I. Introduction

Financial risks can be broadly classified into several categories, namely market risk, credit risk, liquidity risk, operational risk, and legal risk [1]. *Market risk* is the risk of loss arising from changes in the value of tradable or traded assets. *Credit risk* is the risk of loss due to the failure of the counterparty to pay the promised obligation. *Liquidity risk* is the risk of loss arising from the inability either to meet payment obligations (funding liquidity risk) or to liquidate positions with little price impact (asset liquidity risk). *Operational risk* is the risk of loss caused by inadequate or failed internal processes, people and systems, or external events. *Legal risk* is the risk of loss arising from uncertainty about the enforceability of contracts.

VaR is one of the most important and widely used risk management statistics. It measures the maximum expected loss in the value of a portfolio of assets over a target horizon, subject to a specified confidence level [2]. Since the publication of the market-risk-management system RiskMetrics of JPMorgan in 1994, value at risk (VaR) has gained increasing acceptance and now be consider as industry’s standard tool to measure market risk.

In our work, we compared different approaches to computing VaR for a delta hedged portfolio that contains stocks and call options. We investigated the issue of 1-day vs. 10-day VaR measurement, the effect of historical data window size, and the accuracy of VaR models in the 2008 financial crisis.

II. Methods

2.1 Delta hedge portfolio

A delta hedge is a simple type of hedge that is widely used by derivative dealers to reduce or eliminate a portfolio’s exposure to the underlying asset. The dealer calculates the portfolio’s delta with respect to the underlying asset and then adds an offsetting position to make the portfolio’s delta zero.

For example, consider the hedging portfolio with value \( P(t) \), we wrote an European call option with strike price \( K \) and maturity \( T \) at time \( t = 0 \). In order to hedge this short position, we bought \( \alpha(0) \) shares stock, which requires borrowing risk-free money of \( \alpha(0)S(0) \) from the bank. \( \alpha(t) \) is the delta, the
partial derivative of the option’s fair value under Black-Sholes option pricing model with respect to the price of the underlying security at time $t, S(t)$.

In summary, at $t = 0$,

$$P(0) = 0 = -V(0) + \alpha(0)S(0) + B(0)$$

$$\alpha(0) = \frac{\partial}{\partial s} c(S(0), T, K, \sigma)$$

$$B(0) = V(0) - \alpha(0)S(0)$$

Due to the limited exposure to the market, we were able to re-balance our hedged portfolio at discrete time $t = t_i$, we need to:

1. Calculate the new hedge: $\alpha(t_i) = V_S(S(t_i), t_i)$,

2. Adjust the stock position, which costs $S(t_i)(\alpha(t_i) - \alpha(t_{i-1}))$

3. Update bank account: $B(t_i) = e^{rt_i}B(t_{i-1}) - S(t_i)(\alpha(t_i) - \alpha(t_{i-1}))$

4. Finally, the value of the portfolio becomes: $P(t_i) = -V(t_i) + \alpha(t_i)S(t_i) + B(t_i)$

This hedge can maintain the same portfolio value for small changes in underlying stock price, or being balanced at infinitesimal time intervals. In fact, we would like to hedge as infrequently as possible, since in real, there are transaction costs and etc.

2.2 Modeling: Historical VaR

2.2.1 Historical VaR
Historical VaR is the most straightforward VaR calculation method. It treats all historical observations equal, and simply read the percentile off the historical returns of the portfolio. The advantage of historical VaR calculation is that it’s very easy to calculate and understand, and there is no explicit underlying model. The disadvantage is that it assumes independent returns and there is no unconditional covariance structure.

2.2.2 Delta normal VaR
In delta normal VaR calculation, it assumes change in stock $\Delta S$ follows $N(0, \sigma^2)$. Thus the change in option price $\Delta V$ can be approximated as the product of delta and $\Delta S$, hence can be seen as normal distributed random variable. This approach is also relatively simple to evaluate. The drawback of this approach is that empirical financial returns exhibit fat tail or left kurtosis in reality and the approximation here is only first-order.
2.3 Modeling: Monte-Carlo VAR

