1.1 Plot the daily returns of the portfolio to observe general trend

2 Simulate the historical portfolio returns with t distribution

2.1 Treat the the historical portfolio returns as a t distribution sample, and estimate its DoF (Degree of Freedom) using MLE (Maximum Likelihood Estimate)

2.2 Sample portfolio returns from the estimated t distribution

3 Summarize the statistical results of portfolio returns

3.1 Interpret the results from the simulating returns, like maximal value, minimal value, VaR at various confidence levels, distribution function (CDF) and the probability

density function (PDF) of the simulating portfolio returns

And we name it unfiltered MLE method under t distribution assumption in the rest of the paper for identification and comparison.

4.2.2 Filtered Historical Simulation (FHS) and t Distribution of Standardized Residuals

Procedure Using Bootstrapping

1 Examine the historical portfolio returns

1.1 Plot the daily returns of the portfolio to observe general trend

2 Filter the historical portfolio returns

2.1 Plot the ACF of portfolio returns to observe autocorrelation and the ACF of squared portfolio returns to observe heteroskedasticity

2.2 Construct a first order autoregressive (AR(1)) model for autocorrelation and an asymmetric exponential GARCH¹ (EGARCH) model for heteroskedasticity to filter portfolio returns, and apply the Augmented Dickey-Fuller test to justify the AR(1) model

2.3 Fit the AR(1) and EGARCH models based on historical portfolio returns, plot the residuals and volatilities for comparison

2.4 Get the standardized residuals = residuals / sigmas, plot the ACF of standardized residuals and squared standardized residuals to justify that standardized residuals are approximately i.i.d., and apply the Q-Q plot to justify that standardized residuals come from a t distribution with MLE of DoF

3 Simulate the historical portfolio returns with FHS

3.1 Bootstrap from the standardized residuals with replacement to obtain the simulating residuals

3.2 Use the simulating residuals in the simulation of the AR(1) and EGARCH models starting from the latest information (residual, sigma and return of the latest day) to predict the simulating returns

4 Summarize the statistical results of portfolio returns

4.1 Interpret the results from the simulating returns, like maximal value, minimal value, VaR at various confidence levels, distribution function (CDF) and the probability density function (PDF) of the simulating portfolio returns

¹There are three options here: GARCH, GJR and EGARCH. We would compare them with each other in this situation.

And we name it **filtered bootstrap method under t distribution assumption** in the rest of the paper for identification and comparison.

Procedure Using Empirical Distribution

3 \blacklozenge Simulate the historical portfolio returns with FHS

3.1 \blacklozenge Sample from a uniform distribution at random as the quantile in the empirical distribution to obtain the simulating residuals

With all else the same as **Procedure Using Bootstrapping**

And we name it **filtered empirical method under t distribution assumption** in the rest of the paper for identification and comparison.

Procedure Using MLE of Distribution

3 \blacklozenge Simulate the historical portfolio returns with FHS

3.1 \blacklozenge Treat the standardized residuals as a t distribution sample and estimate its DoF (Degree of Freedom) using MLE (Maximum Likelihood Estimate), and sample from the estimated t distribution to obtain the simulating residuals

With all else the same as **Procedure Using Bootstrapping**

And we name it **filtered MLE method under t distribution assumption** in the rest of the paper for identification and comparison.

4.3 GPD Assumption

Input Data

- The daily historical returns of each delta hedging unit composed of 1 stock and corresponding option based on Baseline-HS in Section 3

4.3.1 Filtered Historical Simulation (FHS) and GPD Standardized Residuals

Procedure Using EVT

1 Examine the historical returns of each unit

1.1 Plot the daily returns of each unit to observe general trend

2 Filter the historical returns of each unit

2.1 Plot the ACF of returns of each unit to observe autocorrelation and the ACF of squared returns of each unit to observe heteroskedasticity

2.2 Construct a first order autoregressive (AR(1)) model for autocorrelation and an asymmetric GARCH model for heteroskedasticity to filter returns of each unit, and apply the Augmented Dickey-Fuller test to justify the AR(1) model

2.3 Fit the AR(1) and asymmetric GARCH models based on historical returns of each unit, plot the residuals and volatilities for comparison

2.4 Get the standardized residuals = residuals / sigmas, plot the ACF of standardized residuals and squared standardized residuals to justify that standardized residuals are approximately i.i.d.

3 Estimate the semi-parametric CDFs of each unit

3.1 Fit the upper and lower tails of standardized residuals to a parametric GPD by maximum likelihood and the interior of standardized residuals to an non-parametric kernel CDF estimate, and construct a composite semi-parametric CDF for each unit

3.2 Concatenate graphically and display the three distinct regions of the composite semi-parametric empirical CDF to justify the lower and upper tail regions are suitable for extrapolation, while the kernel-smoothed interior is suitable for interpolation.

