

Extreme Value Theory with High-Frequency Financial Data

Abhinay Sawant

Professor George Tauchen, Faculty Advisor

*Honors Thesis submitted in partial fulfillment of the requirements for Graduation
with Distinction in Economics in Trinity College of Duke University.*

Duke University
Durham, North Carolina
2010

Acknowledgements

I would like to thank Professor George Tauchen for providing his help in the development of this paper and for leading the Honors Finance Seminar. I would like to thank Professor Tim Bollerslev, whose insights led to the foundation of the research in this paper and for his help in the research process. Finally, I would like to thank the rest of the members of the Honors Finance Seminar for their help in providing comments and suggestions to develop the paper: Pongpitch Amatyakul, Samuel Lim, Matthew Rognlie and Derek Song.

Abstract

Extreme Value Theory (EVT) is one of the most commonly applied models in financial risk management for estimating the Value at Risk of a portfolio. However, the EVT model is practical for estimation only when data is independent and identically distributed, which usually does not characterize financial returns data. This paper aims to modify this model by using high-frequency data to standardize financial returns by their realized volatility and then tests the modified model with recent equity data. The results from the paper show an improvement in the EVT model when forward volatility can be properly forecasted.

1. Introduction and Motivation

In the study and practice of financial risk management, the Value at Risk (VaR) metric is one of the most widely used risk measures. The portfolio of a financial institution can be enormous and exposed to thousands of market risks. The Value at Risk summarizes these risks into a single number. For a given portfolio of assets, the N -day X -percent VaR is the dollar loss amount V that the portfolio is not expected to exceed in the next N -days with X -percent certainty. Proper estimation of VaR is necessary in that it needs to accurately capture the level of risk exposure that the firm is exposed to, but if it overestimates the risk level, then the firm will set unnecessarily set aside excess capital to cover the risk, when that capital could have been better invested elsewhere (Hull, 2007).

One method of determining the N -day X -percent VaR of a portfolio is to model the distribution of changes in portfolio value and then to determine the $(100-X)$ -percentile for long positions (left tail) and the X -percentile for short positions (right tail). For simplicity, many practitioners have modeled changes in portfolio value with a normal distribution (Hull, 2007). However, empirical evidence has shown that asset returns tend to have distributions with fatter tails than those modeled by normal distributions and with asymmetry between the left and right tails (Cont, 2001).

As a result, several alternative methods have been proposed for estimating VaR, one of which being the Extreme Value Theory (EVT). EVT methods make VaR estimations based only on the data in the tails as opposed to fitting the entire distribution and can make separate estimations for left and right tails (Diebold et al., 2000). Several studies have shown EVT to be one of the best methods for application to VaR estimation. Ho et al. (2000) found the EVT approach to be a much stronger method for estimating VaR for financial data from the Asian Financial Crisis when compared to fitting data to normal and student distributions and other methods such as using percentiles from historical data. Gencay and Selcuk (2004) found nearly similar results when applying these methods to emerging markets data and found EVT to especially outperform the other methods at higher percentiles such as 99.0, 99.5 and 99.9 percent.

One issue with the implementation of the EVT approach is the requirement that the financial returns data be independent and identically distributed (Tsay, 2005). However, due to the presence of volatility clustering, this may not apply to financial asset

returns data. Volatility, as typically measured by the standard deviation of financial asset returns, tend to “cluster.” Days with high volatility tend to be followed by days with high volatility. Therefore, returns from two days in a sample of asset returns may be correlated due to volatility and changes in volatility environments may significantly impact the distribution of asset returns (Stock & Watson, 2007).

The goal of this paper is to counteract this independent and identically distributed issue by using high-frequency financial data. High-frequency data are data sampled at higher frequency than just daily closing prices. For example, the data set in this paper contains minute-by-minute sampled price data of S&P 100 stocks. Literature has shown that data sampled at high frequency can provide accurate estimates of volatility. This paper improves the VaR model with the EVT approach by first standardizing daily returns by their daily realized volatility. Through this standardization technique, the data become more independent and identically distributed and so more suited for use in the VaR model. This paper also explores other uses of high-frequency data such as the concept of an intraday VaR, which uses shorter time periods such as half-day and quarter-day as independent trials.

2. Description of the Model

2.1: Definition of Value at Risk (VaR)

The Value at Risk (VaR) is usually defined in terms of a dollar loss amount in a portfolio (e.g. \$5 million VaR for a \$100 million portfolio); however, for the purposes of this paper, the value at risk will instead be defined in terms of a percentage loss amount. This way, the metric can be applied to a portfolio of any initial value (Hull, 2007).

Let x characterize the distribution of returns of a portfolio over N days. The right-tail N -day X -percent Value at Risk of the portfolio is then defined to be the value VaR such that:

$$P(x \leq VaR) = \frac{X}{100} \quad (1)$$

Likewise, the daily left-tail X -percent Value at Risk of the portfolio can be defined as the value VaR such that:

$$P(x \leq VaR) = 1 - \frac{X}{100} \quad (2)$$

2.2: Extreme Value Theory (EVT)

Tsay (2005) provides a framework for considering the distribution of the minimum order statistic. Let $\bar{x} = \{x_1, x_2, \dots, x_n\}$ be a collection of serially independent data points with common cumulative distribution function $F(x)$ and let $x_{(1)} = \min(x_1, x_2, \dots, x_n)$ be the minimum order statistic of the data set. The cumulative distribution of the minimum order statistic is given by:

$$F_{x_{(1)}}(x) = P(\min(x_1, x_2, \dots, x_n) \leq x) \quad (3)$$

$$F_{x_{(1)}}(x) = 1 - P(\min(x_1, x_2, \dots, x_n) > x) \quad (4)$$

$$F_{x_{(1)}}(x) = 1 - P(x_1 > x, x_2 > x, \dots, x_n > x) \quad (5)$$

$$F_{x_{(1)}}(x) = 1 - P(x_1 > x) \cdot P(x_2 > x) \cdot \dots \cdot P(x_n > x) \quad (6)$$

