# 567 FINAL PROJECT

Sihan Li, sihanl2@illinois.edu
Yuchen Duan, yuchend3@illinois.edu
Ruozhong Yang, ry8@illinois.edu
Fengkai Xu, fengkai4@illinois.edu
Joseph Loss, loss2@illinois.edu

## 1. Background of the Dispersion Trades

### 1.1 Portfolio construction

$$Portfolio = -Dow\ Jones\ Index\ straddle\ + 30\ components\ straddle$$
$$Straddle = Call + Put$$

In the following paragraph we will use $P$ to denote portfolio and $C_i$ denotes the $i_{th}$ components.

### 1.2 Weights among Dow Jones index

Dow-Jones is a price-weighted index, therefore the weight of the $i_{th}$ components $w_i$ should be:

$$w_i = \frac{C_i}{\sum_i C_i}$$

### 1.3 Arbitrage opportunity

If we sell a straddle, we will win money when implied volatility decrease. When we buy a straddle, we will win money when implied volatility increase. Therefore we want the volatility to be low in the index and to be high among the components. Therefore in this instruments, Vega is a very important indicator.

## 2. Selection of data

### 2.1 Time range

We decided to select data from 2012 to 2017. However, the components are changing over time within the Dow Jones Index. There are 37 companies involved in those years, so we have to get (37+1) pairs of data. And then select the ones we need in different time period through an if-statement in the code.

- *On September 24, 2012, UnitedHealth Group replaced Kraft Foods Inc following Kraft's split into Mondelfiz International and Kraft Foods Group.*
- *On September 20, 2013, Goldman Sachs, Nike, and Visa replaced Alcoa, Bank of America, and Hewlett-Packard. Visa replaced Hewlett-Packard because of the split into HP Inc. and Hewlett Packard Enterprise[10][11]*
- *On March 19, 2015, Apple replaced AT&T, which had been a component of the DJIA since November 1916.[12] Apple became the fourth company traded on the NASDAQ to be part of the Dow.*
- *On September 1, 2017, DowDuPont replaced DuPont. DowDuPont was formed by the merger of Dow Chemical Company with DuPont.[13]*

### 2.2 The selection of options

We use options with at the money strike price and maturities of 3M from daily data. When there is no 3M data, we use options that are approximately 87-95 days before maturity. To select the option that we want form what we derived from WORD, we used python to pre-process data.

```python
import pandas as pd
import numpy as np
import os
from datetime import datetime
from dateutil.parser import parse
import pandas as pd
df=pd.read_csv("F:/file/input.csv")

data=pd.DataFrame(df)
data.describe()
data4use=data.drop(['issuer', 'optionid','cp_flag','exercise_style','index_flag','best_bid','best_offer'], axis=1)
#data4use=data.drop(['best_bid','best_offer'], axis=1)
data4use["exdate"]=pd.to_datetime(data4use["exdate"])
data4use["date"]=pd.to_datetime(data4use["date"])
data4use["price"]=(data['best_bid']+data['best_offer'])/2
data4use["T"]=data4use["exdate"]-data4use["date"]
data4use["T"]=data4use["T"]/np.timedelta64(1,'D')
data4use["strike_price"]=data4use["strike_price"]/1000
data4use=data4use.loc[(data4use['vega'] >0)]

data4output=data4use.loc[(data4use['date'] >= '2015-01-01')]
#data4output=data4use.loc[(data4use['T'] <=92)&(data4use['T'] >=50)]
#data4output.to_csv('F:/file/data4output.csv')
#data4output=pd.read_csv("F:/file/data4output.csv")
alldata=data4output
stockriskfree=pd.read_csv("F:/file/alldata.csv")
alldata=alldata.loc[(alldata['T'] <=199)&(alldata['T'] >=58)]
riskfree=stockriskfree["r"]
stockriskfree=stockriskfree.drop(columns=['r'])
stock=stockriskfree
stock=stock.loc[(stock['Date'] >= '2015/1/1')]
#alldata=alldata.loc[(alldata['ticker'] != 'B')]
grouped = alldata.groupby("ticker")
stock["Date"]=pd.to_datetime(stock["Date"])
output = pd.DataFrame()
for tick, group in grouped:
        groupedsub = group.groupby("date")
        for time, groupsub in groupedsub:
                #print(tick)
                #print(time)

                tempprice=(stock[(stock['Date']==time)][tick])
                output=output.append(groupsub.iloc[(groupsub['strike_price']-float(tempprice)).abs().argsort()[:1]])

os.chdir("F:/file/")
output.to_csv('output.csv')
```

The file after preprocessing is "output.xlsx" , and all the information about stock prices are saved in "stock.xlsx", which are attached in the file.