4 Assess the GPD fit of each unit

4.1 Plot the empirical CDF of the upper tail of standardized residuals along with the CDF fitted by the GPD to visually assess the GPD fit

5 Calibrate the t Copula

5.1 Transforms the standardized residuals to uniform variates by the semi-parametric empirical CDF derived above, and fit the t copula to the transformed quantile with its freedom parameter (DoF) and the linear correlation matrix (R) estimate

6 Simulate the portfolio returns with a t Copula

6.1 Transform the uniform variates to standardized residuals via the inversion of the semi-parametric marginal CDF of each unit by extrapolating into the GP tails and interpolating into the smoothed interior to obtain the simulating residuals

6.2 Use the simulating residuals in the simulation of the AR(1) and asymmetric GARCH models starting from the latest information (residual, sigma and return of the latest day) to predict the simulating returns

6.3 Form an equally weighted portfolio composed of all the units given the simulated returns of each unit to obtain the simulating portfolio returns

7 Summarize the statistical results of portfolio returns

7.1 Interpret the results from the simulating returns, like maximal value, minimal

value, VaR at various confidence levels, distribution function (CDF) and the probability density function (PDF) of the simulating portfolio returns

And we name it filtered empirical method under GPD distribution assumption in the rest of the paper for identification and comparison.

5 Backtesting Results and Methods Comparison

The whole sample period is 01/03/2005 to 04/16/2012, or 1821 trading days, denoted by $T_1, T_2, \dots, T_{1821}$. The data are obtained from Bloomberg, including 41 actively trading stocks and corresponding options in recent years.

We apply the technique of backtesting to compare the approaches in Section 4, regarding roughly half of the data from T_1 to T_{900} as the training set and the rest from T_{901} to T_{1821} as the test set.

Suppose today is T_i , $901 \leq i \leq 1821$, we focus on the VaR corresponding to T_i . Since the historical data available for T_i is $\{T_1, T_2, \dots, T_{i-1}\}$, we can choose the training set for T_i from the range of $\{T_{i-k}, T_{i-k+1}, \dots, T_{i-1}\}$, $1 \leq k \leq i - 1$.

Undoubtedly, when the number of used days or the size of used historical data varies, we could expect different results from the models based on the used training set on T_i . Given the fact that there are 250 trading days in 1 year, here we select $k = 250$ (1 year), 500 (2 years), 750 (3 years) & 900 (size of whole training set) to take the effect of different size of used training set into account.

5.1 Comparison between Unfiltered and Filtered Methods

From Table 5, it is surprising to notice that under normal and t distribution assumption, unfiltered MLE method outperforms filtered MLE method, considering the 99% VaR violation probability. However, this is not the truth if we further take a look at Figure 9 & 10. The relatively stable line indicates that the VaR from unfiltered MLE method is not responsive and sensitive to the historical data, leading to useless VaR measure and low predicting power especially under t distribution assumption.

Moreover, the result varies with different size of used training set, while it does not change the conclusion derived from relative comparison, which is similar to a robust test.

We further take a look at the effect of filtering on a selected example at random from filtered MLE method under t distribution assumption, see Figure 11 & 12. Based on the values of ACF before filtering (returns) and after filtering (standardized residuals), we can

Table 5: Comparison between unfiltered and filtered methods

Distribution Assumption	Training Set Size k	Method	99% VaR Violation Prob
Normal	250	Unfiltered MLE	0.54%
		Filtered MLE	1.85%
	500	Unfiltered MLE	1.41%
		Filtered MLE	2.06%
	750	Unfiltered MLE	1.63%
		Filtered MLE	1.95%
	900	Unfiltered MLE	1.63%
		Filtered MLE	1.74%
t	250	Unfiltered MLE	0.00%
		Filtered MLE	1.30%
	500	Unfiltered MLE	0.00%
		Filtered MLE	1.09%
	750	Unfiltered MLE	0.00%
		Filtered MLE	0.65%
	900	Unfiltered MLE	0.00%
		Filtered MLE	0.76%

see that AR-GARCH filtering model does solve the problem of autocorrelation and heteroskedasticity.

Having applied the Augmented Dickey-Fuller (ADF) test to justify the AR(1) part of the AR-GARCH filtering model, we should reject the null hypothesis of a unit root for each method in most cases when $k = 500, 750$ & 900 , while we cannot reject the null hypothesis of a unit root for each method in most cases when $k = 250$, which indicates the tradeoff between the increased statistical reliability when using longer time periods and the increased stability of the estimates when using shorter periods.

However, the ADF test itself does not necessarily mean that our data fits the AR-GARCH filtering model as a whole. Fortunately, we could supervise the goodness-of-fit and rationality of AR-GARCH filtering model from the feedback of Matlab, see Figure 13.

