

# Evaluating the Accuracy of Value at Risk Approaches

Kyle McAndrews

April 25, 2015

## 1 Introduction

Risk management is crucial to the financial industry, and it is particularly relevant today after the turmoil of the Great Recession. Within risk management, Value at Risk became the gold standard in the mid-to-late 1990s. In 1996, the Basel Committee on Banking Supervision at the Bank for International Settlements required financial firms to use VaR to determine adequate capital requirements (Engle, Manganelli, 2004). Despite its popularity, this risk management method is controversial. The mathematics and assumptions behind Value at Risk have been called into question over the years, and in light of the recent financial downturn, it seems even more fitting to evaluate its success today. The goal of this paper is to empirically evaluate and compare the different approaches of Value at Risk.

### 1.1 What is Value at Risk?

Value at Risk, or VaR, provides a single value to summarize the risk of an asset or portfolio due to general market movements over a specified length of time. VaR can be viewed as the minimum loss for a given probability,  $p$ , over a given time period.

Below is the equation for VaR at time index  $t$ . Note that  $F_t$  (VaR) is the cumulative distribution function (CDF), and  $\Delta V(l)$  is the change in the value of the asset or portfolio from time  $t$  to  $t + l$ . VaR is equivalent to the  $p * 100$ th quantile of the distribution, where  $p$  refers to a probability, by definition between 0 and 1.

$$p = \Pr[\Delta V \leq \text{VaR}] = F_t(\text{VaR}).$$

In words, the above equation says that a portfolio is likely to change in value more than VaR with a given probability,  $p$ . For a portfolio with a 1% VaR of \$1 million, the VaR suggests that the portfolio can expect to lose more than \$1 million 1% of the time. Note that VaR does not indicate what the loss will likely be in these extreme times the VaR value is surpassed. By itself, this statistic does not explain the behavior of the tails of the distribution; it only gives an estimation for how often the given loss will be surpassed.

This paper first introduces the mathematics behind certain approaches to Value at Risk. Next, this paper will “backtest” these different approaches. Backtesting analyzes the accuracy of these models by comparing projected VaR values with historical returns. Backtesting is hugely important for risk managers and regulators to help calibrate effective models in the real world. It is also central to the Basel Committee’s ultimate decision to allow internal VaR models to be used to determine capital requirements.

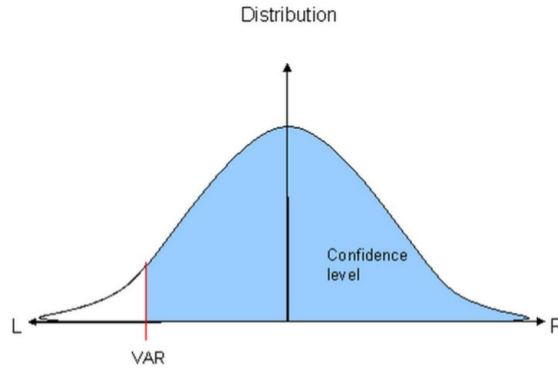


Figure 1: Illustration of Value at Risk. For the above distribution of asset returns, the VaR value is shown for a given confidence level,  $(1-\alpha)$ . Returns will be to the left of the VaR value with probability  $\alpha$ . Note that VaR is understood to refer to the left tail; therefore 99% and 1% VaR are understood to be the same.

## 1.2 Previous research

Value at Risk can be applied to a single asset, such as a stock, or an entire portfolio of varied assets. In practice, banks calculate VaR in two ways. First, the banks can calculate the VaR for individual assets, and then combine these VaR values by calculating the correlations among each pair of two assets. Secondly, banks can map assets onto risk factors (such as interest rate changes, exchange rate changes, etc.). In this approach, the bank determines the sensitivity of each asset to given risk factors (like how a bond responds to a change in interest rate, known as duration), and the bank then determines the correlations among the individual risk factors. These approaches are referred to as multivariate.

While multivariate is common in practice, much of the literature comparing alternate approaches treats portfolios as a single asset; this is known as the univariate approach. This approach does not require calculating the correlations between assets or risk factors. The univariate approach instead looks at the overall profit and loss of the portfolio. For example, a paper might look at the returns of the NASDAQ, without looking at the returns of each individual stock that makes up the exchange. Papers using this approach include Kuester et. al (2006), So and Yu (2005), and Gaglianone et. al (2008). The decision for this paper to use portfolio returns for backtesting is discussed in more detail below.

The literature compares a wide array of different approaches for calculating VaR. The deciding criteria for most of the comparisons is the frequency of VaR breaks, or the frequency that the portfolio returns surpass the VaR value. It is important to note that this paper will evaluate the accuracy of these approaches using the frequency of VaR breaks. For a bank in practice, there is an idiosyncratic trade-off related to financial risk models. The banks balance the benefit of conservative financial risk models with the cost. More specifically, banks balance the benefit of protection in bad market environments with the opportunity cost of unused capital. As these optimization problems are outside the scope of this paper, this paper will focus on the accuracy of the VaR approaches. For example, for a 5% VaR, the approach deemed optimal in the context of this paper will be the approach that results in a frequency of VaR breaks closest to 5%.

