FIN567 PROJECT

Names	NetID
Kai Yang	kaiy2
Renjun Li	renjunl2
Meichen Zhang	meichen6
Danni Chen	dannic3
Ruixin Zhang	ruixin3

Background of our project: Dispersion trade

Nowadays, equity volatility has been one of important signal of market change. As Helen (2018) said that we should pay attention to the theta-flat products, a useful index in a higher volatility market, which means when there is a widespread of theta-flat dispersion, the volatility will go to a higher level. While in low volatility period, vega-flat alternative is the standard and investors more prefer to see the "vega-flat structures with more diversified baskets and illiquid names". In conclusion, a higher volatility, in short period we use theta-flat, while in long-term, vega-neutral dispersion will be more attractive.

Source:

Bartholomew, H(2018). Old dispersion product signals new vol regime. Risk.net.

Introduction

For this project, we get into a dispersion trade, i.e. we write straddles (one at-the-money European Straddles) on the DJX index and buy straddles (k notional amounts of at-the-money American Straddles) on each of the component stocks inside the Dow Jones index. We try to acquire the data of actual listed options(near-the-money) from WRDS/OptionMetrics, but we could only find near-the-money options traded in the actual market. Luckily, on the section of volatility surface in OptionMetrics, there were interpolated at-the-money option data with interpolated implied volatilities and implied strike price for all the underlying stocks we want in our portfolio, which is great! (Options where their delta = 50 and their days to expiration is 30). We will pick June 27th, 2019 as the initial date where the options have about one-month expiration (actually it's 21 trading days). As a reminder, Dow Jones Index is a simple average of prices, then divided by the divisor (This implied that index is price weighted). are using the Dow Jones index weights (which means those component stocks are price-weighted, and the dollar investments are proportional to the stock price of each component stock). So equal number of shares results in price weighting. For the project, the implication of this is that you should hold equal numbers of the component options. Besides, we use at-the-money straddles since we want to avoid the stock-price risks, though there is great possibility that those at-the-money options we use weren't actually traded on the initial day. We create our portfolio by using two methods to get the size(k) of the straddles for each of the component stocks in the Dow Jones Index at time 0:

- Vega Neutral Portfolio;
- Theta Neutral Portfolio.

Once we create these two portfolios(actually what we do here is confirm the size of k for component stocks for Dow Jones), we calculate our profit/loss of the portfolio for each day when we hold this portfolio until a final date(one-day or one-month time horizon) and then compute the VAR (1% and 5%) by simply using the quantile function in R on the simulated data of portfolio profits/losses. The probable profits/losses from our portfolio is computed using the method of Multivariate filtered historical simulation. To be more precise, we try to simulate the log change of implied volatility of all the options and log returns of stock prices by first acquiring a historical database of log difference of both stock prices and implied volatilities for the options in the past 2 years(actually that's 503 trading days) before the initial date and then using the filtered historical simulation to simulate the future log returns of implied volatility and the prices for all the options inside the whole portfolio we have constructed on the initial date(a little similar to what we have done in hw3, instead we are using FHS simulation not Monte Carlo simulation here). We calculate the European option prices(Dow Jones Index Straddle) by using the GBSOption R function and calculate the American option prices(other component stock straddles) by using the CRRBinomialTreeOption R function .We set the numbers of times of Filtered Historical Simulation trials to be 10000 so that we can generate 10000 portfolio profits/losses to calculate the VaR & Expected Shortfall of the portfolio using one-day horizon or 10000*21 daily portfolio profits/losses to calculate the VaR & Expected Shortfall of the portfolio using a 21-day horizon. In addition to this, we calculate some Greek letter risks of our portfolio, namely Delta, Gamma, Vega and Theta (or even Rho) by using the built-in function in R shown in hw2("GBSGreeks" function in R). We will also discuss some other risks (such as stock price risks, etc.) or some disadvantages of our system to measure the risk of the dispersion trade in the end of our project.