# 3. Methodology

### 3.1 Put-call-purity

The pool of options is too large so it will be time-consuming and space-consuming to have both the data of the put options and the call options. Therefore we decided to apply put-call-purity in the model. Assume that put-call-purity for European Option holds, and the American calls always has the same price of European calls in ideal market. We can derive the European Put Option's price by:

$$P = S + C - K \cdot e^{-r(T-t)}$$

Also, according to Black-Scholes formula and put-call purity, the Vega of the put option should be the same as the call option, which is:

$$\frac{dP}{d\sigma_{put}} = \frac{dP}{d\sigma_{call}}$$

And the Theta of put and call has relationship:

$$\frac{dP}{d\Theta_{put}} = \frac{dP}{d\Theta_{call}} + re^{-r(T-t)}$$

With those formula, all the data we need for the put option can be derived from call options.

### 3.2 Vega neutral and Theta neutral

$$\frac{dP}{d\sigma_P} = k \sum_i \frac{dC_i}{d\sigma_{C_i}} \cdot w_i$$
$$\frac{dP}{d\Theta_P} = k \sum_i \frac{dC_i}{d\Theta_{C_i}} \cdot w_i$$

There are two ways of allocating weights between index straddle and the components straddle. The first one is to apply Vega neutral calculation, where the Vega can be found in the database. Another method is to apply theta neutral calculation, where the Theta can be found in the database. We will compare the performance of two allocations.

### 3.3 Simulation Methods: FHS

Assume that the volatility of the portfolio can be described by GARCH(1,1) model, where

$$\mathrm{R_{it}} = z_{it} \cdot \sigma_t$$

$$\sigma_{t+1}^2 = (1 - \alpha - \beta)\sigma^2 + \alpha R_t^2 + \beta \sigma_t^2$$

$\sigma$ is the long-run volatility. Since $R_t$ is very close to normal distribution, so we use MLE method to decide what the parameter $\sigma, \sigma_1, \alpha$ and $\beta$ should be. Then we get $\sigma_1, \sigma_2, \ldots, \sigma_{t+1}$ from the model. Compute the standard $z_{t_i} = R_{t_i}/\sigma_{t_i}$ and then simulate future returns by randomly choose someone from historical $z_t$. Then the rescaled return that we simulate is:

$$R_{t+1} = \frac{R_{t_i}}{\sigma_{t_i}} \cdot \sigma_{t+1}$$

If we assume that the correlation among components and the straddle of index is constant, then this model has already captured the correlation because we use filtered historical data.

### 3.4 Risk measurement

The objective of the model is to forecast one day risk for the position we take by theta neutral method or Vega neutral method. Three parameters are used to measure the risk. The first one is Value at Risk (VaR) of the portfolio, which is 5 percent quantile of the worst behavior of the portfolio. The second one is Expected Shortfall (ES), which is the expected value of the 5 percent worst behavior of the portfolio. The third one is Delta with respect to index levels of that particular day.