$$F_{x_{(1)}}(x) = 1 - \prod_{i=1}^n P(x_i > x) \quad (7)$$

$$F_{x_{(1)}}(x) = 1 - \prod_{i=1}^n [1 - F(x)] \quad (8)$$

$$F_{x_{(1)}}(x) = 1 - [1 - F(x)]^n \quad (9)$$

As n increases to infinity, this cumulative distribution function becomes degenerated in that $F_{x_{(1)}}(x) \rightarrow 0$ when $0 \leq F(x) < 1$ and $F_{x_{(1)}}(x) \rightarrow 1$ when $F(x) = 1$ and hence, has no practical value. In Extreme Value Theory, location series sequence $\{\beta_n\}$ and a scaling factors series sequence $\{\alpha_n : \alpha_n > 0\}$ are determined such that the distribution of $x_{(1)^*} = \frac{r_{(1)} - \beta_{(n)}}{\alpha_n}$ converges to a non-degenerate distribution as n goes to infinity. The distribution of the normalized minimum is given by:

$$F_{x_{(1)^*}}(x) = \begin{cases} 1 - \exp[-(1 + \xi \cdot x)^{1/\xi}] & \text{if } \xi \neq 0 \\ 1 - \exp[-\exp(x)] & \text{if } \xi = 0 \end{cases} \quad (10)$$

This distribution applies where $x < -1/\xi$ if $\xi < 0$ and for $x > -1/\xi$ if $\xi > 0$. When $\xi = 0$, a limit must be taken as $\xi \rightarrow 0$. The parameter ξ is often referred to as the shape parameter and its inverse $\alpha = -1/\xi$ is referred to as the tail index. This parameter governs the tail behavior of the limiting distribution.

The limiting distribution in (10) is called the Generalized Extreme Value (GEV) distribution for the minimum and encompasses three types of limiting distributions:

1) Gumbel Family ($\xi = 0$)

$$F_{x_{(1)^*}}(x) = 1 - \exp(-\exp(x)) \quad -\infty < x < \infty \quad (11)$$

2) Fréchet Family ($\xi < 0$)

$$F_{x_{(1^*)}}(x) = \begin{cases} 1 - \exp[-(1 + \xi \cdot x)^{1/\xi}] & \text{if } x < -1/\xi \\ 1 & \text{if } x \geq -1/\xi \end{cases} \quad (12)$$

3) Weibull Family ($\xi > 0$)

$$F_{x_{(1^*)}}(x) = \begin{cases} 1 - \exp[-(1 + \xi \cdot x)^{1/\xi}] & \text{if } x > -1/\xi \\ 0 & \text{if } x \leq -1/\xi \end{cases} \quad (13)$$

Although Tsay's (2005) framework provides a model for the minimum order statistic, the same theory also applies for the maximum order statistic $x_{(n)} = \max(x_1, x_2, \dots, x_n)$, which is the primary interest for this paper. In this case, the degenerate cumulative distribution function of the maximum order statistic would be described by:

$$F_{x_{(1)}}(x) = [F(x)]^n \quad (14)$$

The limiting Generalized Extreme Value Distribution is then described by:

$$F_{x_{(1^*)}}(x) = \begin{cases} \exp[-(1 + \xi \cdot x)^{-1/\xi}] & \text{if } \xi \neq 0 \\ \exp[-\exp(x)] & \text{if } \xi = 0 \end{cases} \quad (15)$$

3. Description of Statistical Model

3.1: EVT Parameter Estimation (“Block Maxima Method”)

In order to apply Extreme Value Theory to Value at Risk, one must first estimate the parameters of the GEV distribution (15) that govern the distribution of the maximum order statistic. These parameters include the location parameter α_n , scale parameter β_n , and shape parameter ξ for a given block size n . One plausible method of estimating these parameters is known as the block maxima method. In the block maxima method, a large data set is divided into several evenly sized subgroups. The maximum data point in each subgroup is then sampled. With this sample of maximum data points for each subgroup, maximum likelihood estimation is then used to determine a value for each parameter and fit the GEV distribution to these data points. Hence, the assumption is that the distribution of maximum order statistics in subgroups is similar to the distribution of the maximum order statistic for the entire group.

Tsay (2005) outlines a procedure for conducting the block maxima method: Let $\bar{x} = \{x_1, x_2, \dots, x_n\}$ be a set of data points. In the block maxima method, the original data set $\bar{x} = \{x_1, x_2, \dots, x_n\}$ is divided into g subgroups (“blocks”) of block size m : $\bar{x}_1 = \{x_1, x_2, \dots, x_m\}$, $\bar{x}_2 = \{x_{m+1}, x_{m+2}, \dots, x_{2m}\}$, \dots , $\bar{x}_g = \{x_{(g-1)m+1}, x_{(g-1)m+2}, \dots, x_n\}$. For sufficiently large m , the maximum of each subgroup should be distributed by the GEV distribution with the same parameters (for a large enough m , the block can be thought of as a representative, independent time series). Therefore, if the data points $\bar{Y} = \{Y_1, Y_2, \dots, Y_g\}$ are taken such that $Y_1 = \max(x_1, x_2, \dots, x_m)$, $Y_2 = \max(x_{m+1}, x_{m+2}, \dots, x_{2m})$, \dots , $Y_g = \max(x_{(g-1)m+1}, x_{(g-1)m+2}, \dots, x_n)$, then \bar{Y} should be a collection of data from a common GEV distribution. Using maximum likelihood estimation, the parameters in can be estimated with the data from \bar{Y} .