Data

We obtain the option data from WRDS/OptionMetrics. We use the data from past 2 years before the initial date (we choose 06/27/2019 as the initial date). We assume that we determine the composition of the portfolio on 06/27/2019. We will try to use the historical data from OptionMetrics to determine what the cost of holding this portfolio will be in the future by simulating the stock price and implied volatility on the future dates(till one month) and calculating the option prices on every day. To be more precise, we will use Black-Scholes model(for European Options) and Binomial-Tree model(for American options) to compute the prices of all the options inside the portfolio in order to get the value of the portfolio on future days. Also, since we are trying to find at-the-money options to construct the portfolio, initially we thought it was very difficult to find actual at-the-money options for underlying component stocks/Dow Jones Indexes since there might be no actual at-the-money options traded in the market on the initial day. But luckily, as Professor has shown in the previous lecture, OptionMetrics has helped us located the at-the-money stocks with an expiration date of 30 days by doing the interpolation work secretly. To be more specific or clear, OptionMetrics has linearly interpolated and calculated the implied volatility of two options with the prices closest to the strike price(one-month till expiration) and we use that for our computation and assume those are actual at-the-money options that are being traded in the derivatives market.
We obtain the annualized risk-free rate during our holding period of the portfolio(one month after the initial day we have set) through the online data source—Fred Economics Data(whose website url is: "https://fred.stlouisfed.org/series/USD1MTD156N"). We retrieve the 1-month LIBOR rate from the initial date (June 27th, 2019) from this website and denote it as the risk-free rate during the holding period for the calculation of all option prices.
We also obtain the historical stock prices and (continuous) dividend yield of all the underlying stocks inside our portfolio from yahoo finance since our group members learned how to acquire historical closed prices of all the stocks inside our portfolio by just inputting the tickers and certain function in Rstudio. Again, we retrieved 2 years' historical data before

the initial date (2017/06/27-2019/06/27). As for the dividend yield part, we use the way Professor Pearson has told us in the email sent recently. In detail, we only consider the dividend that took place during the life of the options we have set initially. To be more precise, we use stock VZ as an example. If the valuation date is June 27. 2019 and you are valuing the July 2019 options, there is one dividend during the life of the options: a dividend of $0.6025 with ex-dividend date of July 9, 2019. The VZ closing price on June 27, 2019 was $57.25, so the dividend is 1.0524% of the stock price. The remaining time to expiration is 15 trading days (July 4 is a holiday), or 15/252 = 0.059524 year. The dividend yield of 1.0524% per 15 days is roughly like 17% or 18% per year. In fact, the equivalent continuous dividend yield d can be obtained by solving 1 – 0.010524 = exp(-0.059524d), yielding d = 0.17774.
What's more, we found that we couldn't find the historical data for one of the component stocks inside DJX, Dow Inc. (ticker: DOW). During class, Professor Pearson told us to drop DOW and only use other 29 stocks inside DJX, because dealing with the DOW merger will be extremely annoying and time-consuming.

METHODOLOGY

We have talked a lot about our methodologies which have been used in our group's project in the first introduction section. Right now, I just want to highlight some of the tricky steps inside those methodologies that require further explanation.

1. How do we determine the composition of the portfolio (For this specific project, we want to explain how we get the amounts of straddles (k here) for each component stocks inside the Dow Jones index(two cases: Vega-Neutral and Theta Neutral):

Our groups want to answer this question by using the following formula to tell you how many amounts of straddles for each of the component stocks in Dow Jones index that we want to hold for the portfolio on the initial date:

$$-Vega_{index} + k_{vega-neutral} \sum_{i=1}^{T} Vega_i = 0$$

$$-Theta_{index} + k_{theta-neutral} \sum_{i=1}^{T} Theta_i = 0$$

Remember, all the vegas and thetas in the two equations above are the respective greek letter values of all the at-the-money options on the initial date(which we can get the data in a straight-forward way from OptionMetrics). $k_{vega-neutral}$ and $k_{theta-neutral}$ are the amounts of straddles for each component stocks we need to hold for both cases' portfolios.
Furthermore, if we consider the instructions of the whole project inside the **F567.2020.project.pdf** downloaded from Compass, we quote "In each case, you will hold the options on the index components in proportion to their underlying stocks' weights in the Index" as shown in the formula below:

$$-Vega_{index} + k_{vega-neutral} \sum_{i=1}^{T} wi * Vega_i = 0$$

$$-Theta_{index} + k_{theta-neutral} \sum_{i=1}^{T} wi * Theta_i = 0$$

Specifically, if the index weights are w1, w2, w3, etc., then the notional amounts of the component options will be kw1, kw2, kw3, …, where k is chosen to make the trade either vega-neutral or theta-neutral. Here w1=w2=w3=⋯⋯=1/29. The reasons for an equal weight of all the index components present in our portfolio are already shown in the introduction section above. Since we hold equal number of options for the index components, the two ways of calculating K using the two pairs of formulas above are essentially the same. We use the first pair of formulas to calculate K since it's more straight-forward (ignoring the equal weight part).