The first focus of my paper, the historical simulation VaR approach, is widely analyzed in the literature, including by Kuester et. al (2006), Gustafsson and Lundberg (2009), and Raaji and Raunig (1998). The second focus of the paper, the parametric approach that uses volatility modeling, is also widely analyzed in the literature, including by Kuester et al. (2006), Mokni et al. (2009), Gustafsson and Lundberg (2009), Nyssanov (2013), and So and Yu (2005).

Recall that VaR refers to a probability for a loss over a given time period. This time period can vary depending on which time frame is most valuable or relevant to the financial institution. The literature predominantly uses daily returns to calculate daily VaR. The backtesting in this paper will first use daily VaR values, and it will then extend to a 10-day time period.

### 1.3 Portfolio choice

As introduced above, banks calculate VaR using a multivariate approach. Calculating the necessary correlations can require immense computing power. For the multivariate asset approach, banks need to calculate the correlation between every group of two assets within the portfolio. In the case of a portfolio of 50 assets, there would be  $\binom{50}{2}$ , or 1225 different correlations to calculate.

In practice, Berkowitz and O'Brien (2002) noted that some large trading portfolios can have positions in tens of thousands of assets, making it virtually impossible for the risk models to estimate daily VaR values by calculating the relationships between all pairs of variables in their model. To overcome some of the computational difficulty, banks use approximations. In their paper, Berkowitz and O'Brien compare the multivariate VaR models banks use to a standard model that only looks at the returns on the portfolio as a whole, an approach that does not require calculating correlations. Berkowitz and O'Brien show that the complex VaR models are in fact not better than the simple, portfolio-as-an-asset VaR calculations. In their study, even though bank VaR values were more conservative, their analysis showed that the bank-calculated VaR values did not lead to significantly less VaR Breaks than their own model did.

This might seem counterintuitive; if bank VaR values from the multivariate approach provide more conservative VaR values (i.e. provide higher VaR values), wouldn't these, almost by definition, be surpassed and broken less than the more aggressive portfolio model would be? While this would seem to be the case, Berkowitz and O'Brien find that the portfolio-as-an-asset approach performs just as well as the banks' multivariate approach because their method responded more quickly to changing volatility environments. (Their paper used a GARCH volatility model, which will be introduced below). Note that volatility is equivalent to the standard deviation and square root of the variance of the asset. In their paper, Berkowitz and O'Brien find that the returns are not independent. In other words, if today has a very negative return, tomorrow is more likely to have a very negative return. This gives their model an advantage, as their volatility model more quickly adjusted to this changing volatility. If today has a very negative return, the VaR value for tomorrow determined using their model will increase more than the multivariate approach would in response to today's negative return.

Based on the above findings, this paper will use the univariate, portfolio-as-an-asset approach. This analysis can be viewed as helpful in two ways. First, it can be assumed to be a good proxy for a true multivariate approach (largely supported by Berkowitz and O'Brien's findings). Secondly, this paper can be viewed as an evaluation of the accuracy of one component of the multivariate approach, still very important to the overall effectiveness of a VaR model. In other words, this paper can be viewed as a test for calculating the VaR for a single asset, though the banks in practice will then aggregate.

The historical asset returns used will come from Wharton Research Data Services (WRDS). This section will utilize equally weighted S&P 500 log returns from 12/01/1992 through 12/31/2014. Note this data set focuses on larger companies as it uses companies in the S&P 500. Within the index, the equal weighting calculates the return as if an equal amount of capital is invested into each company in the index, regardless of size. This time period was chosen to encapsulate many different market environments. These returns do include dividends.

### 1.3.1 Why do we use log returns?

Using log returns is standard for risk managers. The simple return of an asset is defined as  $R_t = \frac{P_t + D_t}{P_{t-1}}$  where  $P_t$  is the price of the asset at time  $t$  and  $D_t$  is the dividend paid at time  $t$ , if applicable. The continuously compounded return is the natural log of this simple return of an asset. Log return is defined as

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t + D_t}{P_{t-1}}\right).$$

A benefit of using log returns is that it allows multi-period returns to be defined as the sum of individual period returns.

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) \\ &= \ln[(1 + R_t)(1 + R_{t-1}) \cdot (1 + R_{t+k-1})] \\ &= \ln[(1 + R_t)] + \ln[(1 + R_{t-1})] + \dots + \ln[(1 + R_{t+k-1})] \\ &= r_t + r_{t-1} + \dots + r_{t+k-1}. \end{aligned}$$

If each daily return is assumed to follow a log normal distribution, in order for the  $k$ -day distribution to follow a normal distribution, daily returns are assumed to be independent. An additional benefit of using log returns is that if these returns are distributed log normally, the distribution can never lead to a negative price; as the  $\log(P_t/P_{t-1})$  goes to  $-\infty$ ,  $P_t$  goes to 0. (Tsay)(Jorion)

## 2 Different ways to calculate VaR

### 2.1 Historical simulation introduction

The first approach discussed in this paper to calculate Value at Risk is the historical simulation approach. In a survey conducted in 2012 by McKinsey, 75% of the banks interviewed used historical simulation to calculate VaR. This approach calculates VaR for a given asset by looking at the previous returns on that given asset over a specified length of time. In the McKinsey report, 40% of the banks using historical simulation used a one-year window, and the remaining banks used multi-year windows, spanning two to five years. The length of the window indicates how many previous asset returns the VaR calculation deems relevant.