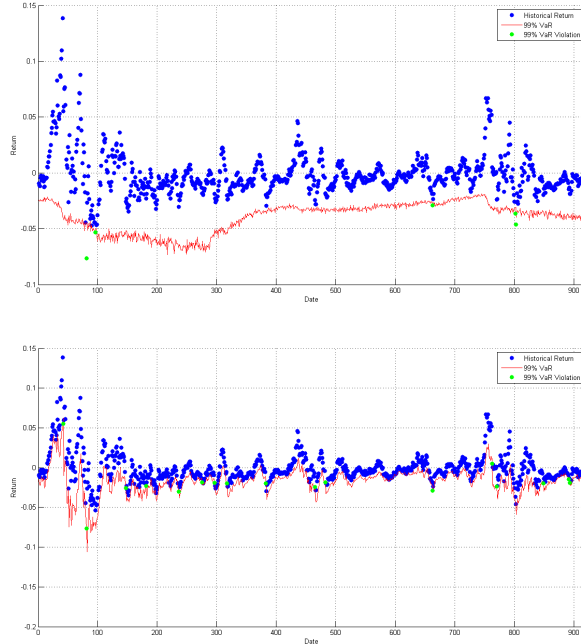


Figure 9: The VaR results of unfiltered and filtered MLE methods under normal distribution assumption with $k = 250$. Top panel: unfiltered MLE. Bottom panel: filtered MLE

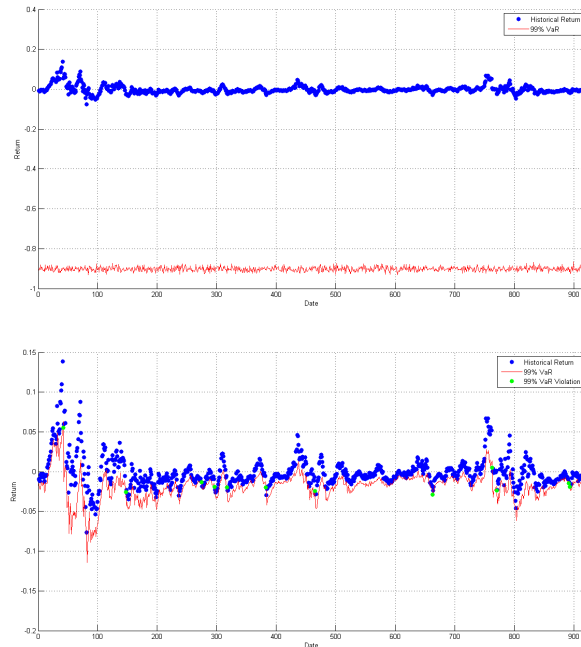


Figure 10: The VaR results of unfiltered and filtered MLE methods under t distribution assumption with $k = 250$. Top panel: unfiltered MLE. Bottom panel: filtered MLE

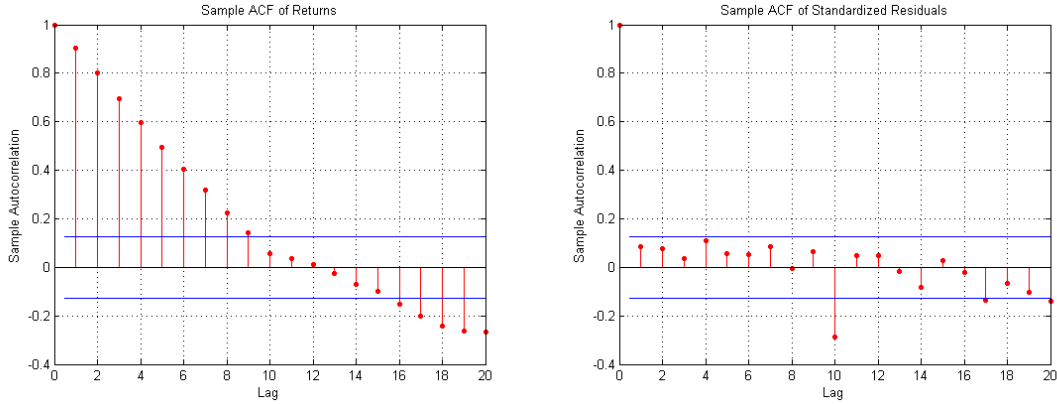


Figure 11: Sample ACF of Returns VS Sample ACF of Standardized Residuals from filtered MLE method under t distribution assumption with $k = 250$. Left panel: ACF of Returns. Right panel: ACF of Standardized Residuals

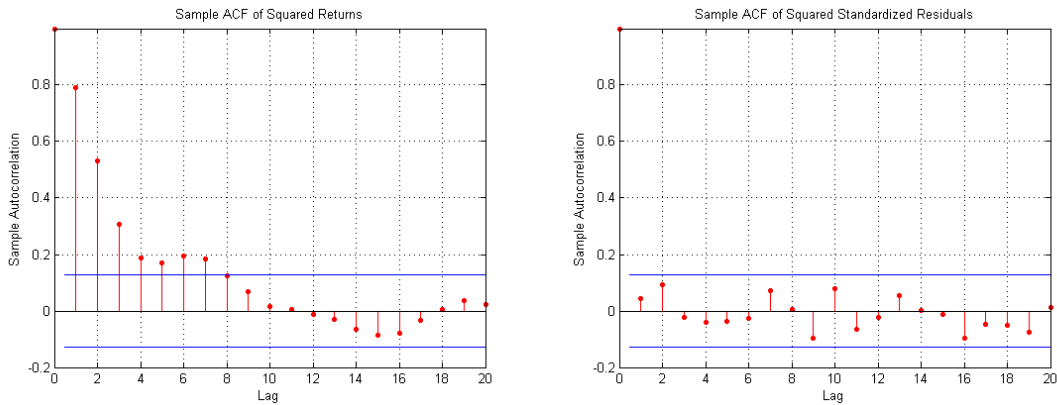


Figure 12: Sample ACF of Squared Returns VS Sample ACF of Squared Standardized Residuals from filtered MLE method under t distribution assumption with $k = 250$. Left panel: ACF of Squared Returns. Right panel: ACF of Squared Standardized Residuals