# 4. Design of the system

We design two main functions to complete the task. The code is attached in the file.

```r
FHS<-function(tickername,time){

  # put-call parity to simulate put price
  sub_option<-OptionDataRA[which(OptionDataRA$ticker==tickername & OptionDataRA$date<=time),]
  sub_stock<-Stock[which(Stock$Date %in% sub_option$date),][tickername]
  put<-sub_stock+sub_option$price-sub_option$strike_price
  put<-put[,1]
  put[which(put<0)]=0.00001
  logreturn_put<-diff(log(put))
  logreturn_call=diff(log(sub_option$price))
  logreturn=logreturn_put+logreturn_call
  logreturn[which(logreturn>10)]=0
  logreturn[which(logreturn<(-10))]=0

  Garch<-function(x){
    sigmasqhat=rep(0,length(logreturn))
    sigmasqhat[1]=x[4]^2
    if(x[1])
      if (x[1]+x[2]>=1 || x[1]<0 || x[2]<0 || x[3]<0.001|| x[4]<0.001){
        NeglogLH = 9999
      } else {
        for (i in 1:(length(logreturn)-1)) {
          sigmasqhat[i+1] = (1-x[1]-x[2])*x[3]^2+x[1]*logreturn[i]^2+x[2]*sigmasqhat[i]
        }
        f <-(1/(sqrt(2*pi*sigmasqhat)))*exp(-0.5*logreturn^2/sigmasqhat)
        NeglogLH = -sum(log(f))
      }
    return(NeglogLH)
  }
  output = optim(c(0.1,0.8,0.01,0.0001),Garch)
  alpha = output$par[1]
  beta = output$par[2]
  sigma = output$par[3]
  sigma1 = output$par[4]
  sigmasqhatF=rep(0,length(logreturn))
  sigmasqhatF[1]=sigma1^2
  for (i in 1:(length(logreturn)-1)) {
    sigmasqhatF[i+1] = (1-alpha-beta)*sigma^2+alpha*logreturn[i]^2+beta*sigmasqhatF[i]
  }
  sigmanew<-sqrt((1-alpha-beta)*sigma^2+alpha*logreturn[length(logreturn)]^2+beta*sigmasqhatF[length(logreturn)])
  standardZ=logreturn/sigmasqhatF
  standardZ[!is.finite(standardZ)] <- NA
  standardZ= standardZ[!is.na(standardZ)]
  SimulationReturn<-sample(standardZ,1000,replace = TRUE)*sigmanew
  SimulationReturn
}
```

```r
PortfolioReturn<-function(date){

    ticker_select(date)
  price<-c()
  vega<-c()
  OptionPrice<-c()
  theta_temp<-c()
  theta<-c()
  for (i in ticker_list){
    price=c(price,as.numeric(St which(x, arr.ind = FALSE, useNames = TRUE) ]))
    vega=c(vega,OptionDataRA[which(OptionDataRA$date==date&OptionDataRA$ticker==i),]$vega*2)
    OptionPrice=c(OptionPrice,OptionDataRA[which(OptionDataRA$date==date&OptionDataRA$ticker==i),]$price)
    theta_temp=c(theta_temp,OptionDataRA[which(OptionDataRA$date==date&OptionDataRA$ticker==i),]$theta)
  }
  theta=theta_temp*2+r*exp(-0.25*r)/0.25
  kvega=sum(weight(price)*vega)/OptionDataRA[which(OptionDataRA$date==date&OptionDataRA$ticker=="DJX"),]$vega
  ktheta=sum(weight(price)*theta)/OptionDataRA[which(OptionDataRA$date==date&OptionDataRA$ticker=="DJX"),]$theta
  print(sprintf("for vega-neutral portfolio k is : %.2f ",kvega))
  print(sprintf("the weight of %s among all the components is %.2f",ticker_list,weight(price)))
  print(sprintf("for theta-neutral portfolio k is : %.2f ",kvega))
  print(sprintf("the weight of %s among all the components is %.2f",ticker_list,weight(price)))

  j=0
  ReturnVega=0
  ReturnTheta=0
  for (i in ticker_list){
    j=j+1
    tempvega=kvega*weight(price)[j]*FHS(i,date)
    temptheta=ktheta*weight(price)[j]*FHS(i,date)
    ReturnVega=tempvega+ReturnVega
    ReturnTheta=ReturnTheta+temptheta
  }
  ReturnVega=ReturnVega-FHS("DJX",date)
  ReturnTheta=ReturnTheta-FHS("DJX",date)
  DeltaS=Stock[which(Stock$Date==date),]$DJX- Stock[which(Stock$Date==date)-1,]$DJX

  VaRVega=quantile(ReturnVega,0.05)
  VaRTheta=quantile(ReturnTheta,0.05)

  EVega=mean(ReturnVega)
  ETheta=mean(ReturnTheta)

  ESVega=mean(ReturnVega[ReturnVega<=VaRVega])
  ESTheta=mean(ReturnTheta[ReturnTheta<=VaRTheta])

  DeltaVega=mean(ReturnVega)/DeltaS
  DeltaTheta=mean(ReturnTheta)/DeltaS
```