Although the block maxima method is a statistically reliable and plausible method of estimating the parameters of the GEV distribution, there a few criticisms have limited its use in EVT literature. One criticism is that large data sets are necessary. The block size m has to be large enough for the estimation to be meaningful, but if it is too large, then there

is a significant loss of data since fewer data points will be sampled. Another criticism is that the block maxima method is susceptible to volatility clustering, the phenomena that days of high volatility are followed by days of high volatility and days of low volatility are followed by low volatility. For example, a series of extreme events may be grouped together in a small time span due to high volatility but the block maxima method would only sample one of the events from the block. In this paper, both problems with the block maxima method are largely minimized. Since high-frequency returns are considered in this paper, the data set is sufficiently large that 10 years of data can produce enough data points for proper estimation. Furthermore, since high-frequency returns are standardized by dividing by their volatility, the effect of volatility clustering is removed. Other common methods of EVT estimation include forms of non-parametric estimation. However, these methods rely on qualitative and subjective techniques in estimating some parameters. Therefore, the block maxima method was used in this paper because its weaknesses have largely been addressed and because it can provide a purely statistical and quantitative estimation (Tsay, 2005).

3.2: Value at Risk Estimation

The value at risk can be estimated from the block maxima method by using the following relationship for block size m :

$$P(x_{(m)} \leq VaR) = P(\max(x_1, x_2, \dots, x_m) \leq VaR) = [P(x_i \leq VaR)]^m.$$

Therefore, to determine the right-tail X -percent Value at Risk, one would find the value of VaR where:

$$P(x_{(m)} \leq VaR) = \left(\frac{X}{100} \right)^m \quad (16)$$

The order statistic $x_{(m)}$ is assumed to be distributed by the GEV distribution.

3.3: Realized Variance

Given a set of high-frequency data where there are M ticks available for each day, let the variable $P_{t,j}$ be defined as the value of the portfolio on the j^{th} tick of day t . The j^{th} intraday log return on day t can then be defined as:

$$r_{t,j} = \log(P_{t,j}) - \log(P_{t,j-1}) \quad j = 2, 3, \dots, M \quad (17)$$

The realized variance over day t can then be computed as the sum of the squares of the high-frequency log returns over that day:

$$RV_t = \sum_{j=2}^M r_{t,j}^2 \quad (18)$$

Andersen and Bollerslev (1998) have shown that the realized variance measure converges in frequency to the integrated variance plus a discrete jump component, due to the theory of quadratic variation:

$$\lim_{M \rightarrow \infty} RV_t = \int_{t-1}^t \sigma^2(s) ds + \sum_{t-1 \leq s \leq t} \kappa^2(s) \quad (19)$$

Therefore, the realized variance metric can intuitively be used as an estimate for the variance over a day t . The realized volatility is defined to be the square root of realized variance. It should be noted that realized variance and realized volatility can be calculated over any time period and not just one day.

3.4: Standardizing by Volatility

Asset prices are typically modeled with the following standard stochastic differential equation:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \quad (20)$$

The variable $dp(t)$ represents the log-price movement of an asset at time t , $\mu(t)$ represents the time-varying drift component, $\sigma(t)$ represents the time-varying volatility component and $W(t)$ represents standard Brownian motion (Merton, 1971). At high-frequencies, the drift is considered small enough that it can effectively be considered zero: $\mu(t) = 0$. Therefore, at high-frequencies the stochastic differential equation can be represented as:

$$dp(t) = \sigma(t)dW(t), \quad (21)$$

Hence, by dividing by a metric for time-varying volatility, the standardized returns can be considered to be identically distributed:

$$r_t^S = \frac{r_t}{\sqrt{RV_t}} \quad (22)$$

This standardization procedure was demonstrated by Andersen et al. (2000), who determined that the overall distribution of standardized returns was approximately unconditionally normally distributed for thirty random selected liquid stocks from the Dow Jones Industrial Average Index.

4. Data

Minute-by-minute stock price data for several S&P 100 stocks were purchased from an online data vendor, price-data.com. This paper presents the results of an analysis of Citigroup (C) share price data from April 4, 1997 to January 7, 2009 (2,921 days) sampled every minute from 9:35 am to 3:59 pm. Although the stock exchange opens as early as 9:30 am, data was collected from 9:35 am and onwards to account for unusual behavior in early morning price data, resulting from several technical factors such as reactions to overnight news. To check the validity of the results, share prices were also analyzed for other stocks although their results were not presented in this paper. The time frame of these share price data are approximately the same as that of Citigroup. The results were presented for Citigroup because it is a representative large capitalization, highly liquid stock and because it had a strong response to the extreme market events in the fall of 2008, making it pertinent in risk analytics studies.

5. Statistical Methods

5.1: Overview of VaR Test

In the VaR test conducted for the paper, share prices from the first 1,000 days were used for the in-sample data. From the in-sample data, daily returns were determined by finding the log difference between the opening and closing prices for each day. These daily returns were then standardized by dividing each daily return by the realized volatility over that same day. Then, using the standardized daily returns data, the block maxima method (see section 3.1) was used to determine the parameters of the distribution.

For each confidence level, 97.5, 99.0, 99.5 and 99.9 percent, a value at risk was determined (See Section 3.2). Since this value at risk only applied to a distribution of a standardized returns, the value at risk had to be multiplied by a volatility metric in order to “un-standardize” the value at risk. In one test, the standardized value at risk was multiplied by the realized volatility on the 1001st day. In a second test, the standardized value at risk was multiplied by a forecasted realized volatility for the 1001st day. After being multiplied by a realized volatility, this new value at risk was referred to as the un-standardized value at risk.

The un-standardized value at risk was then compared to the actual daily return on the 1001st day. If the actual daily return exceeded the un-standardized value at risk, then the out-of-sample trial was recorded as a “break.” The number of breaks and the timing of the breaks were noted. The value at risk was then re-calculated using the first 1,001 days as the in-sample to calculate an un-standardized value at risk which was then compared to the daily return on the 1002nd day. This process repeated itself until all the days in the out-of-sample data were exhausted.

The number and timing of breaks were used to determine statistics that would evaluate the validity of the value at risk test and model. For example, if a 99.0 percent value at risk test was conducted, then breaks would be expected to occur in approximately 1.0 percent of the out-of-sample trials. Two statistical tests, binomial and Kupiec, were used to

evaluate the number of breaks in the test and one statistical test, Christofferson, was used to evaluate for the bunching of breaks in the test.

