3.1 Back testing Design

The purpose of this document is to assess the accuracy and robustness of the Risk Calculation System. In accordance with the current Basel guidelines on VaR backtests, 1-day 99th percentile VaR calibrated against the prior 12-months' data is used. With a 1-day 99th percentile VaR, we expect daily losses to exceed the VaR measure approximately 1% of the time or, equivalently, 2.5 times in a 12-months period (250 trading days).

3.1.1 Accuracy Test

The first test is to examine the accuracy of the VaR measure. That is to assess whether the frequency of exceptions is consistent with the confidence quantile of loss the VaR measure is intended to reflect. For this test, the Basel committee provides the following traffic light classifications:

Backtesting Zone	Number of exceptions	Backtesting dependent multiplier
Green	0-4	1.50
Amber	5-9	1.70-1.92
red	10 or more	2.00

This classification splits the backtesting results into three groups based on the number of instances that the daily loss exceeds the VaR. The green zone does not suggest a problem with the accuracy of the model with no add-on to the capital requirement while the amber zone and the red zone indicates an indefinitive problem or a definitive problem to the risk model, respectively.

Three straightforward statistical tests for accuracy are also included for completeness, which are the Binomial Test, the Kupiec's proportion of failures (POF) test, and the Kupiec's time until first failure (TUFF) test, which all produces reject-or-accept result where the null hypothesis is that the exception probability is consistent with 1-VaR level. The test statistics are calculated as the following:

$$Z_{bin} = \frac{x - Np}{\sqrt{Np(1-p)}}$$

$$LR_{POF} = -2\log\left(\frac{(1-p)^{N-x}p^{x}}{\left(1-\frac{x}{N}\right)^{N-x}\left(\frac{x}{N}\right)^{x}}\right)$$

$$LR_{TUFF} = -2\log\left(\frac{(1-p)^{n-1}p}{\left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^{n-1}}\right)$$

Where x is the number of exceptions, N is the number of observations, n is the number of days until the first exceptions, and p = 1 – VaR level. Specifically, the binomial test is essentially the z-test with the underlying distribution as binomial distributions. The test statistics is distrusted as standard normal. Similarly, for the Kupiec's POF test, x is modelled as following the binomial distribution and the LR_{POF} statistics computes the likelihood ratio between the likelihood of x exceptions given the VaR level implied probability and the likelihood of x exceptions given the

empirical probability. LR_{POF} asymptotically follows Chi-squared distribution with one degree of freedom and can be expressed alternatively as $LR_{POF} = 2\log\{\frac{L(\hat{p}|H_A)}{L(p|H_0)}\}$ where $\hat{p} = \frac{x}{N}$ and $p = 1 - VaR\ level$. H_A stands for the alternative hypothesis that the probability of exceptions is not the VaR implied probability while H_0 stands for the null hypothesis that the probability of exception is the VaR implied probability. The Kupiec's POF test is again a likelihood ratio value while the POF test models the time to the first exception and models it as following the geometric distribution. Again, the statistics is distributed as chi-squared with one degree of freedom.

Both the binomial test and Kupiec's tests suggested that the null hypothesis should be rejected if the test statistics exceed the critical value. For this backtest, a significance level of 0.05 and 0.01 is used. The critical values (one-sided) are summarized as below.

Distribution	$\alpha = 0.05$	$\alpha = 0.01$
Chi-squared distribution with 1 degree of freedoms	3.841	6.635
Standard Normal	1.645	2.326

All test statistics computed for this backtest are unconditional coverage tests which does not capture the clustering of exceptions commonly observed in empirical observations. Conditional coverage tests such as the Christoffersen's test will reflect it. For further backtesting of the risk calculation system, conditional coverage tests will be included.

Three methods of VaR calculation, Monte Carlo method, Historical method, and Parametric Method, are included in the risk system so for each of the method, we will compute the 1-day 99th VaR with 1-year window and the above-mentioned statistics for all three methods over a 10-year window. Detailed graphic demonstrations will be provided. For the Monte Carlo Method, drifts and volatility will be calculated using both methods, unweighted moving average method and the exponentially weighted moving average method.

3.1.2 Robustness Test

The next part of the test plan is to assess the robustness of the risk calculation system. For simplicity, only the traffic light test and the Kupiec's POF test will be provided in this section. For the test against different market conditions and specifications, a long-only, short-only, and long-short portfolios of size 5 will be used, all with a total position of 10000.

To test the stability and robustness of the model, we will consider the following time periods to reflect changes in the market and investigate the robustness of the model under different market conditions.

Early 2000s Recession	2000.01 – 2002.12
Economic Expansion post 2000s Recession	2004.01 – 2005.12
2008 Financial Crisis	2008.01 – 2009.12
Economic Expansion post 2008	2014.01 – 2015.12
COVID-19 Recession	2020.01 – 2021.12

All time periods are two-years window so the level of accuracy should be similar. All three methods will be applied to each of the five market periods with 3 representing market down times with high volatility and correlation and 2 representing market up times with low volatility and correlation. The standard 1-day 99th percentile VaR with 1-year window will be used for this test.

To test the stability and robustness of the model against different specifications, the following sets of parameters will be used over the 10-year window.

1. VaR percentile: 95%, 97.5%, 99%
2. VaR frequency: 1-day, 5-days, 10-days
3. VaR window: 1-year, 2-years

To assess behavior over a large range of inputs, a long only portfolio of size 1, 10, 50, 100 will be considered to compare the model's accuracy. For this part, each stock will have a position of 10000. Again, the standard 1-day 99th percentile VaR with 1-year window will be used for this test.

Test Result

- Summary of different portfolios.
- Summary of results with commentary and evaluation.
- Detailed analysis of informative samples