2.3.1 Semi-parametric Model
There are various ways of modeling the stock return series and generate new samples/predictions to calculate VAR. It is obvious that we need to heavily rely on historical data to provide VAR predictions. However, although being straightforward and model-free, the historical VAR has major drawbacks, one of which is that it assumes independent returns. In reality, returns of different stocks can be highly correlated due to factors like being in the same industry/sector. To improve the VAR estimation, correlation structure of multiple time series must be taken into account properly. Starting from the historical VAR, we want to first explore the potential of semi-parametric models, in which we don’t pose too much assumption on the stock returns, but still want to take full advantage of the empirical data. Bootstrapping can be a promising for a completely nonparametric approach. It is well known that it can be used to reproduce samples from a single time series. Even when the autocorrelation within the time series becomes significant, order statistics can potentially be integrated into the bootstrapping procedure to retain the autocorrelation structure. But in our case, a much more complicated bootstrapping scheme is necessary to deal with cross correlations of multiple time series.

To avoid going into the complications of developing a brand new bootstrapping scheme, we instead use the power of Cholesky decomposition. From the historical returns of multiple stocks, we estimate the empirical covariance matrix $\Sigma$ to which a Cholesky decomposition can be applied. Let $Z_1, Z_2, ..., Z_n$ be normal variables with zero mean and unit variance, by Cholesky decomposition, we can generate new multivariate normal samples $Z'_1, Z'_2, ..., Z'_n$ with zeros mean and covariance matrix $\Sigma$. An example with $n = 3$ is shown below.

$$
\Sigma = LL^T = \begin{bmatrix}
L_{11} & 0 & 0 \\
L_{21} & L_{22} & 0 \\
L_{31} & L_{32} & L_{33}
\end{bmatrix} \begin{bmatrix}
L_{11} & L_{12} & L_{13} \\
0 & L_{22} & L_{23} \\
0 & 0 & L_{33}
\end{bmatrix}, \quad Z' = \begin{bmatrix}
L_{11} & L_{12} & L_{13} \\
0 & L_{22} & L_{23} \\
0 & 0 & L_{33}
\end{bmatrix} \begin{bmatrix}
Z_1 \\
Z_2 \\
Z_3
\end{bmatrix}
$$

Now the covariance between different return series is retained by Cholesky decomposition. The next step will be converting the multivariate normal samples to samples that follow the empirical distributions. This is simply done by matching the quantile of a normal sample to the same quantile of the empirical CDF of the corresponding stock returns. The newly generated samples, individually, follows the empirical distribution of a certain stock return series, while collectively they are correlated in a way that is enforced by the historical covariance structure.

In our semi-parametric approach, there is no model fitting involved so that we minimized the impact of any model misspecifications. However, by using Cholesky decomposition and generating correlated normal samples, we have inherently assumed the returns series to be multivariate Gaussian. Even though we eventually make the marginal distribution of the generated samples aligned with the historical distribution, the covariance structure is still preserved as if it is Gaussian covariance. That is why we consider the model semi-parametric instead of a completely non-parametric one.
2.3.2 Parametric Models

A. Parametric Model I: Geometric Brownian Motion (GBM) and jump-diffusion

Geometric Brownian Motion (GBM) is widely used to model the stock price behavior and is the foundation of the Black-Scholes model. Under risk neutral measure, assuming constant risk-free rate and volatility, the stock price evolves according to the following stochastic differential equation and an analytical solution is available explicitly.

\[ dS_t = rS_t dt + \sigma S_t dW_t \]

\[ S_t = S_0 \exp \left( \left( r - \frac{\sigma^2}{2} \right) t + \sigma W_t \right) \]

Being a systematic model itself, GBM finds it hard to model the behavior of real stock prices. In the real market, stock prices are driven by a lot of forces which cannot be fully captured by GBM. Using GBM to simulate stock prices will result in large bias due to model misspecifications. However, it is very attractive that the correlated Brownian motion is easy to simulate and therefore the correlations between different stock prices can be taken care of automatically when we simulate price series for multiple stocks. In general, we generate Monte-Carlo samples in which each sample consists of \( N \) correlated stock prices driven by \( N \) correlated Brownian motions. Similar to what we did in the semi-parametric model, Cholesky decomposition is conducted on the historical sample covariance matrix. After the multivariate normal samples are generated, we simply make \( W_{t,i} = \sqrt{\tau} Z_i, i = 1,2,\ldots,N \), and the corresponding stock price samples can be obtained.