5.2 Comparison between Filtered Methods under Normal and t Distribution Assumption

From Table 6 & 7, we can see that all of the three filtered methods (bootstrap, empirical & MLE) under t distribution assumption outperform their counterparts under normal distribution assumption, considering the 99% VaR violation probability and the p-values of three related hypotheses (UCT, IND & CCD).

Moreover, the result varies with different size of used training set, while it does not change the conclusion derived from relative comparison, which is similar to a robust test.

Mean: ARMAX(1, 0, 0); Variance: EGARCH(1, 1)

Conditional Probability Distribution: T
 Number of Model Parameters Estimated: 7

Parameter	Value	Standard Error	T Statistic
C	-6.0532e-005	6.1183e-005	-0.9894
AR(1)	0.92152	0.011612	79.3580
K	-0.053552	0.042688	-1.2545
GARCH(1)	0.99561	0.0035402	281.2319
ARCH(1)	0.14103	0.033439	4.2176
Leverage(1)	0.080278	0.020447	3.9261
DoF	6.4308	1.051	6.1188

Figure 13: Goodness-of-fit of AR(1)-EGARCH(1,1) from filtered MLE method under t distribution assumption with $k = 900$

5.3 Comparison between Three Filtered Methods under t Distribution Assumption

From Table 8 & 9, we can see that among all of the three filtered methods (bootstrap, empirical & MLE) under t distribution assumption, MLE method outperforms bootstrap and empirical methods, considering the 99% VaR violation probability and the p-values of three related hypotheses (UCT, IND & CCD).

Moreover, the result varies with different size of used training set, while it does not change the conclusion derived from relative comparison, which is similar to a robust test.

5.4 Comparison between Three GARCH Models in the Filtered Bootstrap Method under t Distribution Assumption

From Table 10 & 11, we can see that among all of the three GARCH models (GARCH, GJR & EGARCH) in the filtered bootstrap method under t distribution assumption, no model outperforms the other two significantly, considering the 99% VaR violation probability and the p-values of three related hypotheses (UCT, IND & CCD).

Moreover, the result varies with different size of used training set, while it does not change the conclusion derived from relative comparison, which is similar to a robust test.

Table 6: Comparison between filtered methods under normal and t distribution assumption

Method	Training Set Size k	Distribution Assumption	99% VaR Violation Prob	UCT pValue	IND pValue	CCD pValue
Bootstrap	250	Normal	1.52% (14/921)	0.1407	0.5109	0.2721
		t	1.41% (13/921)	0.2372	0.1720	0.1957
	500	Normal	2.17% (20/921)	0.0020	0.4503	0.0063
		t	1.74% (16/921)	0.0418	0.2765	0.0696
	750	Normal	1.85% (17/921)	0.0210	0.3165	0.0422
		t	1.85% (17/921)	0.0210	0.3165	0.0422
900	Normal	1.85% (17/921)	0.0210	0.3165	0.0422	
	t	1.63% (15/921)	0.0788	0.4809	0.1664	
Empirical	250	Normal	1.85% (17/921)	0.0210	0.0361	0.0078
		t	1.63% (15/921)	0.0788	0.2390	0.1067
	500	Normal	2.06% (19/921)	0.0046	0.0581	0.0030
		t	1.63% (15/921)	0.0788	0.2390	0.1067
	750	Normal	1.63% (15/921)	0.0788	0.2390	0.1067
		t	1.63% (15/921)	0.0788	0.4809	0.1664
900	Normal	1.52% (14/921)	0.1407	0.2041	0.1508	
	t	1.95% (18/921)	0.0100	0.3589	0.0239	
MLE	250	Normal	1.85% (17/921)	0.0210	0.4239	0.0507
		t	1.30% (12/921)	0.3773	0.5735	0.5780
	500	Normal	2.06% (19/921)	0.0046	0.0581	0.0030
		t	1.09% (10/921)	0.7964	0.0928	0.2356
	750	Normal	1.95% (18/921)	0.0100	0.3589	0.0239
		t	0.65% (6/921)	0.2562	0.7791	0.5047
900	Normal	1.74% (16/921)	0.0418	0.2765	0.0696	
	t	0.76% (7/921)	0.4447	0.7433	0.7078	

5.5 Comparison between Filtered Methods under t and GPD Distribution Assumption

From Table 12, we can see that among all of the three filtered methods (bootstrap, empirical & MLE) under t distribution assumption, MLE method outperforms bootstrap and empir-

