# DAT537 Final Project

Adrian Cao

2022-12-02

**DAT 537 Final Project**

Mingxuan(Adrian) Cao BA 2024

# Q1

I randomly selected 30 stocks from the S&P 500 companies, there are:

W. R. Berkley Corporation WRB

Expeditors International of Washington Inc. EXPD

Caterpillar Inc. CAT

Quanta Services Inc. PWR

O'Reilly Auto Parts ORLY

Prudential Financial Inc. PRU

Aflac Incorporated AFL

Walmart WMT

Becton Dickinson and Company BDX

Southwest Airlines LUV

NetApp Inc. NTAP

PulteGroup PHM

Verizon Communications Inc. VZ

Weyerhaeuser Company WY

FMC Corporation FMC

Pinnacle West Capital Corporation PNW

Baker Hughes Company Class A BKR

Rollins, Inc. ROL

Regions Financial Corporation RF

Citigroup C

Allstate Corporation ALL

Evergy Inc. EVRG

Illinois Tool Works Inc. ITW

The Estée Lauder Companies Inc. EL

Match Group Inc. MTCH

STERIS Plc STE

FedEx FDX

Wells Fargo & Company WFC

Newmont Corporation NEM

Biogen Inc. BIIB

# Q2

```
datdf = getfinmdat(symbols = c("WRB","EXPD","CAT","PWR","ORLY",
                               "PRU","AFL","WMT","BDX","LUV",
                               "NTAP","PHM","VZ","WY","FMC",
                               "PNW","BKR","ROL","RF","C",
                               "ALL","EVRG","ITW","EL","MTCH",
                               "STE","FDX","WFC","NEM","BIIB"),
                   symnames = c("Berkley","Expeditors","Caterpillar","Quanta","OReilly",
                                "Prudential","Aflac","Walmart","BectonDickinson","Southwest",
                                "NetApp","PulteGroup","Verizon","Weyerhaeuser","FMC",
                                "PinnacleWest","BakerHughes","Rollins","Regions","Citigroup",
                                "Allstate","Evergy","Illinois","EstéeLauder","Match",
                                "STERIS","FedEx","WellsFargo","Newmont","Biogen"),
                   from = as.Date("2004-12-31"),
                   to = as.Date("2018-12-31"))
```

```
## Warning: ^IRX contains missing values. Some functions will not work if objects
## contain missing values in the middle of the series. Consider using na.omit(),
## na.approx(), na.fill(), etc to remove or replace them.
```

```
## Warning in to.period(datxts, indexAt = "endof", period = "months"): missing
## values removed from data
```

```
head(datdf)
```

```
##              prmBerkley prmExpeditors prmCaterpillar   prmQuanta  prmOReilly
## 2005-01-31  0.009410642   0.002828004    -0.08405630 -0.06682497  0.01326937
## 2005-02-28  0.074703373  -0.013961067     0.06475210  0.04743877  0.11102841
## 2005-03-31 -0.035080076  -0.036865363    -0.04023134 -0.03027735 -0.02916737
## 2005-04-29 -0.019416019  -0.085191677    -0.03484684  0.04359302  0.03386108
## 2005-05-31  0.088701178   0.038688644     0.06644868  0.12920329  0.08024349
## 2005-06-30  0.005208275  -0.025363396     0.01033853 -0.02788392  0.07138079
##             prmPrudential    prmAflac   prmWalmart prmBectonDickinson
## 2005-01-31  -0.020929654 -0.01010783 -0.009776563      -0.0044653870
## 2005-02-28   0.055290750 -0.02919359 -0.017102763       0.0548137976
## 2005-03-31   0.004766129 -0.03016730 -0.028453507      -0.0235022920
## 2005-04-29  -0.006634460  0.08870347 -0.061548516      -0.0005676212
## 2005-05-31   0.105410823  0.02254971  0.002741555      -0.0206595305
## 2005-06-30   0.034706150  0.03922372  0.018124691      -0.0861613810
##             prmSouthwest   prmNetApp prmPulteGroup  prmVerizon prmWeyerhaeuser
## 2005-01-31  -0.11238995 -0.04336632    0.03391180 -0.11474262   -0.0735295423
## 2005-02-28  -0.04553473 -0.05950127    0.17866303  0.00865057    0.0774450730
## 2005-03-31   0.02627821 -0.08055894   -0.05786560 -0.01531849    0.0212060134
## 2005-04-29   0.04266534 -0.03807063   -0.03188634  0.01772109   -0.0006729582
## 2005-05-31  -0.02455287  0.07598993    0.06760339 -0.01410732   -0.0604311747
## 2005-06-30  -0.04469794 -0.01945072    0.10027576 -0.02587283   -0.0102070597
##                 prmFMC prmPinnacleWest prmBakerHughes  prmRollins   prmRegions
## 2005-01-31 -0.02480623   -0.0522772966     0.01293957 -0.05957515 -0.093290064
## 2005-02-28  0.04395804   -0.0008274181     0.09273603  0.03010807  0.006098514
## 2005-03-31  0.08060846    0.0159518795    -0.06126201  0.09100872  0.002088274
## 2005-04-29 -0.08553398   -0.0054823879    -0.01059485  0.05901118  0.042042256
## 2005-05-31  0.12925771    0.0506077944     0.04724179  0.03367377  0.003297820
## 2005-06-30  0.01003003    0.0050665272     0.10518601 -0.02005982  0.003524821
##             prmCitigroup prmAllstate    prmEvergy  prmIllinois prmEstéeLauder
## 2005-01-31  0.0162324568 -0.02657370  0.016977076 -0.063326824   -0.015589715
## 2005-02-28 -0.0204483565  0.06220837 -0.015761150  0.029819684   -0.027724295
```

```
## 2005-03-31 -0.0605083291  0.01077206 -0.051055233 -0.001615127    0.020486174
## 2005-04-29  0.0525346540  0.03656697  0.055947082 -0.066056756   -0.148343713
## 2005-05-31  0.0008189227  0.03969554  0.005048337  0.004902123    0.015328569
## 2005-06-30 -0.0210933732  0.02421935  0.049592549 -0.055358624   -0.001390366
##               prmMatch    prmSTERIS    prmFedEx prmWellsFargo  prmNewmont
## 2005-01-31 -0.12456227 -0.001824974 -0.03066073 -1.550218e-02 -0.06532424
## 2005-02-28 -0.07301278  0.041396751  0.02024219 -2.576499e-02  0.07996442
## 2005-03-31 -0.01291360  0.017949970 -0.04075302  4.821322e-03 -0.06127128
## 2005-04-29 -0.02607775 -0.064456840 -0.09807439  6.211244e-05 -0.10358099
## 2005-05-31  0.12457990  0.021289329  0.05024401  1.355423e-02 -0.01890608
## 2005-06-30 -0.02200543  0.062462793 -0.09565743  1.695476e-02  0.04565343
##              prmBiogen     hml     smb     rmw     cma
## 2005-01-31 -0.02659606  0.0206 -0.0118  0.0273 -0.0146
## 2005-02-28 -0.40704495  0.0153 -0.0030  0.0145 -0.0005
## 2005-03-31 -0.10936700  0.0204 -0.0142  0.0047  0.0129
## 2005-04-29  0.04785184  0.0006 -0.0398  0.0097 -0.0094
## 2005-05-31  0.07654280 -0.0064  0.0279 -0.0100  0.0030
## 2005-06-30 -0.12133903  0.0282  0.0328  0.0097 -0.0051
```

