Test Plan

I. Backtesting introduction:

Backtesting is a statistical procedure that we use to compare the actual profits and to corresponding VaR estimates. If the confidence level of the daily VaR is 99%, we expect there is only one exception for a 100 days period on average. In the backtesting process we statistically examine whether the frequency of exceptions over some certain time period is less than the number of days of that period times (1-confidence level).

In order to evaluate the quality of the estimates, we choose to use sixes stocks (CLEG, AMD, K, VZ, JPM, MS) and three portfolios that are formed by them to backtest the VaR calculation models with two methods: Basel Regulatory Framework and Kupiec's POF-test.

II. Backtesting Evaluation method

1. Basel Regulatory Framework

From Basel Committee we know that for banks they are required to calculate 10-day VaR using a confidence level of 99%. To make sure banks estimate their risk correctly, an appropriate backtesting approach is necessary and plays a vital role when Basel decide whether the bank's internal VaR models for capital requirements calculation is reasonable or not.

The Basel Regulatory Framework requires us compare last 250-days 99% VaR estimate with corresponding reality payoffs. We evaluate the VaR model's accuracy by counting the exception numbers during this period. And the backtesting results are classified in to three categories: green zone, yellow zone and red zone.

Zone	Number of exceptions	Increase in scaling factor	Cumulative probability
Green Zone	0	0.00	8.11 %
	1	0.00	28.58 %
	2	0.00	54.32 %
	3	0.00	75.81 %
	4	0.00	89.22 %
Yellow Zone	5	0.40	95.88 %
	6	0.50	98.63 %
	7	0.65	99.60 %
	8	0.75	99.89 %
	9	0.85	99.97 %
Red Zone	10 or more	1.00	99.99 %

The green zone including exceptions from 0 to 4. And a green zone model is regard as a accurate model.
The yellow zone including exceptions from 5 to 9. Backtesting results in yellow zone would cause an increase of the multiplication factor when bank need to calculate its regulatory capital.
The red zone including exceptions greater than 10. It indicates that there is a problem with current VaR model and would lead to a rejection.

2. Kupiec's POF-test

Kupiec's Proportion of Failures (POF) test measures whether the number of exceptions is consistent with the confidence level. We assume the number of exceptions as x, the total number of days during the period is T, and p to be the frequency of tail losses

$$p = 1 - confidence\ level$$

Then we can define the failure rate as x/T.

The null hypothesis for POF test is:

$$H_0: p = \hat{p} = \frac{x}{T}.$$

Kupiec conduct POF test as a likelihood-ratio (LR) test.

$$LR_{POF}: -2ln\left(\frac{(1-p)^{T-x}p^x}{[1-(\frac{x}{T})]^{T-x}(\frac{x}{T})^x}\right)$$

Under the null hypothesis that the model is correct, LR_{POF} is asymptotically chi-squared distributed with degree of freedom equal to one. If the value of LR_{POF}-statistic greater than the critical value of the chi-squared distribution, the null hypothesis will be rejected and the model will be regarded as inaccurate.

Probability Level p	VaR Confidence Level	Nonrejection Region for Number of Failures N		
		T = 255 days	T = 510 days	T = 1000 days
0.01	99 %	N < 7	1 < N < 11	4 < N < 17
0.025	97.5 %	2 < N < 12	6 < N < 21	15 < N < 36
0.05	95 %	6 < N < 21	16 < N < 36	37 < N < 65
0.075	92.5 %	11 < N < 28	27 < N < 51	59 < N < 92
0.1	90 %	16 < N < 36	38 < N < 65	81 < N < 120

From nonrejection region above, we can see that the POF test not only reject the model when the number exception days within certain period is too big, but also reject the model when the number of exception days within certain period is too small for certain confidence level, which implies the overestimated VaR.

III. Backtesting Process

1. **Data**

   To test the accuracy of our VaR calculation models, we choose three groups of stocks.

	Low volatility (<20%)	Middle volatility (20%~40%)	High volatility (>40%)
stock	K (Kellogg Company), VZ (Verizon Communications)	JPM (JPMorgan chase), MS (Morgan Stanley)	SPG (Simon Property Group), AMD (Advanced Micro Devices)
portfolio	K, VZ	JPM, MS	SPG, AMD

   In this way, we could examine the models' performance for low volatility stocks, high volatility stocks, and stocks with high correlation. To make sure only consider one variable a time, we select the stocks for correlation study have the middle volatility rate in the market, and the stocks for volatility study are have relative low correlation with each other: the companies are scattered in different industries.

   All stocks mentioned above have over 20 years' trading history and are still trading on the stock market.

2. **VaR Calculation**

   We compute the 99% and 95% 5-day historical VaR, parametric VaR, and Monte Carlo VaR for a $100,000 investment in selected stocks, portfolios, and portfolio option's long and short positions on each day in the past 20 years.

3. **Backtesting**

   Backtesting requires us to compare 5-day trading outcomes and 5-day VaR estimates. To do that we first calculate the 5-day trading outcomes by using $100,000(day t position's value) minus day t-5 position's value. Noted that the day t-5 position has the same number of stocks as day-t position. After that, we could compare each day's 5-day actual payoff with that day's estimated VaR and counting the number of exceptions from that day to one year ago. Then we can get the average VaR exception number for the position in one year period. Finally, we use the Basel Regulatory Framework and Kupiec's POF-test to analysis the results we get and justify the accuracy of the model.

Notice that we also test 95% 5-day VaR, which means we need to calculate the corresponding evaluation standards for Basel Regulatory Framework method and Kupiec's POF-test according to their definition.

95% confidence level standard Basel Regulatory Framework calculation:

Comparing each day's payoff and estimated VaR only generate two results: exceed VaR or not. If we regard payoff exceed the VaR as a "success" and for those do not exceed as a "failure", the we can regard the comparing process as a Bernoulli trial and we can get the number of exceptions x follows a binomial distribution $B(n, p)$.

And when the number of trials n increase, the binomial distribution can be approximated with a normal distribution $N(pT, p(1-p)T)$

We know that the yellow zone for the 99% confidence level VaR begins at the cumulative probability 95% and red zone begins at 99.99%. So, we can calculate the cut-off points for 95% confidence level with 250 observations by applying binomial distribution.

Zone	95%	99%
Green	0-17	0-4
Yellow	18-26	5-9
Red	27+	10+

To better evaluate the accuracy of models, we also perform Pre-Crisis Evaluation and Post-Crisis Evaluation, to see how our risk calculation system perform without the impact of crisis and with the crisis in mind. For the Pre-Crisis Evaluation, we use two years data before the 2008 financial crisis: Dec,2005-Dec,2007. And for the Post-Crisis Evaluation, we use two years data after the 2008 financial crisis: Jan, 2010-Jan, 2012.

The form of the backtesting result would be:

Stock/portfolio	Drift and volatility estimation method	Confidence level	Position Direction	Pre-crisis	Post-crisis	Average yearly exception number
	5-year window	99%	Long			
			Short			
		95%	Long			
			Short			
	Exponential	99%	Long			
			Short			
		95%	Long			
			Short			