The major drawback of this model, of course, is the misspecification that fundamentally limits the performance of VAR prediction. No fitting is involved and so the historical data does not play a big role here. But there is still something we can do to make the model more realistic for true return series. One possible choice is to use stochastic volatilities, which will result in more randomized behavior of stock prices, and hopefully some large moves. Another model that can be an improvement over GBM is the jump-diffusion model. The prices simulated by GBM represent the more ‘general’ behavior and thus are in serious lack of jumps. A jump-diffusion model will take care of the large moves and result in more price samples on the tails, which are essential in VAR calculations. In this paper, however, we did not include results for these models because the stochastic volatility and tail distributions will be discussed in the second parametric model which involves extreme value theory. The GBM model, serving as a comparison, is mainly used to show the impact on the VAR prediction when the model itself is misspecified.

B. Parametric Model I: Extreme Value Theory (EVT)

GBM model showed us the extent of performance when the model is misspecified. However, another obstacle we only briefly mentioned is how we can generate enough price/return samples from the upper and lower tails even with a better model. This becomes extremely important if we want to have a robust VAR that can take the extreme circumstances, such as the crisis in 2008, into account. The problem is, however, we do not have sufficient historical data in the tails on both ends. That is where the EVT comes into play. We want to find an appropriate model that can fit well to the rare tail events and generate new samples in the tails from the fitted model.
Before going into our model, we want to introduce the student’s t-copula that is used to deal with the correlation between historical returns. Copula itself is a joint distribution function that describes the dependence structure between random variables.

Consider random vector \((X_1, X_2, \ldots, X_n)\) for which the marginal CDFs are continuous, we have

\[
(U_1, U_2, \ldots, U_n) = (F_1(X_1), F_2(X_2), \ldots, F_n(X_n))
\]

The copula of \((X_1, X_2, \ldots, X_n)\) is defined by

\[
C(X_1, X_2, \ldots, X_n) = P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_n \leq u_n)
\]

\(C\) is usually a multivariate distribution function that contains the dependence structure. Sample from the joint distribution function \(C\) to generate samples of \((U_1, U_2, \ldots, U_n)\) we can get

\[
(X_1, X_2, \ldots, X_n) = (F^{-1}_1(U_1), F^{-1}_2(U_2), \ldots, F^{-1}_n(U_n))
\]

Compared with the Cholesky decomposition described in the previous section, copula is a more refined way to deal with dependence structure by distribution fitting.

In our parametric model, we proceed in three major steps.

**Step 1**: Fit marginal CDF to the filtered residual of each historical return series

In this step, we first fit an asymmetric GARCH model to historical volatilities and an AR(1) to the returns. The asymmetric GARCH model is formulated as

\[
\begin{align*}
    r(t) &= c + \theta r(t-1) + \epsilon(t) \\
    \sigma^2(t) &= \kappa + \alpha \sigma^2(t-1) + \phi \epsilon^2(t-1) + \psi[\epsilon(t-1) < 0] \epsilon^2(t-1)
\end{align*}
\]

The normalized residual \(\epsilon(t)/\sigma(t)\) is assumed i.i.d. t-distributed.

After that, we apply Gaussian kernel smoothing to the residual and fit a Generalized Pareto Distribution (GPD) to both end of its tails. The GPD gives

\[
f(x | k, \sigma, \theta) = \left(\frac{1}{\sigma}(1 + k \frac{x - \theta}{\sigma})\right)^{-\frac{1}{k}}
\]

GPD is a right-skewed distribution in which \(k\) is the shape parameter. Larger \(k\) values lead to fatter tails.

**Step 2**: Fit student’s t copula and generate samples from the estimated copula.

In this step, we first estimate the degrees of freedom and linear correlation matrix of the t-copula with the normalized residuals using maximum log-likelihood [3]. The t-copula is given by

\[
C^t_{\nu, \rho}(u) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_d} \frac{\Gamma\left(\frac{\nu + d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu^{(d/2)}|P|} \left(1 + \frac{x^TP^{-1}x}{\nu}\right)^{-\frac{\nu + d}{2}} dx
\]
We then generate correlated uniform random variables by sampling the fitted t-copula, and transform them back to normalized return residuals from the inversion of the fitted empirical CDF.