In addition to varying the time window of previous returns, banks also vary the weights given to each return. While some banks use equal weighting (implying each return within the window is equally important), other banks use weighting in order to emphasize the more recent returns. Banks might use weighting under the assumption that recent returns on an asset are more indicative of future asset behavior. For the historical simulation in this paper, equal weighting will be used.

As mentioned above, Value at Risk looks at the  $p * 100$ th percentile of the distribution of an asset return. Parametric approaches (such as the variance-covariance approach introduced below) require assumptions about the distributions of the assets in order to estimate this percentile. The strength of the historical simulation approach is that the distributional assumption is not required; this is an example of a non-parametric approach. To calculate VaR using historical simulation, the firm will order the past returns, from lowest return to highest return, for the given asset. In order to calculate the  $p * 100$ th VaR value, the firm will look at the  $p * 100$ th percentile in the distribution of past returns. For example, if there are 100 previous returns included in the relevant historical window, a 5 % VaR will look at the 5th percentile or the 5th lowest return.

This approach inherently assumes that future behavior of an asset will be similar to historical behavior of that asset over a previous time window. While this is a potential weakness, a strength of this approach is that it does not require any assumptions about the underlying distribution of asset returns.

## 2.2 Variance-covariance approach introduction

Unlike historical simulation, the variance-covariance approach requires an assumption for the distribution of asset returns; the standard assumption, and the assumption used in this paper, is that the asset returns follow a normal or log normal distribution. In order to apply a log normal distribution, as done in the backtesting of this paper, the mean, denoted  $\mu_t$  and standard deviation, denoted  $\sigma_t$ , need to be specified. With these values, the Value at Risk can be defined. To show how the mean and standard deviation are used in this model, first consider our original definition of Value at Risk.

$$\Pr[\Delta V \leq \text{VaR}] = p.$$

In order to determine the VaR value, we “standardize” the distribution by subtracting the mean from both sides of the inequality, and then dividing by the standard deviation.

$$\Pr\left[\frac{\Delta V - \mu}{\sigma} \leq \frac{\text{VaR} - \mu}{\sigma}\right] = p$$

By assuming that  $\Delta V$  follows a normal distribution, the new standardized variable,  $\frac{\Delta V - \mu}{\sigma}$ , follows a normal distribution with a mean of 0 and standard deviation of 1. The cumulative distribution function of a standard normal for a probability  $p$  is also known as a Z score. Plugging in Z for  $\frac{\Delta V - \mu}{\sigma}$  and solving for VaR, we get

$$\frac{\text{VaR} - \mu}{\sigma} = Z,$$

$$\text{VaR} = Z\sigma + \mu.$$

For a normal distribution, the approximate values for Z are known for given probabilities. For  $p = .05$ , the Z score is approximately 1.645, and for  $p = .01$ , the Z score is approximately 2.33.

As mentioned, for the variance-covariance approach assuming a normal distribution, there are two inputs:  $\mu$  and  $\sigma$ . There are two general ways to determine these inputs. First, the statistics can be determined historically, by looking back at previous asset returns. Secondly, the statistics can be determined looking “forward”, using current market prices to determine implied volatility. The first approach, using historical asset prices, is referred to as “backward looking”, as it looks backward to determine current variance. The second approach is “forward looking” in that it determines these current parameters from what is implied in the market. In order to use market prices, the

risk manager needs a market pricing model that can observe the price and determine the implied desired statistics. For example, Black-Scholes is a derivative-pricing model that analysts can use to determine the implied volatility of an underlying stock by observing the prices of derivatives on that stock. The advantage of this approach is that it is “forward looking” and reflects the current investor sentiment surrounding the assets. The drawback of this approach is that the approach is only as good as the model, as the statistic is a complete function of the model used. Additionally, while some derivatives have widely accepted pricing models, for certain assets, it is not obvious which model to use.

With these drawbacks in mind, this paper will determine the volatility and the mean using historical returns. This paper will determine the two relevant parameters through both volatility modeling and a simple trailing variance and mean calculation (discussed more below). Looking at historical variance in order to talk about current variance is motivated by the idea of volatility clustering. This concept was introduced by Mandelbrot in 1963 when he commented on stock returns. He explained that “large changes tend to be followed by large changes of either sign, and small changes tend to be followed by small changes”. If a stock was actively traded and saw a large return, either positive or negative, one day, it is likely to be actively traded and have a large return, either positive or negative, the following day. This concept is fundamental in risk management.

### 2.2.1 Volatility modeling: GARCH

The common model used in volatility modeling is the generalized autoregressive conditional heteroskedastiy, or GARCH, model. The GARCH model assumes that current volatility, or the current standard deviation of an asset, is a linear combination of previous squared stock returns above the mean and previous stock variance. See the equation of a GARCH(m,s) below.  $a_t$  is the “shock” value of an asset, which is the return of the asset above the mean. Therefore, the return of an asset is defined as  $r_t = \mu + a_t$ . Risk managers often assume that returns when measured on a daily frequency have a mean of 0. If this is the case, the shock of the asset is just its return. (The effect of this assumption will be tested in this paper). The shock value of an asset is function of a changing standard deviation,  $\sigma_t^2$ , and a noise term,  $\varepsilon_t$ , which follows a normal distribution with a mean of 0 and standard deviation of 1.

$$a_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2.$$

In practice, the coefficients of the model,  $m$ ,  $s$ ,  $\alpha_i$ , and  $\beta_i$ , will be optimized using maximum likelihood. This optimization determines the parameters that would make recreating the data the most likely.