Table 7: Summary for Table 6

Distribution Assumption	Average 99% VaR Violation Prob	UCT Rejection Prob ($\alpha = 0.05$)	IND Rejection Prob ($\alpha = 0.05$)	CCD Rejection Prob ($\alpha = 0.05$)
Normal	1.84%	9/12	1/12	7/12
t	1.44%	3/12	0/12	2/12

Table 8: Comparison between three filtered methods under t distribution assumption

Training Set Size k	Method	99% VaR Violation Prob	UCT pValue	IND pValue	CCD pValue
250	Bootstrap	1.41% (13/921)	0.2372	0.1720	0.1957
	Empirical	1.63% (15/921)	0.0788	0.2390	0.1067
	MLE	1.30% (12/921)	0.3773	0.5735	0.5780
500	Bootstrap	1.74% (16/921)	0.0418	0.2765	0.0696
	Empirical	1.63% (15/921)	0.0788	0.2390	0.1067
	MLE	1.09% (10/921)	0.7964	0.0928	0.2356
750	Bootstrap	1.85% (17/921)	0.0210	0.3165	0.0422
	Empirical	1.63% (15/921)	0.0788	0.4809	0.1664
	MLE	0.65% (6/921)	0.2562	0.7791	0.5047
900	Bootstrap	1.63% (15/921)	0.0788	0.4809	0.1664
	Empirical	1.95% (18/921)	0.0100	0.3589	0.0239
	MLE	0.76% (7/921)	0.4447	0.7433	0.7078

Table 9: Summary for Table 8

Method	Average 99% VaR Violation Prob	UCT Rejection Prob ($\alpha = 0.05$)	IND Rejection Prob ($\alpha = 0.05$)	CCD Rejection Prob ($\alpha = 0.05$)
Bootstrap	1.66%	2/4	0/4	1/4
Empirical	1.71%	1/4	0/4	1/4
MLE	0.95%	0/4	0/4	0/4

Table 10: Comparison between three GARCH models in the filtered bootstrap method under t distribution assumption

Method	Training Set	Model	99% VaR	UCT	IND	CCD
	Size k	Type	Violation Prob	pValue	pValue	pValue
Bootstrap	250	GARCH	1.41% (13/921)	0.2372	0.5418	0.4129
		GJR	1.85% (17/921)	0.021	0.0361	0.0078
		EGARCH	1.41% (13/921)	0.2372	0.1720	0.1957
	500	GARCH	1.30% (12/921)	0.3773	0.5735	0.5780
		GJR	1.41% (13/921)	0.2372	0.1720	0.1957
		EGARCH	1.74% (16/921)	0.0418	0.2765	0.0696
	750	GARCH	1.85% (17/921)	0.0210	0.4239	0.0507
		GJR	1.41% (13/921)	0.2372	0.5418	0.4129
		EGARCH	1.85% (17/921)	0.021	0.3165	0.0422
	900	GARCH	1.52% (14/921)	0.1407	0.5109	0.2721
		GJR	1.63% (15/921)	0.0788	0.4809	0.1664
		EGARCH	1.63% (15/921)	0.0788	0.4809	0.1664

Table 11: Summary for Table 10

Model	Average 99% VaR	UCT Rejection	IND Rejection	CCD Rejection
Type	Violation Prob	Prob ($\alpha = 0.05$)	Prob ($\alpha = 0.05$)	Prob ($\alpha = 0.05$)
GARCH	1.52%	1/4	0/4	1/4
GJR	1.58%	1/4	1/4	1/4
EGARCH	1.66%	2/4	0/4	1/4

ical methods, while GPD appears to be in the middle, considering the 99% VaR violation probability and the p-values of three related hypotheses (UCT, IND & CCD). Besides, the 99% VaR results from these four methods are fairly close to each other, see Figure 14.

Moreover, the result varies with different size of used training set, while it does not change the conclusion derived from relative comparison, which is similar to a robust test.

We further take a look at the piecewise effect of fitting standardized residuals to a GPD distribution on a selected example at random from filtered method, see Figure 15. Based on

Table 12: Comparison between filtered methods under t and GPD distribution assumption

Training Set Size k	Distribution Assumption	Method	99% VaR Violation Prob	UCT pValue	IND pValue	CCD pValue
750	t	Bootstrap	1.85% (17/921)	0.0210	0.3165	0.0422
		Empirical	1.63% (15/921)	0.0788	0.4809	0.1664
	GPD*	MLE	0.65% (6/921)	0.2562	0.7791	0.5047
		Empirical	1.41% (13/921)	0.2372	0.1720	0.1957
900	t	Bootstrap	1.63% (15/921)	0.0788	0.4809	0.1664
		Empirical	1.95% (18/921)	0.0100	0.3589	0.0239
	GPD*	MLE	0.76% (7/921)	0.4447	0.7433	0.7078
		Empirical	1.74% (16/921)	0.0418	0.2765	0.0696

* The return of each unit might be rejected by the built-in AR-GARCH model of Matlab with reporting parameter limit errors, which indicates that the AR-GARCH model does not fit the return of a specific unit to some extent. In order to simplify the problem and carry out the comparison, here we replace the unavailable result of GPD with corresponding one of MLE. Besides, for $k = 750$ the failing probability is $22 / 921 = 2.39\%$, for $k = 900$ the failing probability is $17 / 921 = 1.85\%$.