Finally, we conduct GARCH simulation to reproduce return samples from the residual generated in the previous step.

**Step 3:** Aggregate portfolio values and read VAR.

In this final step, we only need to use the generated price/return samples to calculate sample portfolio returns and get the aggregated distribution. And then the remaining work is just the read the VAR from the distribution, as simple as it can be.

In this parametric model, stochastic volatility is integrated into the GARCH model, and GPD takes care of the tail distributions and ensures us to have sufficient samples from both ends of the tails. Therefore, it is expected to provide better VAR predictions over the GBM and semi-parametric models we described in the previous sections.

## 2.4 Backtesting

In order to gauge the performance of our VaR calculations, we backtested our models using historical stock data from 2004 to the present.

We collected historical stock returns from Bloomberg for 26 stocks since 1992 (tickers are AAPL, ABT, AEP, AMGN, APA, APC, BA, BAC, BAX, BHI, BK, BMY, MSFT, NKE, ORCL, OXY, PEP, PFE, XLNX, XRAY, VRTX, WFM, SYMC, TEVA, ROST, SBUX). These stocks were chosen with maximum positive or negative variations, as well as least variation so that this ensures extreme cases in two tails.

We included the 3-month at-the-money (ATM) call option for each stock in our portfolio, and option prices were calculated using implied volatility since 2004 and the Black-Scholes formula assuming a 3% risk-free rate. We backtested for periods of 250, 500, 750, 1000, and 1700 trading days; a 1700 day run ends in mid-2012. Overall, this is a rigorous testing period spanning almost four years before and after the 2008 financial crisis.

Ideally, a nominally 95% VaR model over some time period should see violations close to 5% of the time. For a financial institution, a conservative model where the actual probability of a violation is significantly lower than 5% can tie up too much money in reserves whereas an aggressive model where the actual probability of a violation is significantly higher than 5% can lead to excessive losses. The predicted violation frequency should be accurate, not conservative or aggressive.

In addition, the realized violations, when they occur, should never cluster or disperse too much. Even if our 95% VaR model has good unconditional coverage – when the realized probability of a violation is close to 5% – the model may not be particularly useful if all of the violations happen within the same week or month; the violations should occur independently of each other.

As such, we run two statistical tests during backtesting in order to measure the unconditional coverage and the independence of violations: these are the Kupiec for the former and the Ljung-Box test for the
latter[4]. In the Kupiec test the null hypothesis states that the probability of an actual violation equals the nominal percentage (100% - X% VaR). In the Ljung-Box test, violations over a period of time are a sequence of 1’s and 0’s; the null hypothesis states that the autocorrelation between every sequence of X 1’s and 0’s is 0 (we test sequences of 1 and 5). For both tests, we reject the null hypothesis if the actual result is statistically different, that is, the p-value is less than the critical value alpha = 0.05. All tests were run in R with freely available packages, and all models are 1-day, 95% VaR calculations unless otherwise stated.
III. Results

Figure 1a depicts the backtesting results (actual percent violations, and Kupiec, Ljung-Box(1), and Ljung-Box(5) p-values at each duration) for the Historical model and the EVT, GBM, and MAP (semi-parametric) models with the same parameters. For the statistical tests, boxes with green highlights indicate that the null hypothesis was not rejected and boxes with yellow highlights indicate that the null hypothesis was narrowly rejected. A 0 indicates that the p-value was below $10^{-4}$, and that the null hypothesis was subsequently rejected by a very large margin.