## 5. Results

```
> PortfolioReturn("2017-04-03")
[1] "for vega-neutral portfolio k is : 1.07 "
 [1] "the weight of MMM among all the components is 0.07"  "the weight of AXP among all the components is 0.03"
 [3] "the weight of T among all the components is 0.01"    "the weight of BA among all the components is 0.07"
 [5] "the weight of CAT among all the components is 0.04"  "the weight of CVX among all the components is 0.04"
 [7] "the weight of CSCO among all the components is 0.01" "the weight of KO among all the components is 0.02"
 [9] "the weight of DIS among all the components is 0.04"  "the weight of XOM among all the components is 0.03"
[11] "the weight of AA among all the components is 0.01"   "the weight of HD among all the components is 0.06"
[13] "the weight of IBM among all the components is 0.07"  "the weight of INTC among all the components is 0.01"
[15] "the weight of JNJ among all the components is 0.05"  "the weight of JPM among all the components is 0.03"
[17] "the weight of MCD among all the components is 0.05"  "the weight of MRK among all the components is 0.02"
[19] "the weight of MSFT among all the components is 0.03" "the weight of HPQ among all the components is 0.01"
[21] "the weight of PFE among all the components is 0.01"  "the weight of PG among all the components is 0.03"
[23] "the weight of TRV among all the components is 0.05"  "the weight of UTX among all the components is 0.04"
[25] "the weight of UNH among all the components is 0.07"  "the weight of VZ among all the components is 0.02"
[27] "the weight of BAC among all the components is 0.01"  "the weight of WMT among all the components is 0.03"
[29] "the weight of WBA among all the components is 0.03"
[1] "for theta-neutral portfolio k is : 1.07 "
 [1] "the weight of MMM among all the components is 0.07"  "the weight of AXP among all the components is 0.03"
 [3] "the weight of T among all the components is 0.01"    "the weight of BA among all the components is 0.07"
 [5] "the weight of CAT among all the components is 0.04"  "the weight of CVX among all the components is 0.04"
 [7] "the weight of CSCO among all the components is 0.01" "the weight of KO among all the components is 0.02"
 [9] "the weight of DIS among all the components is 0.04"  "the weight of XOM among all the components is 0.03"
[11] "the weight of AA among all the components is 0.01"   "the weight of HD among all the components is 0.06"
[13] "the weight of IBM among all the components is 0.07"  "the weight of INTC among all the components is 0.01"
[15] "the weight of JNJ among all the components is 0.05"  "the weight of JPM among all the components is 0.03"
[17] "the weight of MCD among all the components is 0.05"  "the weight of MRK among all the components is 0.02"
[19] "the weight of MSFT among all the components is 0.03" "the weight of HPQ among all the components is 0.01"
[21] "the weight of PFE among all the components is 0.01"  "the weight of PG among all the components is 0.03"
[23] "the weight of TRV among all the components is 0.05"  "the weight of UTX among all the components is 0.04"
[25] "the weight of UNH among all the components is 0.07"  "the weight of VZ among all the components is 0.02"
[27] "the weight of BAC among all the components is 0.01"  "the weight of WMT among all the components is 0.03"
[29] "the weight of WBA among all the components is 0.03"
                                                          5%                   5%
names  "Mean(Vega)"          "Mean(Theta)"         "VaR(Vega)"          "VaR(Theta)"         "ES(Vega)"
output "-0.0348223731387228" "-0.0165249375018926" "-1.34354642699037" "-1.60782298990385" "-2.1779623906158"

names  "ES(Theta)"         "Delta(Vega)"       "Delta(Theta)"
output "-2.76592898628129" "0.267663455272929" "0.127019541497223"
```

The result shows the risk measurement in the situation that we write one share of Dow Jones straddle and buy components straddle with weights we compute from either vega-neutral method or theta-neutral method. Apparently that with Vega neutral method the model forecast a lower risk, but with theta-neutral method, the sensitivity of portfolio to index level is much smaller. We can also conclude that the VaR and ES loss is actually severe respect to the money we invest.

## 6. Risk not captured

The changes of correlation among components are not considered in the model. A better model with DCC can be used to fix this problem. And one-day forecasting is not enough, the term structure of risk should be simulated and considered to improve the model. Also, the parameter of GARCH model can be defined and selected in many ways, which may cause small difference in the result. The put-call purity will not actually hold in reality, which is a big risk that the portfolio is exposed to. As well as the estimation of vega and theta, which can be inaccurate in reality too.