2. How do we use the Multivariate filtered historical simulation methods to simulate the profits/losses of the portfolio we have constructed in detail?

First, I want to talk about the exact reasons why we use Multivariate filtered historical simulation to forecast stock prices and implied volatilities over a one-day/one-month horizon: Filtered Historical Simulation combines the strengths of historical simulation with the strengths of dynamic risk models. Specifically, FHS combines model-based forecasts of variance with model-free approaches to the distribution of shocks. **Most importantly, we use the past returns and past implied volatility(change) data to tell us about the distribution without making any assumptions about the specific distribution. For this project, we assume constant correlations for Multivariate FHS.**

Second, what market factors are we trying to simulate inside the Multivariate FHS? To be more precise, just pay attention to the market_factor matrix we have generated in our R scripts submitted along with this project. This matrix has 503 rows and 90 columns, which means that we have 90 market factors with 503 rows of historical data. These 90 market factors include the log returns of all the stocks (1 index and 29 component stocks), the log returns of the implied volatilities for all the call options and all the put options. (30*3=90).

Reminder, when we exchange emails to ask questions about the project with Professor Pearson, he mentioned that we can save on 30 market factors by assuming that the implied vols of the matching call and put should be the same since put-call parity says they should be the same. In the data, they differ only because the midpoint is only an estimate of the fair value. After intense discussion among the group members, we decided to stick with the original plan and go with 90 market factors to have a more accurate estimate of the put/call option prices inside our portfolio. The reason for that is when we still get access to the Bloomberg terminal at the market lab inside the Gies College of Business. We remember when we need to find the data of implied vol on the Bloomberg terminal, we need to manually input the type of options(put/call) to get the implied vol data. Based on our memory, especially for American options, the implied vol of the put options is different from that of the call options for the same underlying stock while all else equal.

Also, we want to highlight that we consider filtered historical simulation with constant correlations is a perfectly good approach. We understand that if we use a constant correlation matrix, then of course you can simulate the process of stock returns and option implied vols separately. We still construct a matrix of all the market factors including the returns of the stock prices and log change in implied vols of all the options inside the portfolio since first, we believe changes in implied vols are correlated with stock log returns and two, it saves us a lot of time and energy to input all the market factors at once inside our model.

Since we include a thorough steps-by-steps introduction of how we use Multivariate FHS to simulate all the future market factors in our write-up for project part 1, and we quote Professor Pearson's comments "Your write up is much longer than it needs to be---you have a lot of stuff you don't need". We decided not to include all the details of the FHS model to forecast risk over a 1-day or a 1-month horizon. Please check the lecture notes of SimTSRisk-03.pptx posted on compass. To be more precise, page 4-6(actually these pages of slides tell us of how we use univariate FHS) and 11-12 inside this specific powerpoint slide include all the theories and steps of how we execute the Multivariate FHS assuming constant correlations for all the market factors exactly in this project.

I want to highlight one small point when we want to calculate the VaR of both portfolios using a horizon of one-month (21 trading days). Since after 21 trading days, all the options inside our portfolio expired simultaneously, which means we only need to simulate 30 market factors(log return of the 30 stocks' prices) to get the prices after 21 trading days for all the underlying stocks inside our portfolio and simply use the option payoff to calculate the value of all the options so that we can finally get the profits/losses for both portfolios in 21 trading days' time. In short, we don't care about the implied volatilities of all the options at the expiration date.

RESULTS

- **Vega-Neutral Portfolio:**

1. Composition

Our Vega-Neutral Portfolio consists of writing one index straddles and buying 0.06891891 **(see the variable k_vega_neutral in our R scipts)** notional amounts of all the 29 component straddles inside DJX.