As an example, let the volatility structure be best described by a GARCH(2,2). This model structure suggests that current volatility is best modeled by looking at two previous squared shock values and the two previous stock volatilities. Note that volatility is equivalent to the standard deviation. To calculate the VaR value for a given day, the GARCH(2,2) will model today’s variance. The square root of that variance will then be multiplied by the Z score to determine the VaR. In the context of this paper, a GARCH(1,1) will be used.

### 2.2.2 Historical variance

An additional way to determine current volatility is to calculate the sample variance and then take the square root. Below, the sample variance is calculated over the previous  $m$  returns. This approach is unweighted, as each return in the window has equal importance.

$$\sigma_n^2 = \frac{1}{m-1} \sum_{i=1}^m (a_{n-i} - \bar{a})^2$$

Recall that using historical variance is unlike the historical simulation approach discussed above, as this approach still requires a distributional assumption. Additionally, when using this approach, the risk manager needs to choose the window size, and the weighting scheme. These decisions reflect how many returns the risk manager deems relevant, and whether or not the risk manager deems each of those returns equally important. This paper will test a range of windows with equal-weighting.

### 2.3 Criteria for backtesting: Kupiec test

The Kupiec test will be used to evaluate the success of the different approaches. This test was introduced by Paul Kupiec in 1995. This statistic is used to test the null hypothesis that the probability of a VaR break for the proposed model is actually the desired VaR probability; i.e., for a 5% VaR run below, the null hypothesis will be that the true probability of a VaR break is 5%. This model treats the backtesting as a Bernoulli trial, solely measuring the outcomes as a discrete distribution of successes and failures. Note that this framework assumes each return is independent. The number of exceptions, or VaR breaks, is assumed to follow the binomial distribution

$$f(x) = \binom{T}{x} p^x (1-p)^{T-x},$$

where  $p$  is the probability of a VaR break for the designed test. The expected value of a binomial distribution in this form is  $pT$  with a variance is  $p(1-p)T$ . The Kupiec test uses the above framework and maximizes the power of the test, meaning it maximizes the probability of correctly rejecting the null hypothesis. The Kupiec test finds a 95% confidence region. This test is defined by the log-likelihood ratio

$$LR_{uc} = -2 \ln((1-p)^{T-N} p^N) + 2 \ln((1-(N/T))^{T-N} (N/T)^N).$$

This ratio follows a chi-square distribution with one degree of freedom. If the LR statistic returned is greater than 3.84, the null hypothesis, that the probability of the test is the desired probability, is rejected. (Jorion)

## 3 Backtesting: daily VaR values

The first backtesting looks at daily returns to predict changes in the value of a portfolio over a single day. Multiple approaches are used to provide a comparison. This length for VaR is often used by financial firms whose portfolios change daily due to intraday trading.

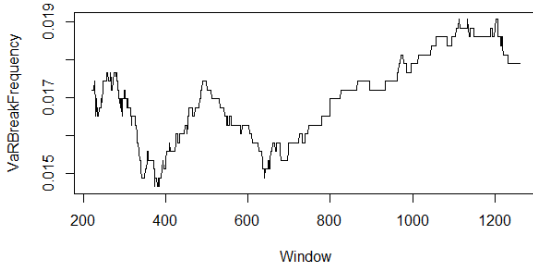
### 3.1 Historical simulation

There are two tests conducted for the historical simulation approach. For the first, the windows used range from 220 to 1260, approximately ranging between one and five years (as there are approximately 252 business days in a year). For the second approach, the window for the estimation stretches all previous data; therefore, the window size increases by 1 for each subsequent VaR estimation.

For the first approach, for a given window,  $n$ , the first day tested for a VaR break is  $n + 1$ . In order to be consistent, the first day tested in every simulation was the 1261st, day as this is the first day the largest window can be applied to. By this constraint, every approach in this paper will be tested on 4303 daily returns. This simulation uses log returns, as is standard in practice.

Below is the VaR break frequency for the range of windows used to calculate 1 % VaR values. The minimum VaR break frequency is 1.464095%, with 63 breaks. This minimum value occurred for multiple values between 373 and 384, translating to a window of roughly a year and one-half. The minimum value has a LR statistic of 8.189647, surpassing the 3.84 cutoff. As the minimum does not satisfy the Kupiec test, each window size fails, providing evidence that the historical simulation approach is not appropriate for a 1% VaR. This result makes some sense; a big challenge in financial math is simulating the tail of a distribution for an asset, and this issue is exacerbated when dealing with smaller VaR predictions.