The test procedure was then repeated for time periods of less than one day. That is, rather than using just 1,000 days of data for the first in-sample period, daily returns over half-days were produced, resulting in 2,000 data points. The return and realized volatility were then determined over half-day periods and value at risk was computed and compared to forward half-day return (i.e. the first 2,000 half-day returns were used to compute a VaR for the 2,001th half-day). This process was repeated for quarter-day trials, and eighth-day trials. In all of these “intraday” value at risk tests, the first 1,000 days of data were always used for the first in-sample trial. Furthermore, in the intraday value at risk tests, the first version of the test was used, in which the realized volatility of the next out-of-sample was known. For intraday volatility tests, the intraday volatility was never forecasted because intraday volatility tends to produce unusual patterns and its dynamics are not well understood.

5.2: Computing Realized Volatility

One of the critical steps in the test is to standardize daily returns and un-standardize the value at risk by using realized volatility. As shown in Section 3.3, realized volatility can be computed from the sum of the squares of the intraday returns. The finer the interval in the intraday returns, the closer the realized volatility is to the actual integrated volatility.

Although price data was available as fine as 1-minute intervals, a larger time interval was used to calculate realized volatility. The reason for this was due to the presence of market microstructure noise. At very high frequencies, the sampled price does not reflect the fundamental value as market agents require longer time periods to accurately price financial assets. Furthermore, the existence of market frictions such as the presence of the bid-ask spread within the data, further mask the fundamental value (Bandi & Russell, 2005).

Therefore, a signature volatility plot (Figure 1) was created which plotted the average calculated daily realized volatility across the entire sample for each time interval. In the absence of market microstructure noise and in the presence of absolute market efficiency, the signature volatility plot should appear as a horizontal line since the sampling time interval should have no correlation with the computed realized volatility (Anderson et al., 1999). However, as shown in the plot, at very high-frequencies, the presence of market microstructure noise creates an upward bias in the computed average realized volatility and tends to smooth out more for less fine time intervals.

Therefore, a sampling interval of larger than 1 minute was necessary to avoid complications presented by market microstructure noise. However, if too large a time interval was chosen, then the benefits to using high-frequency data would be lost as realized volatility approximates integrated volatility at fine time intervals. Therefore, to balance between these two tradeoffs, a sampling interval of 10 minutes was used, as it lessened the effects of market microstructure noise but preserved information through a fine enough sampling frequency.

5.3: Sub-Sampling for Realized Volatility Computation

A “sub-sampling” procedure was developed for calculating realized volatility in order to minimize the loss of information from using a 10-minute interval and to smooth out the calculations. As an example, for a daily realized volatility, ten realized volatility measures were calculated and then averaged to determine the realized volatility over that day. Each measure had a different initial time step: 9:35 am, 9:36 am, ... 9:44 am. If there was a remaining intraday return that was shorter than 10-minutes, then it was scaled appropriately to be used in the realized volatility calculation.

5.4: Standardization of Daily Returns

Daily returns were standardized by their daily realized volatility in order to make the data independent and identically distributed. The model in Section 3.4 establishes that standardizing by realized volatility would make the data identically distributed. To test for independence, an autocorrelation plot (Figure 2) was constructed for the standardized

daily returns data. Figure 2 of Appendix A suggests that the standardized data are very weakly correlated and so the independent and identically distributed assumption is validated for standardized data.

5.5: Block Size in Block Maxima Method

In order to implement the block maxima method from section 3, a block size m has to be chosen for estimating the parameters of the GEV distribution. A large block size is required in order for the statistical estimation method to hold but if the block size is too large, then there are fewer data points for estimation. In all tests, a block size of 70 trials was used. The tests were repeated using other block sizes and there were no significant changes in the final results. Figure 3 shows the estimated shape parameter as a function of block size in the case of sixteenth-day trails. The figure shows that the parameter estimation tends to stabilize somewhat after block sizes of 20-30.

5.6: Forecasting Volatility (“HAR-RV Model”)

Volatility clustering has been a well-observed phenomenon observed in empirical studies on asset prices. Periods of high volatility tend to cluster together and likewise, similar trends occur for periods of low volatility. Because of this effect, predictive models of volatility have been developed since volatility in the near past history can provide information on volatility in the near future. That is, volatility has strong autocorrelation for shorter lags (Stock & Watson, 2007). One such predictive model is the heterogeneous autoregressive realized variance (HAR-RV) model developed by Corsi (2003) which regresses forward average realized volatility on average realized volatility for one day, one week and one month prior:

$$RV_{t,t+h} = \beta_0 + \beta_1 RV_{t-1,t} + \beta_2 RV_{t-5,t} + \beta_3 RV_{t-22,t} + \varepsilon_{t+1} \text{ .where } RV_{t,t+h} = h^{-1} \sum_{j=1}^h RV_{t+j} \quad (23)$$

Corsi demonstrates that the HAR-RV model tends to provide superior forecasts for one day ahead, one week ahead and two weeks ahead when compared to other standard predictive volatility models such as the GARCH, J.P. Morgan’s RiskMetrics, and AR(1) and AR(3) models of realized volatility (2003). The model that was used in the test was to forecast one-day ahead volatility ($h = 1$).

5.7: Statistical Methods for Evaluating VaR Model

During the VaR test, the VaR was calculated for one-day forward and then was compared with the actual daily return. If the actual daily return was larger than the estimated VaR, then a break was recorded. For an X -percent VaR test, it would be expected that breaks occur $(1-X)$ -percent of the time. Therefore, the number of breaks can be thought of as a binomial random variable with probability $p = (1 - X)/100$ over the number of out-of-sample trials n . Using a two-sided test with the null hypothesis that the number of breaks equals the expected value $n(1 - X)$, a p -value can be determined. This test was referred to in the paper as the “binomial” test. With the same general idea, Kupiec (1995) proposed a powerful two-test that for a valid VaR model, the statistic