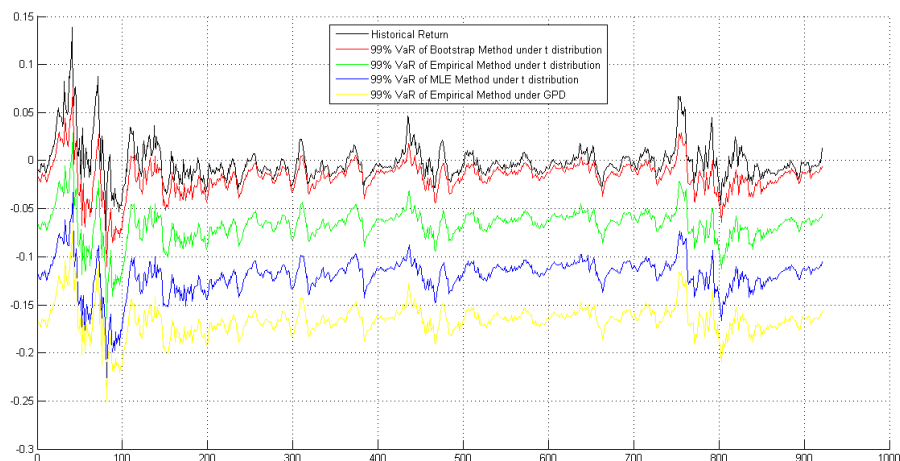


Figure 14: GPD cumulative distribution function versus empirical cumulative distribution function of the upper tail of standardized residuals in filtered method with $k = 900$. In order for clear distinction, the VaR curves shift downwards along Y-axis by 0 (bootstrap under t), 0.05 (empirical under t), 0.10 (MLE under t), 0.15 (empirical under GPD)

the plots of fitted GPD cumulative distribution function (CDF) and empirical cumulative distribution function (CDF) of standardized residuals, we can see that fitted GPD CDF is fairly close to empirical CDF, see Figure 16.

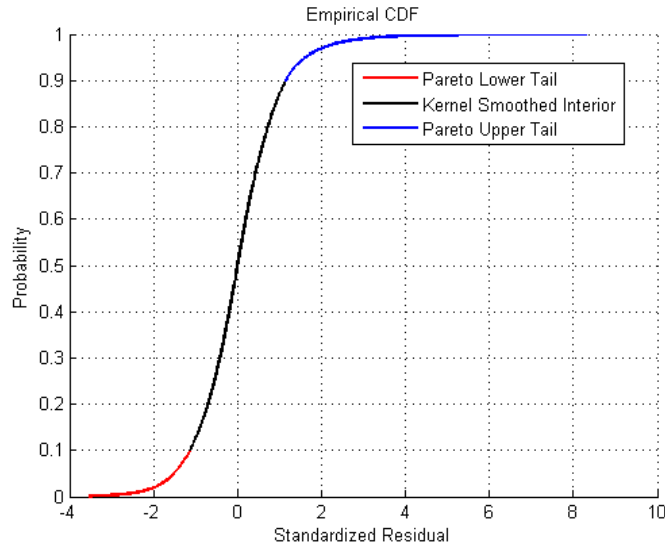


Figure 15: Comparison of 99% VaR results from four filtered methods with $k = 900$

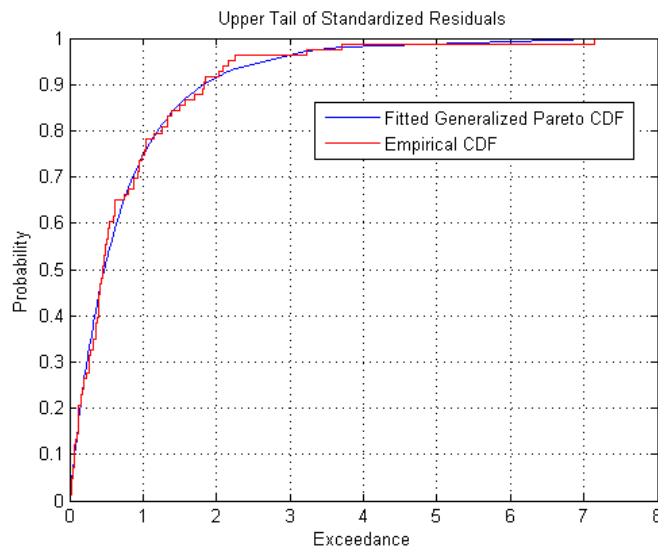


Figure 16: GPD cumulative distribution function versus empirical cumulative distribution function of the upper tail of standardized residuals in filtered method with $k = 900$

6 Conclusion

Summary for Section 5 is as below

- **Section 5.1 justifies the advantages of AR-GARCH model filtering.**
- **Section 5.2 justifies the advantages of t distribution assumption.**
- **Section 5.3 justifies the advantages of MLE method in filtered methods under t distribution assumption.**
- **Section 5.4 demonstrates insignificant difference between three GARCH models in filtered methods under t distribution assumption.** It indicates that they have similar capability to capture the characteristics of underlying returns and predicting power.
- **Section 5.5 demonstrates the middle position of filtered method under GPD assumption in filtered methods under t distribution.** However, taking the high computation complexity of GPD into account, there is not much gain resulting from the sophisticated technique to deal with the heavy tail as a separate distribution.