Figure 2: Var breaks for changing windows, 1 %



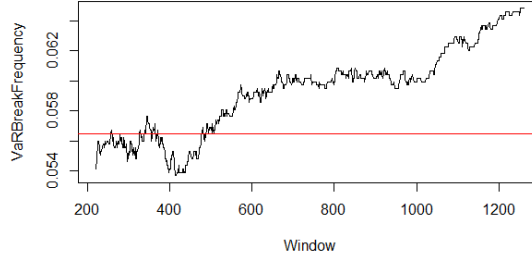
Next, the 5% VaR was tested for the same window range. The minimum VaR break frequency was 5.368348%, occurring with window sizes 413 through 416, or with a window a little more than a year and one-half. This minimum value occurs with 231 breaks, leading to a LR statistic of 1.201536. The null hypothesis that the probability of a VaR break is 5% is not rejected.

For the above 5% VaR using historical simulation, the last number of VaR breaks that does not reject the null for the Kupiec test is 243. Ultimately, only 30.125% of the windows do not reject the null hypothesis. VaR break frequencies below the red line in the graph do not reject the null hypothesis.

For the second approach of historical simulation, the VaR estimate is created using all previous information. As mentioned, for consistency, the first day estimated was the 1261<sup>st</sup> day. Therefore, each data point is created using at least 5 years of previous information. For the 1% VaR value, the VaR break frequency was 2.556356% with 110 breaks. This leads to a LR statistic of 73.6066, easily rejecting the null hypothesis. The 5% VaR value performed even poorer. With a VaR break



Figure 3: Var breaks for changing windows, 5 %



frequency of 8.435975%, the LR statistic is 89.46004, clearly rejecting the null. These last two simulations provide support for the idea of volatility clustering. The underlying idea of this concept is that recent market movements are most relevant when predicting current market movements. By including all previous information, the effect of recent market activity is diluted.

### 3.2 Variance-covariance approach: GARCH(1,1)

This section explores the success of the variance-covariance approach using a GARCH(1,1) to model volatility. As mentioned above, a GARCH(1,1) assumes that current volatility is a linear combination of yesterday's squared shock value and yesterday's volatility. The formula is below.

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

The first iteration of this model assumes that the mean on each return is 0. This assumption is sometimes used by risk managers when dealing with daily returns as it decreases computational burden. Recall that when the mean is 0, the shock value,  $a_t$ , is just the return,  $r_t$ . Therefore, with a mean of 0, GARCH is a linear combination of previous volatility and squared returns.

The Berkowitz O'Brien paper mentioned above used a GARCH(1,1) and ARMA(1,1) to model the volatility and mean, respectively. An autoregressive moving average model, or ARMA, is very similar to a GARCH model, though it is applied to the return series directly in order to model the mean return,  $\mu_t$ . The ARMA(1,1) structure assumes that the current expected return is a linear combination of yesterday's return,  $r_{t-1}$ , and yesterday's "noise", or error term.

$$r_t = \phi_1 r_{t-1} + \theta_1 \varepsilon_{t-1}.$$

As shown in the original derivation of Value at Risk, the VaR estimate will be different if there is an estimated mean value, as  $\text{VaR} = Z\sigma + \mu$ . In this section, this assumption that the mean is zero will be tested, to see if including an ARMA structure for the mean return has a significant effect on the accuracy of the model.

For any volatility model, the optimal coefficients are determined by a subset of the data. The risk manager must decide what subset of the data to use for this optimization, a similar decision to

the size of the window used in historical simulation. Recall that the driving force behind volatility modeling is volatility clustering; if volatility is changing over time, isn't it possible the best model to describe the volatility structure changes as well?

Berkowitz and O'Brien estimate new GARCH and ARMA parameters for each day throughout the sample using all data up to that point. Therefore, on the 1000th day, the GARCH and ARMA parameters will be estimated on the previous 999 days. To keep consistency, the first VaR value estimated was for the 1261st day.

Note that later GARCH coefficient estimations will have more data to work with. This is common in the literature. Recall that the VaR estimation for each day will still be dependent on the same number of previous squared shock values and volatilities; all that is changing is the size of the sample used to estimate the proper GARCH and ARMA coefficients.

The initial GARCH(1,1) for a 1% VaR without modeling the mean resulted in a VaR break frequency of 1.975366%. With 85 breaks, the LR statistic is 32.20299, easily rejecting the null. When adding the ARMA(1,1) structure to model the mean, the VaR break frequency was the same. As it also rejected the null, adding the ARMA structure does not help accurately model the desired VaR value.

Next, for a 5% VaR without modeling the mean, the VaR break frequency was 5.809900%. With 250 breaks, the LR statistic is 5.660683, and the null is rejected. When modeling the mean, the 5% VaR produced a VaR break frequency of 5.623983%, a clear improvement over the GARCH(1,1) run by itself. With 242 breaks, the LR statistic is 3.396242, and the null is not rejected.

There are two interesting conclusions from this section. The first, similar to the results from the historical simulation, is that these models perform better on 5% tests than 1% tests. Secondly, these tests provide evidence against the decision by some risk managers to ignore the mean for daily returns. While there didn't seem to be an improvement for the 1% test, modeling the mean did provide enough of an improvement for the 5% VaR to change the result of the Kupiec test.

### 3.3 Standard historical variance

For the final variance-covariance approach, a standard historical variance was calculated. This approach also serves as a good comparison for the GARCH volatility model. Similar to the historical simulation above, the sample variance used windows ranging from 220 to 1260.