$$-2 \ln[(1 - p)^{n-m} p^m] + 2 \ln[(1 - m/n)^{n-m} (m/n)^m] \quad (24)$$

should have a chi-square distribution with one degree of freedom where m is the number of breaks, n is the number of trials, and p is $(1 - X)/100$. For either the binomial or Kupiec statistic, a low p -value would indicate the number of breaks was much different than expected and thus the VaR model was inappropriate. Another issue to test the validity of a VaR model would be to test for bunching. A valid VaR model should have breaks spread relatively uniformly across the out-of-sample region. A test proposed by Christofferson (1998) indicates that the test statistic

$$-2 \ln[(1 - \pi)^{u_{00}+u_{10}} \pi^{u_{01}+u_{11}}] + 2 \ln[(1 - \pi_{01})^{u_{00}} \pi_{01}^{u_{01}} (1 - \pi_{11})^{u_{10}} \pi_{11}^{u_{11}}] \quad (25)$$

$$\pi = \frac{u_{01} + u_{11}}{u_{00} + u_{01} + u_{10} + u_{11}} \quad \pi_{01} = \frac{u_{01}}{u_{00} + u_{01}} \quad \pi_{11} = \frac{u_{11}}{u_{10} + u_{11}}$$

should have a chi-square distribution with one degree of freedom if there is no bunching. The variable u_{ij} is defined as number of observations in which a day moves from state i and j where $i, j = 0$ indicates a day without a break and $i, j = 1$ indicates a day with a break. These statistics were calculated to test the validity of the VaR model. In addition to these statistics, the average calculated VaR across the out-of-sample period was also recorded (Hull, 2007).

6. Description of Findings

Table 1 displays the statistical results of a VaR test for Citigroup's stock in which historical standardized 1-day returns and one day ahead realized volatility to determine one day ahead VaR. The results show that except for the 97.5% VaR test on the left tail, the procedure described above in the methods section results in a relatively sound VaR model. However, since the forward realized volatility is known in the test, this reveals nothing about the predictive nature of the VaR model. This test only suggests that EVT provides a relatively good estimation of VaR, assuming that the realized volatility is known. Therefore, in order to apply the VaR model developed above in a predictive sense, forecasting future volatility is an integral component.

Table 2 repeats the same test from Table 1, but this time the forward volatility is forecasted using the HAR-RV model. At first glance, the model proposed by this paper with forecasted volatility appears valid for the right tail. However, for the left tail, this model appears to largely underestimate the VaR since more breaks occur than expected and this is especially true for the higher quantiles of 0.5% and 0.1%. Since the procedure appeared valid for known volatility as shown in Table 1, these results might suggest that the problem with the prediction lies with improper volatility estimation by the HAR-RV model. Considering that the out-of-sample region included data from the highly volatility period of fall 2008, it is likely that considerable errors were made in one-day ahead volatility forecasting. The volatility estimation appears to have been more of an issue for the right tail than for the left tail. However, upon running the identical test on several other stocks, it is apparent that this is an issue specifically directed towards Citigroup (C) stock. In other stocks such as Goldman Sachs (GS) and Wal-Mart (WMT), forecasting problems appear to be spread relatively evenly between both left and right tails. One further observation is that the average VaR is higher for the test with forecasted forward volatility than the test with known forward volatility, once again highlighting the errors due to forward volatility estimation.

One problem with the tests conducted so far is that the magnitude of the numbers involved has been relatively small. For example in Tables 1 and 2, many tests have less than 20 breaks. Therefore, the tests could have high sensitivity to the number of recorded

breaks. For example, if no breaks were recorded in the 99.9% VaR test, the model could be deemed valid. If three or more breaks were recorded in the 99.9% VaR, the model could be deemed invalid. This is especially troublesome since the break is binary variable, and a break could be determined if it the daily return is only slightly above the VaR or be left out if it is only slightly less than the VaR. Therefore, one method of working around this issue would be to consider VaR tests on intraday returns rather than daily returns. That is, to compute and standardize half-day returns and then to treat each half-day as a trial under the original VaR test. For half-day returns, this would generate double the amount of trials and for quarter-day returns, quadruple the number of trials. With more trials, there would also be more expected breaks, making the VaR tests more likely to determine the true validity of the proposed model. Furthermore, this would also test the concept of an intraday VaR in which capital allocation could be readjusted during the day as opposed to only reallocating capital overnight.

Table 3 shows the results of the VaR test in which the days are divided into 1, 2, 4 and 8 parts and in which the forward volatility is known. The results show that when the volatility is known, the VaR model outlined in the paper is valid, even with returns at higher frequencies than daily returns. The VaR model test is also more robust for high-frequency returns. For example, the 99.9% VaR test requires around 8 breaks for quarter-day returns and 15 breaks for eighth-day returns. Since there are more data points, the magnitudes of the numbers involved are larger and so these tests provide a better way of measuring the validity of the tests.

This also substantiates the concept of an intraday VaR. In right-tail 99.5% VaR in Table 4, one can simply observe that by daily returns one would on average expect a 3.81% VaR but by half-day returns on average would expect 2.53%. One can interpret this as follows: a risk manager can allocate capital separately for the morning and the afternoon and average 2.53% VaR. On the other hand by just setting the capital allocation once daily, the morning and afternoon VaR would average to 3.81%. Since capital allocation is proportional to the level of VaR, one could imagine an intraday VaR as a method for the risk manager to more efficiently allocate capital.

Table 4 repeats one of the tests from Table 3, only this time using a normal distribution instead of the GEV distribution from EVT. For daily returns, the normal

distribution cannot be immediately rejected. However, when considering half-day and higher-frequency intraday returns it becomes clear that the normal distribution does not appear to be a good fit for the data. This is in contrast with the results in Table 3, which once again shows the superiority of the EVT method over using the normal distribution. Since this contrast was only noticed by using high-frequency data, Table 4 demonstrates that high-frequency data can be used to better test the robustness of VaR models.