We further take a look at the essence of different distribution assumptions (normal, t & GPD) in filtered methods, which leads to different capability of models to capture the characteristics of standardized residuals, and therefore different predicting power in backtesting.

Having applied the Kolmogorov-Smirnov (KS) test to judge whether that standardized residuals come from a standard normal distribution, we should reject the null hypothesis of a standard normal distribution for each method in most cases when $k = 500, 750$ & 900 , while we cannot reject the null hypothesis of a standard normal distribution for each method in most cases when $k = 250$, which indicates the tradeoff between the increased statistical reliability when using longer time periods and the increased stability of the estimates when using shorter periods.

However, the KS test itself does not necessarily mean that normal distribution is undesirable since we do not have rigorous and precise tests for t distribution and GPD at hand. Part of the reason is that t distribution differs significantly with different degree of freedom and it is hard to decide that a symmetric centered distribution does not come from a t distribution, while GPD is a generalized semi-parametric distribution and it lacks suitable measures as consistent traditional distributions. Therefore, we turn to Q-Q plots for consistent relative comparison.

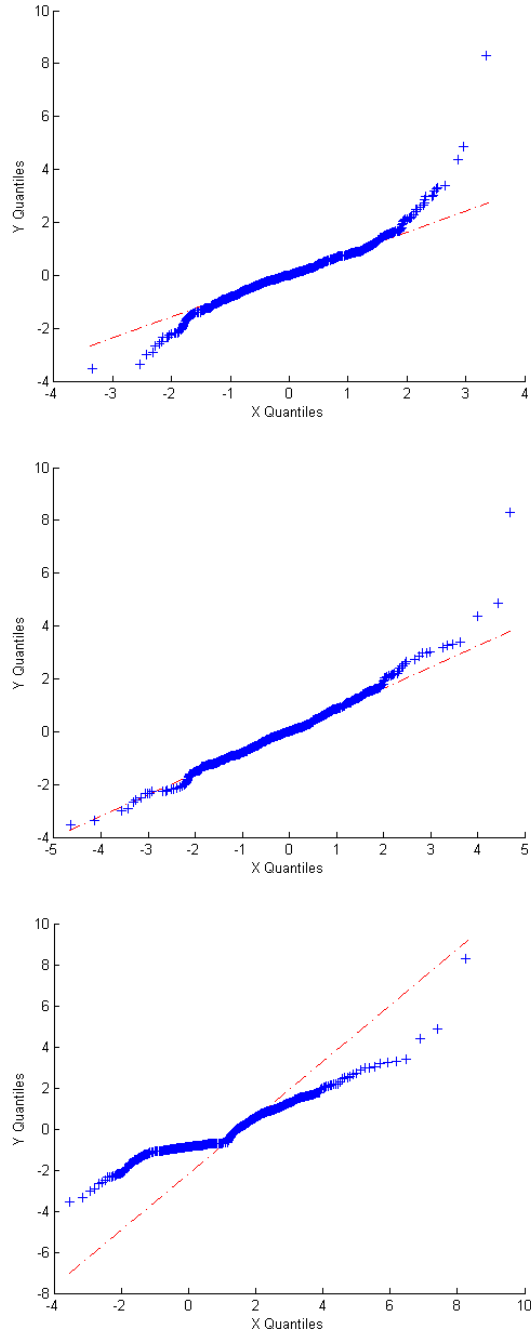


Figure 17: Q-Q plots of standardized residuals of AA versus different distributions (normal, t & GPD) in filtered methods with $k = 900$. Top panel: normal distribution. Middle panel: t distribution. Bottom panel: GPD

From Figure 17, we can see that t distribution and GPD outperform normal distribution remarkably, since they capture the most of the distribution characteristics of standardized residuals. On the other hand, t distribution displays slightly tails than standardized residual-

s, GPD displays slightly heavier tails than standardized residuals, while normal distribution deviates from the distribution characteristics of standardized residuals due to fairly heavy tails. Therefore it justifies the advantages of t distribution and GPD in all three distribution assumptions from a deeper theoretical level.

7 Future Work

There are several potential extensions of our current work

1 Confidence Interval of VaR

In this work we have provided point estimation of VaR. Although this satisfies the common demand, an ideal result should include estimation on the confidence interval.

Since VaR is the estimated quartile value (therefore non-linear in expected mean), the confidence interval cannot be obtained in common way. We have found the approach in doing that, by employing Sectioning techniques which combine the estimation results from a group of sections. This method is quite computational costly. If we use only 20 sections with 500 runs in each section, that is 10000 run for a single estimation. Considering the simulation of about 900 days and 10 day VaR, it typically takes several days to run the program.