<table>
<thead>
<tr>
<th>hst 1d option</th>
<th>rb20 lkb5</th>
<th>LjungBox(1)</th>
<th>LjungBox(5)</th>
<th>gbm 1d option</th>
<th>rb20 lkb5</th>
<th>LjungBox(1)</th>
<th>LjungBox(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>% Violation</td>
<td>Kupiec</td>
<td>LB(1)</td>
<td>LB(5)</td>
<td>Duration</td>
<td>% Violation</td>
<td>Kupiec</td>
</tr>
<tr>
<td>250</td>
<td>3.2</td>
<td>0.163</td>
<td>0</td>
<td>0.002665</td>
<td>0</td>
<td>250</td>
<td>1.2</td>
</tr>
<tr>
<td>500</td>
<td>1.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>500</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>750</td>
<td>1.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>750</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>1.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>1.4</td>
<td>0.133</td>
</tr>
<tr>
<td>1700</td>
<td>1.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1700</td>
<td>5.3</td>
<td>0.581</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>evt 1d option</th>
<th>rb20 lkb5</th>
<th>LjungBox(1)</th>
<th>LjungBox(5)</th>
<th>map 1d option</th>
<th>rb20 lkb5</th>
<th>LjungBox(1)</th>
<th>LjungBox(5)</th>
<th>rb20 lkb5</th>
<th>LjungBox(1)</th>
<th>LjungBox(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>% Violation</td>
<td>Kupiec</td>
<td>LB(1)</td>
<td>LB(5)</td>
<td>Duration</td>
<td>% Violation</td>
<td>Kupiec</td>
<td>LB(1)</td>
<td>LB(5)</td>
<td>Duration</td>
</tr>
<tr>
<td>250</td>
<td>3.6</td>
<td>0.266</td>
<td>0.5509</td>
<td>0.02</td>
<td>250</td>
<td>0.4</td>
<td>0</td>
<td>0.9489</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>5.2</td>
<td>0.838</td>
<td>0.1344</td>
<td>0.11</td>
<td>500</td>
<td>0.8</td>
<td>0</td>
<td>0.8562</td>
<td>0.9964</td>
<td></td>
</tr>
<tr>
<td>750</td>
<td>6.3</td>
<td>0.125</td>
<td>0.2015</td>
<td>0.02</td>
<td>750</td>
<td>1.2</td>
<td>0</td>
<td>0.7386</td>
<td>0.00712</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>3.5</td>
<td>0.022</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1700</td>
<td>7.9</td>
<td>0</td>
<td>0.0001325</td>
<td>0</td>
<td>1700</td>
<td>4.6</td>
<td>0.459</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Fig 1a. Testing results: historical VaR and evt/GBM/MAP (rebalance 20, lookback 5)

Between these four models, the Historical is clearly the worst, posting just one fail-to-reject, and one narrow rejection. By this same metric, the EVT appears to be the best, but by a much smaller margin: it has just as many fail-to-rejects as the GBM, and one more narrow rejection. EVT further edges out the other two by being more consistent: it does relatively well in all tests through 750 days, while GBM and MAP don’t post fail-to-rejects in the Kupiec test until later.
Figure 1b shows the test results graphically. The green line is the VaR calculation, and the blue spikes are the realized profit and loss of the portfolio (same in every test here), centered around zero. When a realized loss exceeds the estimated VaR, it becomes a red spike, indicating a VaR violation. The huge increase in portfolio volatility in the center of the graph corresponds to the 2008 financial crisis.

Again, we can immediately see that the Historical VaR model performs poorly: it is far too conservative everywhere except during 2008, where instead, almost all of the violations occur. It is much more difficult to visually distinguish between the GBM, EVT, and MAP models: in all three the estimated VaR much more closely follows actual portfolio performance, though many violations still occur during the crisis.
One facet of this project involved creating a hedged portfolio for which we could compute VaR models. We decided to compare how well certain VaR models performed between a hedged and not hedged portfolio comprising the same core securities (the not hedged portfolio is simply missing the options); figure 2a shows these results.

We chose our two best models, the EVT and GBM, based on the previous comparison, but even these performed much worse with the stock-only portfolio, especially with regard to independence of violations. Surprisingly, the EVT model for the not hedged portfolio managed to fail-to-reject the null hypothesis in the Kupiec test for all durations.