Given the increase in size of the data set due to the use of high-frequency data, improved parametric estimation can be considered. The main parameter of importance for the GEV distribution is the shape parameter ξ since it is independent of block size and has the largest role in determining the overall shape of the GEV distribution. For larger block sizes, it is expected that the shape parameter ξ converges to a particular value. For daily returns, this result doesn't appear readily apparent because for large block sizes, there are less data points and so the estimation of the shape parameter is less precise. However, due to the large data set generated by intraday returns, the shape parameter may appear to converge for large block sizes. Figure 3 shows the estimation of the shape parameter as a function of block size for 16 returns per day. The plot appears to show slight convergence for higher block sizes, which could in turn lead to a more precise estimation of the shape parameter ξ . However, it is unclear as to whether the estimated shape parameter for 16 returns per day provides any information about the shape parameter for longer time periods such as one day returns.

7. Conclusion

In the majority of the literature regarding the application of EVT to financial data for VaR models, only closing price and their respective log returns are used for EVT estimation. However, since the log returns are not independent and identically distributed, this would provide an improper estimation of the VaR. This paper illustrates a procedure for estimating EVT by first standardizing the log returns by the realized volatility, as estimated with high-frequency data, and so by making the returns data independent and identically distributed prior to EVT estimation. The results empirically show that if the forward volatility is known, then this procedure can provide a valid VaR measure. By using the HAR-RV model to predict forward volatility, the predictive model doesn't perform as well, suggesting that proper volatility prediction is an integral component for the application of this procedure. The paper also demonstrates that the outlined method applies to intraday VaR models.

There are several areas of research that originate from this paper. First of all, this paper demonstrates that the proposed model appears consistent as long as volatility can be forecasted accurately. Therefore, the focus of further research should be on proper volatility forecasting. Although this paper used daily realized volatility forecasting, it is extremely difficult to forecast one-day ahead volatility. Time horizons of longer than one-day should be explored as they might provide smoother volatility forecasting. Another area of research would be to investigate parameter estimation at high-frequencies. There might be information from high-frequency returns that could be used to estimate parameters for daily GEV distributions or to forecast value at risk.

A. Tables and Figures

Table 1

Daily VaR Test for Citigroup Stock with Known Forward Realized Volatility

Left Tail

VaR Level	Expected Breaks	Actual Breaks	Break Ratio	Binomial p -value	Kupiec p -value	Christoff. p -value	Average VaR
97.5 %	48	64	3.33 %	0.0208	0.0262	0.3731	-2.97 %
99.0 %	19	20	1.04 %	0.7415	0.8572	0.5164	-3.42 %
99.5 %	10	12	0.62 %	0.3441	0.4559	0.6977	-3.69 %
99.9 %	2	2	0.10 %	0.6039	0.9548	0.9485	-4.17 %

Right Tail

VaR Level	Expected Breaks	Actual Breaks	Break Ratio	Binomial p -value	Kupiec p -value	Christoff. p -value	Average VaR
97.5 %	48	51	2.65 %	0.5996	0.6668	0.0952	2.87 %
99.0 %	19	19	0.99 %	0.9172	0.9615	0.5377	3.45 %
99.5 %	10	11	0.57 %	0.5178	0.6592	0.7218	3.81 %
99.9 %	2	1	0.05 %	0.8554	0.4638	0.9742	4.50 %

Table 1 shows that when one day ahead realized volatility is known, the model described in the paper performs relatively well. In the above, VaR tests are conducted at four different VaR levels for both the left and right tails. The VaR tests use 2,921 days of data and start testing for breaks on the 1001st day. The number and timing of the breaks are recorded and used to compute statistics to determine the validity of the VaR model. The only unexpected results occur at the 97.5% level but EVT methodology is generally designed for very high VaR levels such as 99.5% and 99.9%. Therefore, this table suggests that if volatility can be forecasted successfully, the model described in the paper can serve as a reasonable VaR model.

Table 2

Daily VaR Test for Citigroup Stock with Forecasted Realized Volatility

Left Tail

VaR Level	Expected Breaks	Actual Breaks	Break Ratio	Binomial <i>p</i> -value	Kupiec <i>p</i> -value	Christoff. <i>p</i> -value	Average VaR
97.5 %	48	49	2.55 %	0.8122	0.8871	0.0002	-3.34 %
99.0 %	19	23	1.20 %	0.3237	0.3993	0.0292	-3.84 %
99.5 %	10	19	0.99 %	0.0043	0.0074	0.0128	-4.14 %
99.9 %	2	12	0.62 %	0.0000	0.0000	0.0622	-4.68 %

Right Tail

VaR Level	Expected Breaks	Actual Breaks	Break Ratio	Binomial <i>p</i> -value	Kupiec <i>p</i> -value	Christoff. <i>p</i> -value	Average VaR
97.5 %	48	37	1.93 %	0.1154	0.0934	0.1997	3.22 %
99.0 %	19	17	0.88 %	0.7190	0.6052	0.5816	3.88 %
99.5 %	10	13	0.68 %	0.2158	0.2975	0.6738	4.29 %
99.9 %	2	2	0.10 %	0.6039	0.9548	0.9485	5.06 %

Table 2 repeats the tests described in Table 1, only this time using the HAR-RV model to predict one day ahead realized volatility. As would be expected, the results shown here are not as desirable as those in the previous table. Although the complications appear concentrated on the left tail, tests on other stocks such as Goldman Sachs and Wal-Mart show that complications are not limited to either tail. Therefore, in order for the model described in the paper to have predictive capabilities, the focus should be on developing proper volatility forecasting methods.