# Q3

```
data(factor12)
data = factor12
nug = c(4,5,6)

thetamls = mapply(FUN="CZZscant",
                  nu = nug,
                  MoreArgs = list(data = data,mustinclude = "Mkt",m = 2000,nclust = 4),
                  SIMPLIFY = FALSE)
```

```
## starting Chib, Zeng and Zhao (2020) model scan with student-t errors ...
## there are 2047 models in the model space
```

```
## Loading required package: snow
```

```
## model scan started ...
## model scan finished ...
## preparing output ...
## starting Chib, Zeng and Zhao (2020) model scan with student-t errors ...
## there are 2047 models in the model space
## model scan started ...
## model scan finished ...
## preparing output ...
## starting Chib, Zeng and Zhao (2020) model scan with student-t errors ...
## there are 2047 models in the model space
## model scan started ...
## model scan finished ...
## preparing output ...
```

```
scanord = list()
scanord[[1]] = thetamls[[1]]$scanord
scanord[[2]] = thetamls[[2]]$scanord
scanord[[3]] = thetamls[[3]]$scanord

scanloglik = c()
```

```
scanloglik[1] = scanord[[1]][1,]['logmarg']
scanloglik[2] = scanord[[2]][1,]['logmarg']
scanloglik[3] = scanord[[3]][1,]['logmarg']
#nu = 5 is the best out of these three

xbest = names(which(scanord[[which.max(scanloglik)]][1,]==1))
xbest
```

```
## [1] "Mkt"  "SMB"  "ROE"  "PEAD" "MGMT"
```