2. Initial Portfolio Value

5.267989 **(see the variable port_time0_vega_neutral in our R scripts)**

3. Greek Letter risks of this portfolio

Delta: -0.07140155 **(see delta_port_vega)**
Gamma: 0.1889099 **(see gamma_port_vega)**
Vega: 0 **(see vega_port_vega)**
Theta: -31.0887 **(see theta_port_vega)**

4. 1%/5% value-at-risk (different time horizons) of this portfolio

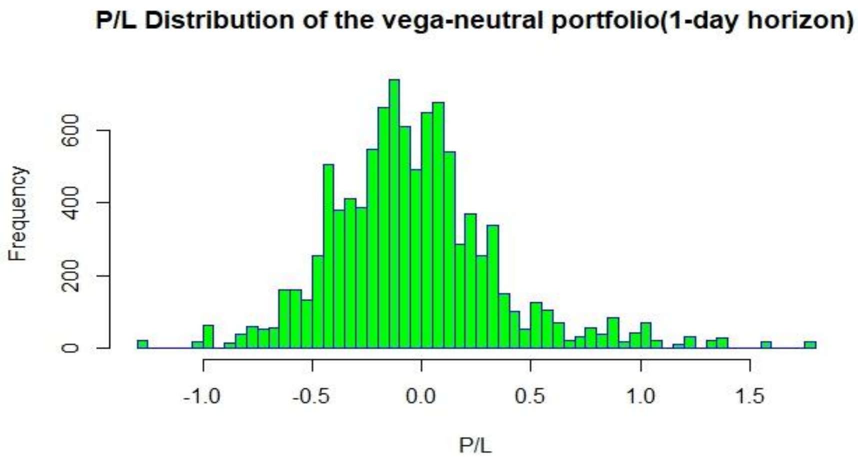
1% Value-At-Risk using a horizon of 1 day: 0.8803429 **(see VaR_vega_1perc)**

5% Value-At-Risk using a horizon of 1 day: 0.5913177 **(see VaR_vega_5perc)**

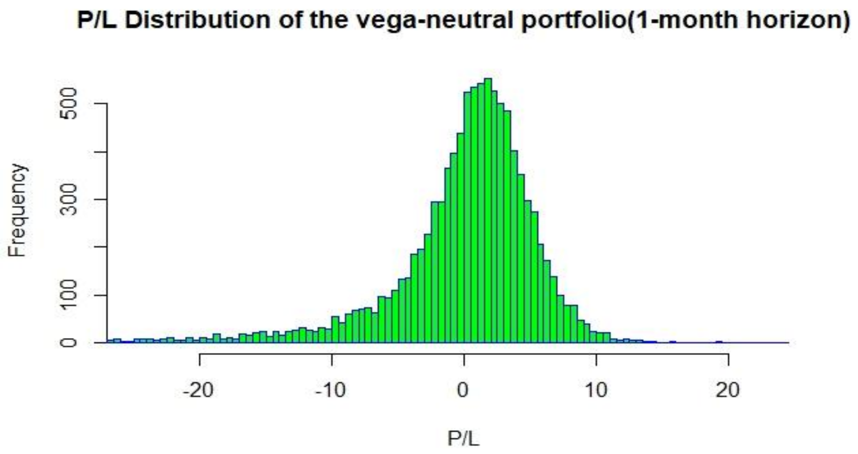
1% Value-At-Risk using a horizon of 1month (21 trading days): 31.84948 **(see VaR_vega_1perc_21)**

5% Value-At-Risk using a horizon of 1month (21 trading days): 12.25691 **(see VaR_vega_5perc_21)**

5. Distribution of Profits/Losses for this portfolio



The distribution of variable **PL_vega_sim1** in our R scripts.



The distribution of variable **PL_vega_sim21** in our R scripts.

6. Expected Shortfalls for different cases using different VaRs

Expected Shortfall using 1% Value-At-Risk and time horizon of 1 day: 1.046993 (see **ES_1perc_vega_1**)

Expected Shortfall using 5% Value-At-Risk and time horizon of 1 day: 0.7578389 (see **ES_5perc_vega_1**)

Expected Shortfall using 1% Value-At-Risk and time horizon of 1 month: 58.61216 (see **ES_1perc_vega_21**)

Expected Shortfall using 5% Value-At-Risk and time horizon of 1 month: 26.63228 (see **ES_5perc_vega_21**)

- **Thega-Neutral Portfolio:**

1. Composition

Our Vega-Neutral Portfolio consists of writing one index straddles and buying 0.04309384 **(see the variable k_theta_neutral in our R scipts)** notional amounts of all the 29 component straddles inside DJX.