For calculating a 1% VaR value, the minimum VaR break frequency was 1.975366%, occurring with windows of 335 and 336, or roughly a year and a quarter. With a LR statistic of 32.20299, the null is rejected. Therefore, the entire range of window sizes rejects the null hypothesis.

For the 5% VaR value, the minimum VaR break frequency was 5.135952%, which occurred with a window size of 245. With 221 breaks and a LR statistic of 0.1660168, the null is not rejected. In fact, 87.0317% of window sizes tested do not reject the null. Every VaR break frequency at or below the red line does not reject the null.

The standard variance provides further support for the strength of the parametric approach and for the improved accuracy of testing for the 5% versus the 1% VaR.

### 3.4 Summary for daily return backtesting

The above backtesting provides valuable insight into the performances of these different VaR approaches. One common thread is the poorer performance for the 1% VaR values. This is what would be expected; as tail behavior in finance is such a challenge to model, the more extreme the

Figure 4: Var breaks, standard variance, 1%

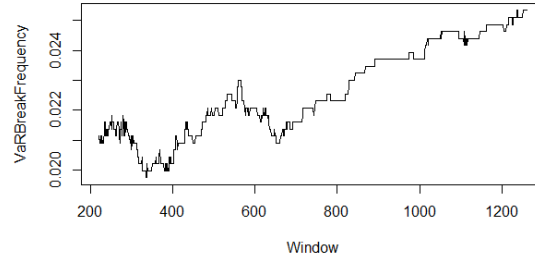
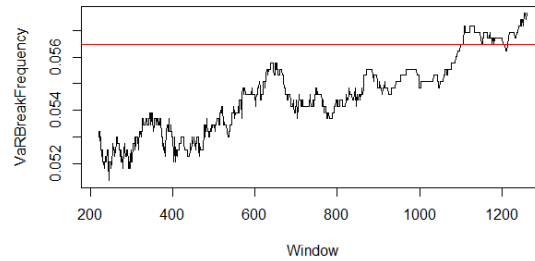


Figure 5: Var breaks, standard variance, 5%



VaR value, the less accurate it is likely to be. It is interesting to note that both volatility models, even though they require the distributional assumption, largely outperformed the historical simulation. More tests should be run to determine which of the two volatility models performs better. As mentioned above, this empirical backtest was used solely on an index, treating a portfolio as a single asset. For future analysis, it would be interesting to see if these results hold when using a multivariate approach.

## 4 Backtesting: 10-day VaR values

The second section of backtesting uses two approaches to estimate 10-day VaR values. These VaR values estimate a change in value that will be exceeded over a 10-day threshold. This additional backtesting is interesting for a few reasons. First, for some firms, 10-day VaR estimates can be more helpful and are the standard; VaR values of this length or longer are particularly relevant to financial firms that do not trade daily, like commercial banks that predominately issue loans. Secondly, this section will test the independence of returns, a fundamental assumption when calculating Value at Risk.

Figure 6: Summary table

| Approach                                      | Minimum VaR Break Frequency | Min Window | Reject the null? (percentage not rejected) |               |
|---|-----------------------------|------------|--|---------------|
| Historical simulation, ranging windows        | 1%                          | 1.464%     | 373-384                                    | Yes (0%)      |
|   | 5%                          | 5.368%     | 413-416                                    | No (30.125%)  |
| Historical simulation, all available info     | 1%                          | 2.556%     | NA   | Yes           |
|   | 5%                          | 8.436%     | NA   | Yes           |
| Standard historical variance, ranging windows | 1%                          | 1.975%     | 335-336                                    | Yes (0%)      |
|   | 5%                          | 5.136%     | 245  | No (87.0317%) |
| GARCH(1,1)                                    | 1%                          | 1.975%     | NA   | Yes           |
|   | 5%                          | 5.810%     | NA   | Yes           |
| GARCH(1,1) + ARMA(1,1)                        | 1%                          | 1.975%     | NA   | Yes           |
|   | 5%                          | 5.624%     | NA   | No            |

#### 4.1 How to calculate k-day VaR

In order to calculate VaR for a k-day period, recall how to add expected values and variances of random variables.

$$E[X_1 + X_2] = E[X_1] + E[X_2],$$

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2).$$

In order to calculate VaR for multiple periods, we assume that daily returns are independent. This is a very important assumption. In words, this implies that the likelihood of a given VaR break tomorrow is  $p$ , no matter what happened today, or yesterday, or so on. Therefore, we are assuming that VaR breaks are randomly distributed through the data set and are not abnormally clustered. This assumption is based on the idea of efficient markets, which states that all public information is properly priced into the asset, and prices only change in response to unanticipated news. In other words, this assumes that stocks follow a random walk. Looking back at the variance of two random variables, independence assumes the covariance term is equal to zero. Additionally, this relationship also assumes that the expected value and variance of a return are constant over time; in other words,  $Var(X_t) = Var(X_{t-1}) = Var(X)$  and  $E[X_t] = E[X_{t-1}] = E[X]$ .