# Q4

```
data = factor12[373:540,]
factor22 = cbind(datdf[,1:30],data)
colnames(factor22)
```

```
##  [1] "prmBerkley"         "prmExpeditors"      "prmCaterpillar"
##  [4] "prmQuanta"          "prmOReilly"         "prmPrudential"
##  [7] "prmAflac"           "prmWalmart"         "prmBectonDickinson"
## [10] "prmSouthwest"       "prmNetApp"          "prmPulteGroup"
## [13] "prmVerizon"         "prmWeyerhaeuser"    "prmFMC"
## [16] "prmPinnacleWest"    "prmBakerHughes"     "prmRollins"
## [19] "prmRegions"         "prmCitigroup"       "prmAllstate"
## [22] "prmEvergy"          "prmIllinois"        "prmEstéeLauder"
## [25] "prmMatch"           "prmSTERIS"          "prmFedEx"
## [28] "prmWellsFargo"      "prmNewmont"         "prmBiogen"
## [31] "Mkt"                "SMB"                "HML"
## [34] "RMW"                "CMA"                "MOM"
## [37] "IA"                 "ROE"                "PEAD"
## [40] "FIN"                "MGMT"               "PERF"
```

# Q5

```
pout = pricing(xnames = xbest,
               data = factor22)
```

```
## starting Bayesian pricing test ...
## there are 37 assets in the data set
## pricing of assets started ...
## computations finished ...
## preparing output ...
```

```
pout
```

```
##                   logmarg0 logmarg1   diffM0mM1 priced2:1 priced3:1 priced4:1
## prmBerkley         237.7574 235.7629  1.99447532         1         1         1
## prmExpeditors      189.9998 187.8480  2.15183199         1         1         1
## prmCaterpillar     181.1443 179.7757  1.36851779         1         1         0
## prmQuanta          160.4284 158.2217  2.20675804         1         1         1
## prmOReilly         190.4204 190.8519 -0.43146815         0         0         0
## prmPrudential      167.4915 166.4332  1.05832374         1         0         0
## prmAflac           174.1605 173.1466  1.01395910         1         0         0
## prmWalmart         240.2974 238.3269  1.97044840         1         1         1
```

```
## prmBectonDickinson 237.9117 236.5991  1.31262710         1         1         0
## prmSouthwest       160.0869 158.6253  1.46168533         1         1         1
## prmNetApp          144.2322 142.4021  1.83009132         1         1         1
## prmPulteGroup      110.1588 108.5677  1.59111805         1         1         1
## prmVerizon         230.0649 228.4985  1.56636513         1         1         1
## prmWeyerhaeuser    187.9307 186.5003  1.43038981         1         1         1
## prmFMC             190.1145 189.9548  0.15961614         0         0         0
## prmPinnacleWest    222.7100 221.8223  0.88768777         1         0         0
## prmBakerHughes     141.0880 139.3043  1.78366017         1         1         1
## prmRollins         222.3032 223.6552 -1.35200070         0         0         0
## prmRegions         125.9727 125.3411  0.63164915         0         0         0
## prmCitigroup       115.9097 116.6235 -0.71383398         0         0         0
## prmAllstate        203.1006 201.6123  1.48832907         1         1         1
## prmEvergy          234.9158 234.1546  0.76116670         1         0         0
## prmIllinois        252.7317 250.8677  1.86397818         1         1         1
## prmEstéeLauder     190.6472 190.4554  0.19178550         0         0         0
## prmMatch           140.9347 140.4893  0.44539996         0         0         0
## prmSTERIS          194.5259 193.6609  0.86500824         1         0         0
## prmFedEx           207.0913 205.3647  1.72659669         1         1         1
## prmWellsFargo      183.6324 182.7661  0.86632677         1         0         0
## prmNewmont         109.2832 108.4144  0.86877759         1         0         0
## prmBiogen          135.2575 134.2709  0.98662015         1         0         0
## HML                353.3944 350.6114  2.78296939         1         1         1
## RMW                406.4091 407.6872 -1.27814669         0         0         0
## CMA                424.9476 423.4273  1.52024947         1         1         1
## MOM                288.8872 288.8576  0.02959259         0         0         0
## IA                 421.7434 420.1455  1.59786326         1         1         1
## FIN                365.0065 363.9270  1.07942283         1         0         0
## PERF               311.2233 311.8414 -0.61806785         0         0         0
```

```
wp = sum(pout[1:30,3] > .69)
wp
```

```
## [1] 23
```

```
wnp = sum(pout[1:30,3] < -.69)
wnp
```

```
## [1] 2
```