2. Initial Portfolio Value

0.05179817 **(see the variable port_time0_theta_neutral in our R scripts)**

3. Greek Letter risks of this portfolio

Delta: -0.06012681 **(see delta_port_theta)**
Gamma 0.09050375 **(see gamma_port_ theta)**
Vega: -22.86394 **(see vega_port_ theta)**
Theta: 0 **(see theta_port_ theta)**

4. 1%/5% value-at-risk (different time horizons) of this portfolio

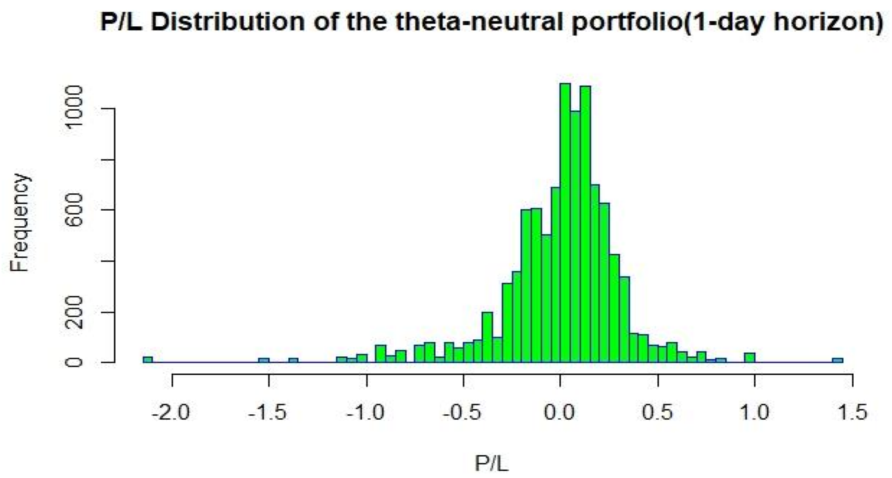
1% Value-At-Risk using a horizon of 1 day: 1.006811 **(see VaR_theta_1perc)**

5% Value-At-Risk using a horizon of 1 day: 0.5563275 **(see VaR_theta_5perc)**

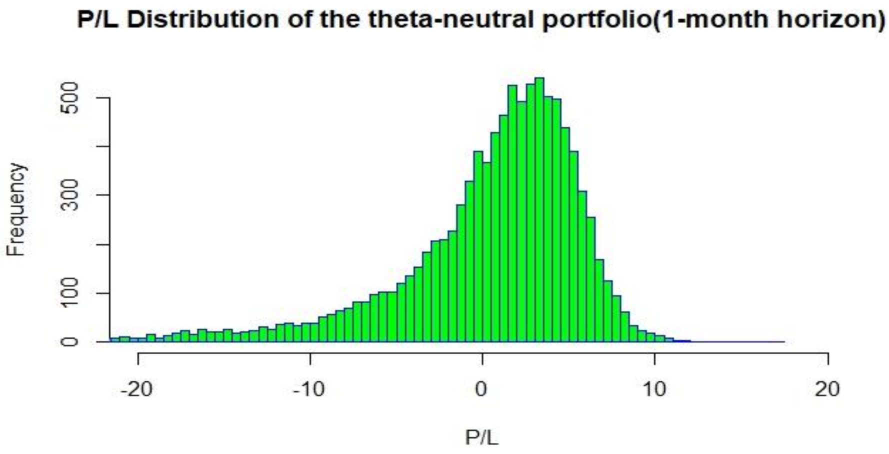
1% Value-At-Risk using a horizon of 1month (21 trading days): 34.49412 **(see VaR_theta_1perc_21)**

5% Value-At-Risk using a horizon of 1month (21 trading days): 13.6145 **(see VaR_theta_5perc_21)**

5. Distribution of Profits/Losses for this portfolio



The distribution of variable **PL_theta_sim1** in our R scripts.



The distribution of variable **PL_theta_sim21** in our R scripts.

6. **Expected Shortfalls for different cases using different VaRs**

Expected Shortfall using 1% Value-At-Risk and time horizon of 1 day: 1.337795 (see **ES_1perc_theta_1**)

Expected Shortfall using 5% Value-At-Risk and time horizon of 1 day: 0.8805674 (see **ES_5perc_theta_1**)

Expected Shortfall using 1% Value-At-Risk and time horizon of 1 month: 61.9856 (see **ES_1perc_theta_21**)

Expected Shortfall using 5% Value-At-Risk and time horizon of 1 month: 28.87142 (see **ES_5perc_theta_21**)

Short analysis of the results for VaR(ES): The 1%/5% VaRs(ESs) for both portfolios using a horizon of 21 trading days are much larger than the respective VaRs(ESs) for both portfolios using a horizon of 1 trading days due to the combination of the longer horizon.