Under these considerations, the variance of a return over  $k$  days is

$$Var\left(\sum_{i=0}^{k-1} R_{t-i}\right) = k\text{VaR}(R),$$

and the expected return is  $E[\sum_{i=0}^{k-1} R_{t-i}] = kE[R]$ . As Value at Risk is calculated using the standard deviation, the  $k$ -day VaR is

$$VaR = Z\sqrt{k}\sigma + k\mu.$$

If the daily return has an expected return of 0, notice that  $VaR_k = \sqrt{k}\text{VaR}$ . This section will make this assumption. For this backtesting, there are two ways to calculate a 10-day VaR, and both use historical simulation. The first two tests will use historical simulation and look at past 10-day returns directly. The second test will use the historical simulation approach to find a 1-day VaR, and then extend it to  $k$  days.

## 4.2 Historical simulation: 10-day overlapping samples

Similar to the historical simulation approach above, this approach looks at the  $p * 100$ th percentile of the empirical distribution of assets returns. Instead of looking at daily returns, this approach adds up 10 daily log returns to determine the 10-day return. The  $i$ th element in the 10-day return data set is the 10-day return ending on the  $i$ th day. This is calculated by adding the daily returns for the  $i - 9$  to  $i$  return (as these are log returns, it is a simple sum). For the  $i + 1$  day, the return will be calculated by summing the  $i - 8$  to  $i + 1$  daily returns. Notice that these 10-day return windows are overlapping.

To run this test, for the VaR value for the  $j$ th day, the algorithm will look at the  $j - 1$  10-day return to the  $j - \text{window}$  10-day return. For this test, like the historical simulation above, the window will range from 220 to 1260 previous 10-day returns. Over these previous returns, the algorithm will order the 10-day returns and find the  $p$ th quantile. This VaR value will be compared to the 10-day return starting on the  $j$ th day, i.e. it will be compared to the sum of returns from  $j$  to  $j + 9$ . It should be noted that the below tests were used on 4285 returns, instead of 4303 returns like above. This is a function of the length of the returns and how it limits the use of the data.

The minimum VaR break frequency for calculating a 5% VaR value was 6.184364% with 265 breaks. This occurs with a window size of 398 and a LR statistic of 11.80607, clearly rejecting the null hypothesis that the true probability of a VaR break is the intended 5%. As this is the minimum break, each window size rejects the null.

Figure 7: 10-Day VaR breaks, calculated with 10-day historical simulation, 5%

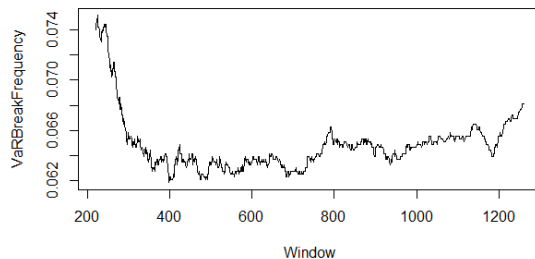
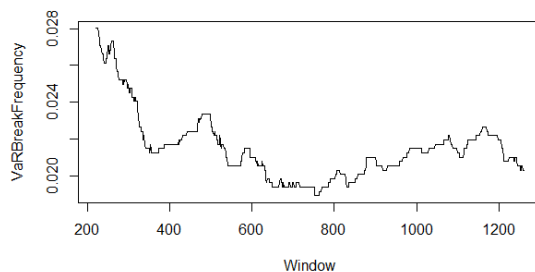


Figure 8: 10-Day VaR breaks, calculated with 10-day historical simulation, 1%



The minimum VaR break frequency for calculating a 1% VaR value was 1.890315%, with 81 breaks. This occurred with window sizes ranging from 752 to 763. The LR statistic is 27.19657 and the null is rejected. As this is the minimum VaR break frequency, each window size is rejected.

### 4.3 Historical simulation: 10-day non-overlapping samples

This test uses historical simulation for 10-day VaR values, but the historical returns do not overlap. The advantage of an overlapping sample is that it provides more data points from a given sample. This said, overlapping samples can cause statistical issues when estimating parameters, as this technique adds dependency amongst the data points. With an overlapping sample, consecutive 10-day returns “share information”, as they include 9 of the same daily returns.

This test hopes to provide a comparison to isolate the effect of adding a dependency structure. This algorithm will look at all groups of 10-day returns back to the first 10-day return that can be calculated. For the  $j$ th day, the first return will be calculated from the  $j - 1$  daily return to the  $j - 10$  daily return, and the second return will be calculated from the  $j - 11$  daily return to  $j - 20$  daily return. Unlike the historical simulation above, this algorithm does not impose a range of window sizes. As the non-overlapping constraint greatly decreases the number of historical data points, this test uses the maximum number of previous returns. As a standard historical simulation,

once the historical sample is created, the algorithm will look at the  $p$ th quantile, and this VaR value will be compared to the 10-day return.

The first day tested is the 1270th, just as the other 10-day VaR tests. With this constraint, there are 126 initial data points to estimate the first VaR. At the end of the sample, there will be 428.