Table 3

Intraday VaR Test for Citigroup Stock with Known Forward Realized Volatility

Left Tail – 99.5% VaR

Returns Per Day	OOS Trials	Expected Breaks	Actual Breaks	Break Ratio	Binomial p -value	Kupiec p -value	Christof. p -value	Average VaR
1	1921	10	12	0.62 %	0.3441	0.4559	0.6976	-3.69 %
2	3842	19	22	0.57 %	0.4416	0.5329	0.6146	-2.58 %
4	7684	38	46	0.60 %	0.1968	0.2345	0.9383	-1.68 %
8	15368	77	79	0.51 %	0.7480	0.8058	0.3662	-1.06 %

Right Tail – 99.5% VaR

Returns Per Day	OOS Trials	Expected Breaks	Actual Breaks	Break Ratio	Binomial p -value	Kupiec p -value	Christof. p -value	Average VaR
1	1921	10	11	0.57 %	0.5178	0.6592	0.7218	3.81 %
2	3842	19	23	0.60 %	0.3248	0.4005	0.5986	2.53 %
4	7684	38	43	0.56 %	0.4061	0.4674	0.2440	1.65 %
8	15368	77	78	0.51 %	0.8352	0.8947	0.3723	1.04 %

Left Tail – 99.9% VaR

Returns Per Day	OOS Trials	Expected Breaks	Actual Breaks	Break Ratio	Binomial p -value	Kupiec p -value	Christof. p -value	Average VaR
1	1921	2	2	0.10 %	0.6039	0.9548	0.9485	-4.17 %
2	3842	4	3	0.08 %	0.9297	0.6548	0.9454	-3.05 %
4	7684	8	9	0.12 %	0.4899	0.6438	0.8845	-1.89 %
8	15368	15	15	0.10 %	0.9391	0.9249	0.8641	-1.15 %

Right Tail – 99.9% VaR

Returns Per Day	OOS Trials	Expected Breaks	Actual Breaks	Break Ratio	Binomial p -value	Kupiec p -value	Christof. p -value	Average VaR
1	1921	2	1	0.05 %	0.8554	0.4638	0.9742	4.50 %
2	3842	4	4	0.10 %	0.6806	0.9362	0.9272	2.97 %
4	7684	8	6	0.08 %	0.7067	0.5272	0.9229	1.85 %
8	15368	15	16	0.10 %	0.7431	0.8727	0.8551	3.34 %

Table 3 demonstrates the viability of an intraday VaR model by using high-frequency data. Rather than simply determining a VaR model for daily returns, it should be possible to determine a VaR model for half-day returns and for quarter-day returns. The above demonstrate the results of 99.5% and 99.9% VaR tests for both the right and left tails when one day ahead realized volatility is known. For each day, the data are sliced up into multiple trials. For example, “2 returns per day” signifies that each half-day is treated as a separate trial. The above results show that the model described in the paper is a valid VaR model even when days are divided into multiple trials. It also suggests that since more trials are tested when more divisions are made, intraday tests can provide a more robust method of testing VaR models.

Table 4

Intraday VaR Test for Citigroup Stock with Known Forward Realized Volatility
and Normal Distribution

Right Tail – 99.5% VaR

Returns Per Day	OOS Trials	Expected Breaks	Actual Breaks	Break Ratio	Binomial <i>p</i> -value	Kupiec <i>p</i> -value	Christof. <i>p</i> -value	Average VaR
1	1921	10	8	0.42 %	0.7570	0.5929	0.7958	3.92 %
2	3842	19	11	0.29 %	0.0622	0.0411	0.8015	2.74 %
4	7684	38	6	0.08 %	0.0000	0.0000	0.9229	1.85 %
8	15368	77	0	0.00 %	0.0000	-	-	1.27 %

Table 4 once again demonstrates the superiority of the EVT methodology with respect to the normal distribution method. The test from Table 3 is repeated for the 99.5% VaR with the right tail, but this time the normal distribution is used instead of the EVT methodology. While the results of normal distribution and EVT VaR models appear similar when only considering daily returns, it becomes clear that the EVT VaR model is far superior when considering VaR models for intraday returns.

Figure 1

Signature Realized Volatility Plot for Citigroup Stock

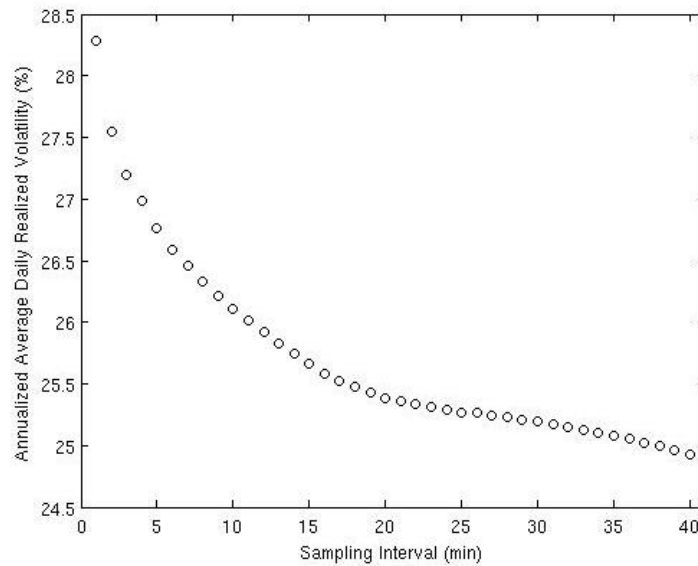


Figure 1 shows the average daily computed realized volatility as a function of the sampling interval. Under ideal circumstances, the plot would appear as a straight line since the estimated volatility should be independent of the sampling interval. However, due to the presence of market microstructure noise, using smaller sampling intervals result in much larger realized volatility values than would be expected. Therefore, a higher time interval would be desirable to mitigate the effects of microstructure noise at high frequencies. However, if too large an interval is used, then much of the data is lost due to infrequent sampling. To balance these two tradeoffs, a time interval of 10 was chosen to lessen the effects of market microstructure noise without a loss of too much information.