<u>Risks not captured by our risk measurement system</u>

1. According to professor, there are two strategies to construct our portfolios. To be more specific, we really want to use at-the-money straddles to construct our portfolios to mitigate the stock price risks. However actual options traded in the market are not exactly at the money because the stock price on time 0 is usually not the same as the strike price of the options in one months (21 trading days). One way is to imagining making an over-the-counter deal with our favourite derivatives dealer (GS, JP

Morgan). Then all the straddles should be exactly at-the-money. The other way is to use actual listed (near-the-money) options to construct our model. Basically, in our project, we adopt the first method since we can downloaded the data for all the at-the-money options on time 0 and we imagine we are making an OTC deal with the derivatives dealer to buy/short those imaginary options. One shortcoming of this strategy is that as what have discussed above in the section of **Introduction and Data**, OptionMetrics has linearly interpolated and calculated the implied volatility/strike price of two options with the prices closest to the strike price(one-month till expiration) and we use those historical interpolated data for our computation of other parameters(for example, Greek letter risks). We may assume those are actual at-the-money options that are being traded in the derivatives market but actually they aren't. We think the historical data(implied volatilities, implied strike prices) of those imaginary options may lead to estimation errors of the valuation of options but will only hurt the accuracy of our estimates for the portfolio value in a limited way.

2. As we have already discussed above in the Data Section (last paragraph). We didn't include Dow Inc. (ticker: DOW) in both our portfolios based on what Professor Pearson has said during class. The reason for that is because dealing with the DOW merger will be extremely annoying and time-consuming. However, we can't ignore the fact that The DJX index option contract is based on 1/100th (one-one-hundredth) of the current value of the Dow Jones Industrial Average(DJIA) and The Dow Jones Industrial Average (DJIA) is an index that tracks 30 large, publicly-owned blue chip companies trading on the New York Stock Exchange (NYSE) and the NASDAQ, which means to replicate the Dow Jones index, you should hold 1/divisor shares of each of the 30 component stocks. For convenience, we just drop the straddles of DOW Inc out of our holding portfolios. I believe there is a better way than just dropping one of the component stocks options from the portfolio to ensure the accuracy of all the estimates in our project. Besides, we cannot rule out the possibility that during the past 2 years prior to the initial date we have set, some of the 30 component stocks weren't included inside the DJX two years ago. What I am trying to say is that during those two years, the composition of the 30 component stocks inside the DJX might be changing. Since the value of Dow Jones index is equal to 1/divisor shares of each component stock of Dow Jones index, it might hurt our estimates of the values for the whole portfolios during our holding period.

3. We have asked about this problem when we exchange emails with Professor Pearson last week. Though he told us that it's a very trivial problem and we would be fine since this problem barely influence our estimates of VaR/ES in our project, we decided to talk about this problem inside this section. As shown above, our portfolios are comprised of one European straddles and K amounts of American straddles for each of the 29 component stocks inside the DJX. Though we can find the GBSGreeks R function specifically used to calculate the Greek letter risks of the European options, we couldn't find a specific R function to calculate those risks of the American options. During our research process, we found that the CRRBinomialTreeOptions can only generate the value(prices) of the American options, but this function didn't have the outputs of the Greek letter risks for American options (only the price). We have tried other new functions such as AmericanOption function from the RQuantLib package(you can check the introduction of this function on url: https://rdrr.io/rforge/RQuantLib/man/AmericanOption.html for better knowledge) but it's very unstable and flawed since it didn't always give us correct calculations of the

Greek letter risks we want. So, we have to use GBSGreeks functions even for the calculation of the Greek Letter Risks for all the American options, which may cause trivial errors in measuring the Greek Letter Risks of the whole portfolios for both cases.