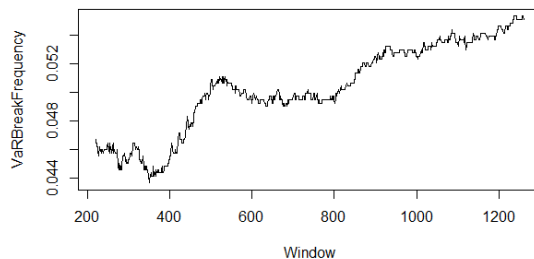
For a 1% VaR value, the VaR break frequency was 2.357%. With an LR statistic of 57.69857, the null is clearly rejected. The 5% VaR value did not perform much better, as the VaR break frequency was 7.491%. With an LR statistic of 48.88321, the null is again rejected. While the non-overlapping sample does not have the issue of adding correlation among the sample, it did not improve the VaR estimation.

#### 4.4 Historical simulation: extending daily VaR

For the second 10-day VaR approach, the daily VaR was calculated and then multiplied by  $\sqrt{10}$ . As mentioned, this assumes returns are independent. As these VaR estimations are based on daily VaR values, there is no issue from overlapping samples.

For the 5% VaR test, the minimum VaR break was actually 4.364061%, far below the 5% VaR value. Note that there were 4285 VaR values estimated. With 187 breaks, the LR statistic is 3.804941, not rejecting the null. This occurred with a window size of 350. This is the first test in this paper that almost rejects as a result of too conservative (or too high) of a VaR value. Ultimately, every window tested did not reject the null hypothesis.

Figure 9: 10-Day VaR breaks, extending 1-day VaR, 5%

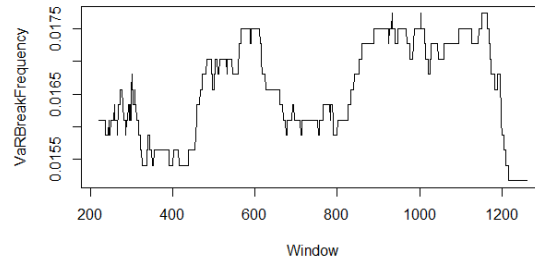


When looking at the daily VaR extended to 1% VaR values, the minimum VaR break frequency is 1.516919%. This occurred with window sizes from 1217 until the maximum window, 1260. With 65 breaks out of 4285, the LR statistic is 9.984467, clearly rejecting the null hypothesis. In fact, every window size rejects the null hypothesis.

#### 4.5 Conclusion

Ultimately, the 10-day VaR value was better calculated when extending the 1-day VaR value. This is surprising as the 1-day VaR value puts emphasis on the independence assumption, an assumption that can cause issues when calculating VaR. It should be noted that as both the overlapping and non-overlapping historical simulations performed poorly, it is not clear if the added dependency

Figure 10: 10-Day VaR breaks, extending 1-day VaR, 1%



structure causes issues. For further research, it would be interesting to see if these results hold up in non-parametric VaR estimation, and to see if these results are replicated in the multivariate approach.

## 5 Bibliography

Berkowitz, Jeremy, and James O'Brien. "How Accurate Are Value-at-Risk Models at Commercial Banks?" *The Journal of Finance* 57.3 (2002): 1093-111. Web.

De Raaji, G., and B. Raunig. "A Comparison of Value at Risk Approaches and Their Implications for Regulators." (n.d.): n. pag. Web.

Gustafsson, M., and C. Lundberg. "An Empirical Evaluation of Value at Risk." University of Gothenburg (n.d.): n. pag. Web.

Jorion, Philippe. *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill, 2001. Print.

Kuester, K., S. Mittnik, and M. Paoletta. "Value-at-Risk Prediction: A Comparison of Alternative Strategies." *Journal of Financial Econometrics* 4.1 (2005): 53-89. Web.

Mehta, A., M. Neukirchen, S. Pfetsch, and T. Poppensieker. "Managing Market Risk: Today and Tomorrow." McKinsey & Company (n.d.): n. pag. Web.

Mokni, Khaled, Zouheir Mighri, and Faysal Mansouri. "On the Effect of Subprime Crisis on Value-at-Risk Estimation: GARCH Family Models Approach." *International Journal of Economics and Finance* 1.2 (2009): n. pag. Web.

Nyssonov, A. "An Empirical Study in Risk Management: Estimation of Value at Risk with GARCH Family Models." Uppsala University (n.d.): n. pag. Web.

So, Mike K.p., and Philip L.h. Yu. "Empirical Analysis of GARCH Models in Value at Risk Estimation." *Journal of International Financial Markets, Institutions and Money* 16.2 (2006): 180-97. Web

Tsay, S. Ruey. "Analysis of Financial Time Series". John Wiley & Sons



Figure 11: Summary table

| Approach                       | Minimum VaR Break Frequency | Min Window | Reject the null? (percentage not rejected) |
|--------------------------------|-----------------------------|------------|--|
| 10-day sample, overlapping     | 1%                          | 1.890%     | 752-763<br>Yes (0%)                        |
|                                | 5%                          | 6.184%     | 398<br>Yes (0%)                            |
| 10-day sample, non-overlapping | 1%                          | 2.357%     | NA<br>Yes                                  |
|                                | 5%                          | 7.491%     | NA<br>Yes                                  |
| 1-day sample extended          | 1%                          | 1.517%     | 1217-1260<br>Yes (0%)                      |
|                                | 5%                          | 4.364%     | 350<br>No(100%)                            |