---

<table>
<thead>
<tr>
<th>Duration</th>
<th>% Violation</th>
<th>Kupiec</th>
<th>LB(1)</th>
<th>LB(5)</th>
<th>Duration</th>
<th>% Violation</th>
<th>Kupiec</th>
<th>LB(1)</th>
<th>LB(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>3.6</td>
<td>0.286</td>
<td>0.5609</td>
<td>0.01747</td>
<td>250</td>
<td>1.2</td>
<td>0.001</td>
<td>0.8462</td>
<td>0.9934</td>
</tr>
<tr>
<td>500</td>
<td>5.2</td>
<td>0.838</td>
<td>0.1344</td>
<td>0.1063</td>
<td>500</td>
<td>1.2</td>
<td>0</td>
<td>0.7849</td>
<td>0.02629</td>
</tr>
<tr>
<td>750</td>
<td>6.3</td>
<td>0.125</td>
<td>0.2015</td>
<td>0</td>
<td>750</td>
<td>1.5</td>
<td>0</td>
<td>0.655</td>
<td>0.0004</td>
</tr>
<tr>
<td>1000</td>
<td>8.0</td>
<td>0</td>
<td>0.02348</td>
<td>0</td>
<td>1000</td>
<td>4</td>
<td>0.133</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1700</td>
<td>7.9</td>
<td>0</td>
<td>0.00001</td>
<td>0</td>
<td>1700</td>
<td>5.3</td>
<td>0.581</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Duration</th>
<th>% Violation</th>
<th>Kupiec</th>
<th>LB(1)</th>
<th>LB(5)</th>
<th>Duration</th>
<th>% Violation</th>
<th>Kupiec</th>
<th>LB(1)</th>
<th>LB(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>2.6</td>
<td>0.083</td>
<td>0.06041</td>
<td>0.1571</td>
<td>250</td>
<td>0.4</td>
<td>0</td>
<td>0.9469</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>4</td>
<td>0.288</td>
<td>0</td>
<td>0</td>
<td>500</td>
<td>1.6</td>
<td>0</td>
<td>0.013</td>
<td>0</td>
</tr>
<tr>
<td>750</td>
<td>4</td>
<td>0.194</td>
<td>0</td>
<td>0</td>
<td>750</td>
<td>1.3</td>
<td>0</td>
<td>0.01596</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>5.6</td>
<td>0.393</td>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>3.4</td>
<td>0.014</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1700</td>
<td>5.3</td>
<td>0.581</td>
<td>0</td>
<td>0</td>
<td>1700</td>
<td>3.8</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---

**Figure 2a.** Testing results: EVT and GBM (rebalance portfolio every 20 trading days, using 10-year historical data window), with and without hedging

---

![EVT graphs](image1.png)

**Figure 2b.** EVT graphs with and without hedging

Figure 2b shows the EVT results graphically. The actual portfolio volatility is much higher without the hedge, and it is correspondingly more difficult for the VaR model to follow it. Furthermore, while there are fewer actual violations for the stock-only portfolio (5.3% versus 7.9% at 1700 days), the average magnitude appears to be much larger, though this is not quantified in our testing.
Finally, we looked at adjusting the parameters of the EVT model (our best model so far) and Figure 3a shows those results. A 99% VaR calculated with rebalance 20 and lookback 5 posts the best result overall, failing to reject all tests through 750 days, though it just edges out rebalance 5/lookback 5, which in turn just edges out rebalance 20/lookback 6.

All of the 99% VaR models generally do better than the 95% VaR models. Nonetheless, the difference in model performance resulting from changes in parameters is small with the exception of the 10-day (all models up to this point have been for 1-day VaR), which does significantly worse than its 1-day counterpart in both unconditional coverage and independence.

Figure 3b highlights the poor performance of the EVT 10-day model relative to the EVT 1-day model. The VaR estimate more loosely follows actual profit and loss and the violations are more clustered around 2008.
IV. Conclusion
In conclusion, the results of our testing were somewhat of a mixed bag. The historical model was definitively worse than the three Monte-Carlo methods, but between those three, the EVT model only stands out slightly. Altering the parameters of the EVT model showed that the 99% VaR models did slightly better than their 95% counterparts, but once again, no particular model stood out, with the exception of the poor 10-day model.

More glaringly, none of the models were able to well predict the 2008 financial crisis. Even our best EVT model (rebalance portfolio every 20 trading days and use 10-year historical data window) posted failed-to-rejects in all tests only up to 1000 days, the effective start of the crisis. We see this sort of performance in all of our models to some extent, indicating that our current set of VaR calculation tools may need improvement before they can accurately model these Black Swan events.

Acknowledgement
We thank Prof. Kay Giesecke and Gerald Teng for the great instructions and insightful suggestions during the weekly meetings.

Contribution
Jiajing Xu: VaR model developing, coding

Xiaomeng Zhang: VaR model developing, coding

Derek Lim: Backtest method developing, coding

Reference