Figure 2

Autocorrelation of Standardized Daily Returns for Citigroup Stock

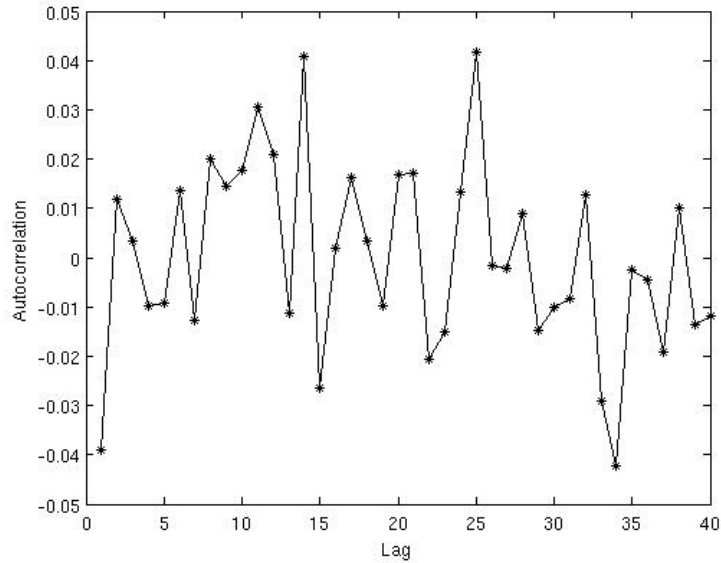


Figure 2 provides a correlogram of daily log returns of Citigroup stock after the returns have been divided by their corresponding daily realized volatility. The purpose of standardizing the returns by their volatility was to make the data appear more independent and identically distributed for EVT methodology. Since the magnitude of the autocorrelations are below 0.05 for the lags considered, this figure suggests that standardized data are weakly correlated and hence can be treated as though they are an independent random sample.

Figure 3

Shape Parameter Estimation for Citigroup Stock with 16 Returns Per Day

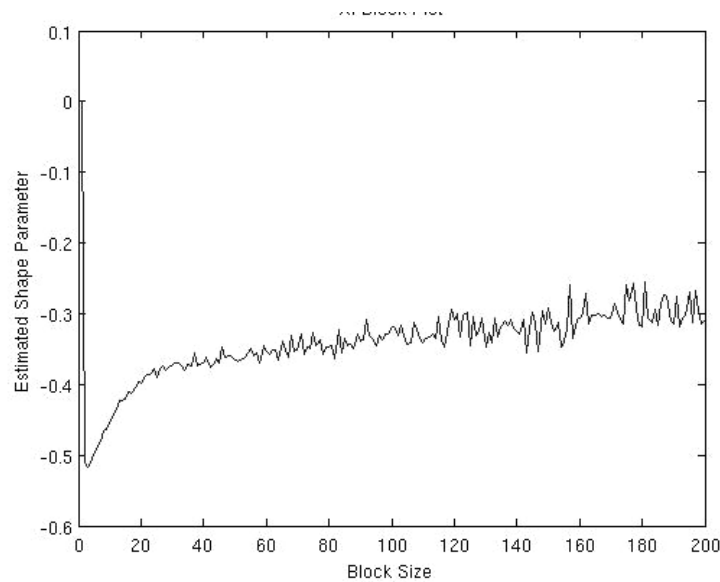


Figure 3 shows using high-frequency data could result in better parameter estimation. The primary estimated variable is the shape parameter ξ and is independent of block size. Large block sizes are expected to provide a better estimation of the parameter. However, the larger the block size, the less data points are sampled for parameter estimation. Therefore, in order to use block maxima estimations with large block sizes, a large data set is required. By obtaining 16 returns a day from 2,921 days of Citigroup stock price data, the data set increases to 46,736 total returns.

In the above graph, the shape parameter appears to be slightly converging as block size grows larger. Such a pattern as above cannot be produced for lower-frequency returns such as daily or half-day returns. The figure above suggests high-frequency returns can provide better parameter estimation which in turn has the potential to provide superior VaR models.

B. References

- Andersen, T.G. & Bollerslev, T. (1998). "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review*, 39, 885-905.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., & Ebens, H. (2001). "The Distribution of Stock Return Volatility." *Journal of Financial Economics*, 61, 43-76.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., & Labys, P. (1999). "(Understanding, Optimizing, Using, and Forecasting) Realized Volatility and Correlation." Manuscript, Northwestern University, Duke University and University of Pennsylvania.
- Bandi, F.M. & Russell, J.R. (2008). "Microstructure noise, realized variance and optimal sampling." *Review of Economic Studies*, 2008, 75, 339-369.
- Christofferson, P.F. (1998). "Evaluating Interval Forecasts." *International Economic Review*, 39, 841-862.
- Cont, R. (2001). "Empirical properties of asset returns: stylized facts and statistical issues." *Quantitative Finance* 1, 223-236.
- Corsi, F. (2003). "A simple long memory model of realized volatility." Unpublished manuscript, University of Southern Switzerland.
- Cotter, J. & Longin, F. (2004). *Margin Setting with High-Frequency Data*. Unpublished. Retrieved April 29, 2009 from <http://mpa.ub.uni-muenchen.de/3528/>
- Diebold, F.X., Schuermann, T., & Stroughair, J.D (2000). "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management." *Journal of Risk Finance* 1, 2, 30-35
- Gencay, R., Selçuk F. (2004). Extreme Value Theory and Value-at-Risk Relative Performance in Emerging Markets . *International Journal of Forecasting*, 20, 287-303.
- Gencay, R., Selçuk F., & Ulugülyağci A. (2001). EVIM: Software Package for Extreme Value Analysis in MATLAB. *Studies in Nonlinear Dynamics and Econometrics*, 5, 3, 214-240.
- Hull, J.C. (2007). *Risk Management and Financial Institutions*. Upper Saddle River, New Jersey: Pearson Education.

- Ho, L., Burridge, P., Cadle, J., & Theobald, M. (2000). "Value-at-risk: Applying the extreme value approach to Asian markets in the recent financial turmoil." *Pacific Basin Finance Journal*, 8, 249-275.
- Kupiec, P.H. (1995). "Techniques for verifying the accuracy of risk measurement models." *The Journal of Derivatives*, 3, 2, 73-84.
- McNeil, A. (1999). *Extreme Value Theory for Risk Managers*. Unpublished manuscript, Department of Mathematics, ETH Zentrum.
- Stock, J.H., & Watson, M.W. (2007). *Introduction to Econometrics*, 2nd ed.. Boston: Addison Wesley.
- Tsay, R.S. (2005). Extreme Values, Quantile Estimation and Value at Risk. (2nd Ed.), *Analysis of Financial Time Series* (pp. 287-316). Hoboken, New Jersey: John Wiley & Sons, Inc.