However, confidence interval may be useful in applications. Our early evaluations show that the upper and lower bounds of VaR do have impact on exceptions, especially when the margin is tight between VaR and realized gain. Simulation techniques can be explored to reduce the required runs to estimate confidence interval.

2 CVaR

One critics at VaR is that it only cares the presence of exceptions without noting the potential loss for the exceptions. CVaR is designed to address this problem, and it is useful when the tail is fat and long. Our data set does not exhibit such extreme shape of distributions (recall that even GPD's tail is too fat), but CVaR shall be quite useful if we explore certain extremely asymmetric strategies like carry trade.

3 Automatic Model Selection

Different portfolio weights and strategies will lead to different distributions of gain and loss. These distributions may have different skewness and kurtosis, which fits different residual models. Ideally, we wish to work out a framework that evaluates a group of scenarios and select the best model. This framework, if implemented, can be very helpful in portfolio management.

References

- [1] Ryozi Miura and Shingo Oue (2000), Statistical Methodologies for Market Risk Measurement, *Asia Pacific Financial Markets*, **7**, 305-319.
- [2] Daniel B. Nelson (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, **59**, 347-370.
- [3] Tim Bollerslev (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.
- [4] Roger B. Nelsen (1999), *An Introduction to Copulas*, Springer.
- [5] Giovanni Barone-Adesi and Kostas Giannopoulos (2000), Non parametric VaR Techniques. Myths and Realities, *Economic Notes*, **30**, 167-181.
- [6] G. Barone-Adesi, F. Bourgoin, K. Giannopoulos (1998), "Don't look back", *Risk*, **11**, 100-104.
- [7] Kerry Back (2005), *A Course in Derivative Securities: Introduction to Theory and Computation*, Springer.
- [8] VolX, <http://www.volx.us/realizedvolatility.htm>
- [9] VolX, <http://www.volx.us/VolFormula.htm>
- [10] VIX and More, <http://vixandmore.blogspot.com/2009/08/gap-between-vix-and-realized-volatility.html>
- [11] MathWorks, Using Bootstrapping and Filtered Historical Simulation to Evaluate Market Risk, http://www.mathworks.com/products/econometrics/demos.html?file=/products/demos/shipping/econ/Demo_RiskFHS.html
- [12] MathWorks, Using Extreme Value Theory and Copulas to Evaluate Market Risk, http://www.mathworks.com/products/econometrics/demos.html?file=/products/demos/shipping/econ/Demo_RiskEVT.html
- [13] MathWorks, Pricing American Basket Options by Monte Carlo Simulation, http://www.mathworks.com/products/econometrics/demos.html?file=/products/demos/shipping/econ/Demo_AmericanBasket.html

- [14] TradingFloor, <http://www.tradingfloor.com/posts/delta-hedging-spot-up-or-down-who-cares-431809581>
- [15] J. Rank and T. Sigel (2002), *Applied Quantitative Finance*, Springer.
- [16] Romain Berry (2009), An overview of value-at-risk: part III - Monte Carlo Simulation- s VaR, http://www.jpmmorgan.com/tss/General/Risk_Management/1159380637650, J.P. Morgan Investment Analytics & Consulting.
- [17] Simone Manganelli and Robert F. Engle (2001), Value at Risk Models in Finance, *Working Paper*, European Central Bank.
- [18] P. Embrechts, A. McNeil and D. Straumann (1999), Correlation: Pitfalls and Alternatives, *RISK*, 69-71.
- [19] A. Sklar (1959), Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ.*, **8**, 229-231.
- [20] Christoph Hartz, Stefan Mittnik and Marc Paoella (2006), Accurate value-at-risk forecasting based on the normal-GARCH model, *Computational Statistics & Data Analysis*, **51**, 2295-2312.
- [21] Ahmed Shamiri and Abu Hassan (2007), Modeling and Forecasting Volatility of the Malaysian and the Singaporean stock indices using Asymmetric GARCH models and Non-normal Densities, *J. Malaysian Math. Sci.*, **1**, 83-102.
- [22] Peter Christofferson (2003), Elements of Financial Risk Management, *Elsevier Science*.
- [23] M. Pritsker (2001), The Hidden Dangers of Historical Simulation, *Working Paper*.
- [24] William F. Sharpe (1966), Mutual Fund Performance, *The Journal of Business*, **39**, 119-138.
- [25] J. D. Jobson and Bob M. Korkie (1981), Performance hypothesis testing with the Sharpe and Treynor measures, *Journal of Finance*, **36**, 889-908.
- [26] Paul H. Kupiec (1995), Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivative*, **2**, 73-84.
- [27] P. F. Christofferson (1998), Evaluating Interval Forecast, *International Economic Review*, **39**, 841-862.

- [28] G. Casella and Roger L. Berger (2001), *Statistical inference*, 2nd Ed, Duxbury Press.
- [29] Tomas Björk (2009), *Arbitrage Theory in Continuous Time*, 3rd Ed, Oxford University Press.
- [30] Philippe John (2011), *Financial Risk Manager Handbook*, 6th Ed, Wiley.