4. We are using the Multivariate FHS approach assuming constant correlations among market factors. Though, for this project, our approach might turn out to be a perfectly good approach. However, in some of the cases taking place in reality, similar to volatilities, correlations change over time, which means the combinations of the FHS and DCC models might be more in line with reality. Furthermore, we use GARCH (1,1) models to simulate and forecast the volatilities of all the market factors. Again, though GARCH (1,1) models are already considered by many as perfect models used for option valuation when implied volatilities are not available, these models overlook the leverage effect**.** The case that a negative return increases variance by more than a positive return of the same magnitude is referred to as the leverage effect. All in all, we might need to modify the GARCH (1,1) models to some of the complex GARCH generations including NGARCH models.

Explanations of some important variables in our R scripts

There are approximately 400 lines of codes in our submitted R scripts so it's unreasonable to attach so much code to our final write-up. But in case you want to examine our R codes, we want to explain some of the variables we have deemed important during the course of completing this project so that you won't get lost when running our R codes. Since we have already talked some of the variables in the RESULTS section, we will exclude those variables in this section. Hope this section can help you understand our thinking process.

market_factor:

It's a very large matrix with 503 rows and 90 columns. It's one of the most important variables in our R codes. It represents the historical data of 90 market factors(two years' historical log returns of all the stock prices, log change of implied volatilities for all the call/put options prior to the initial date) which act as inputs to simulate future stock prices and option implied volatilities in order to finalize the overall portfolio values on future dates.

sigma_tplus1:

It's a vector with 90 numbers. Those 90 numbers are the simulated standard deviation of 90 market factors on time 1 based on the 90 separate fitted GARCH (1,1) models.

daily_shock

It's a very large matrix with 503 rows and 90 columns too like market_factor. It can be denoted as the historical database of daily shocks for all 90 market factors (historical Zs in professor's slides). Since we are using multivariate FHS approaches assuming constant correlations. The first step is to draw a vector (across assets) of historical shocks from a particular day in historical sample of shocks(actually in our project we draw a random number from 1 to 503, and draw the row m from this matrix, this row is actually a vector of historical shocks from the same random day in our two-year time horizon prior to time 0), and use that to simulate tomorrow's shock.

Function: multivariate_FHS_1day

We write a R function of our own to illustrate the main idea of how we are going to use multivariate FHS approach for one-day horizon. To utilize this function, you need to assign values to the input parameters of this function, which include FH(the number of times we draw from the standardized residuals on each future date, the number of FHS trials on each future date, we use 10,000 for this project), shock_database(the historical database of daily shocks for all the market factors you want to simulate, like daily_shock), sigma_tplus1(the simulated standard deviation of all the market factors on time 1, like sigma_tplus1 mentioned above). If we input the correct arguments for this function, the function will return a vector(sim_ret_tplus1) of values whose amounts is the same as the number of columns for the input argument: shock_database. The values inside this vector are the respective simulated log returns of all the factors on time 1whose names and orders are in line with those of column names inside the input argument shock_database matrix.

Function: multivariate_FHS_ 1month

We write a R function of our own to illustrate the main idea of how we are going to use multivariate FHS approach over one-month (21 trading days actually) horizon. To utilize this function, you need to assign values to the input parameters of this function, which include FH(the number of times we draw from the standardized residuals on each future date, the number of FHS trials on each future date, we use 10,000 for this project), shock_database(the historical database of daily shocks for all the market factors you want to simulate, like daily_shock_stock in our R script since we only care about the 30 market factors of log returns of stock prices here), sigma_initial(the simulated standard deviation of all the market factors on time 1, like sigma_tplus1 mentioned above) and garch_coef(the coefficients of all the fitted GARCH(1,1) models for the market factors we care about, like garch_stock_coef in our R script). If we input the correct arguments for this function, the function will return a very large matrix(sim_ret_cul) whose number of rows are the number of factors we care about(the column numbers of shock_database) and number of columns is the same as the number of FH, one of the input argument. This matrix represents the sum of simulated log changes of the market factors we care about over 21 trading days after the initial date in each of the multivariate FHS trial, which is very useful. Like what we have done here in our final project, we use this function to generate 10,000 simulated the sum of log changes of all the stock prices over 21 trading days. Originally, we can get all the stock prices at time 0, we can easily get the simulated stock prices exactly after 21 trading days (which is exactly the expiration day). Since we also have the exercise price of all the options on the expiration day, we can then use the option payoff to calculate the value of all the put/call options on the expiration date and finally get the value of the whole portfolio on the expiration day in each of the 10,000 FHS trials.