Problem2: Tests of Return Predictability

Limin Peng

2023-02-08

```
# import dataset
setwd("C:\\Users\\admin\\OneDrive\\LSE\\FM437 Financial Econometrics\\P2[Tests of Predictability]")

library(readr)
# monthly FF Factors
FF_m <- read_csv("F-F_Research_Data_Factors.csv")
```

```
## Rows: 1158 Columns: 5
## -- Column specification --------------------------
## Delimiter: ","
## dbl (5): ym, Mkt-RF, SMB, HML, RF
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# daily FF Factors
FF_d <- read_csv("F-F_Research_Data_Factors_daily.csv")
```

```
## Rows: 25399 Columns: 5
## -- Column specification --------------------------
## Delimiter: ","
## dbl (5): dt, Mkt-RF, SMB, HML, RF
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# monthly size portfolios
size_m <- read_csv("Portfolios_Formed_on_ME_Wout_Div.csv")
```

```
## Rows: 1158 Columns: 20
## -- Column specification --------------------------
## Delimiter: ","
## dbl (20): ym, <= 0, Lo 30, Med 40, Hi 30, Lo...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# daily size portfolios
size_d <- read_csv("Portfolios_Formed_on_ME_daily.csv")
```

```
## Rows: 25399 Columns: 20
## -- Column specification --------------------------
## Delimiter: ","
## dbl (20): dt, <= 0, Lo 30, Med 40, Hi 30, Lo...
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

(a) CRSP Market Returns

```
#  sample mean, standard deviation, autocorrelations up to lag 5, Box-Pierce Q-statistic with five auto
get_stat<- function(ts){
    T <- length(ts)
    mean <- mean(ts)
    sd <- (sum((ts-mean)^2)/(T-1))^0.5
    auto_corr <- acf(ts,lag.max=5, plot=FALSE)$acf[1:6]
    auto_corr_sq <- auto_corr^2
    BP_Q = T*sum(auto_corr_sq[-1])

    ans <- list(mean, sd, auto_corr, BP_Q)
    names(ans) <- c("mean","sd","auto_corr","BP_Q")

    return(ans)
    }

# market returns
Rm_m = FF_m$`Mkt-RF` + FF_m$RF
Rm_d = FF_d$`Mkt-RF` + FF_d$RF

stat.Rm_m <- get_stat(Rm_m)
stat.Rm_d <- get_stat(Rm_d)
```

```
# CRSP market returns, daily
print(c("CRSP market returns, daily",stat.Rm_d))
```

```
## [[1]]
## [1] "CRSP market returns, daily"
##
## $mean
## [1] 0.04183858
##
## $sd
## [1] 1.08263
##
## $auto_corr
## [1]  1.0000000000  0.0479173255 -0.0272699678
## [4]  0.0005158829  0.0163854148  0.0063774573
##
## $BP_Q
## [1] 85.06484
```

```
# CRSP market returns, monthly
print(c("CRSP market returns, monthly",stat.Rm_m))
```

```
## [[1]]
## [1] "CRSP market returns, monthly"
##
## $mean
## [1] 0.9340155
##
```

```
## $sd
## [1] 5.342443
##
## $auto_corr
## [1]  1.00000000  0.09194058 -0.02536270
## [4] -0.08495695  0.01841728  0.05792049
##
## $BP_Q
## [1] 23.16926
```

(b) 5 Size Portfolios

```
# 5 size portfolios, daily
stat.size_d.Lo20 <- get_stat(size_d$`Lo 20`)
stat.size_d.Qnt2 <- get_stat(size_d$`Qnt 2`)
stat.size_d.Qnt3 <- get_stat(size_d$`Qnt 3`)
stat.size_d.Qnt4 <- get_stat(size_d$`Qnt 4`)
stat.size_d.Hi20 <- get_stat(size_d$`Hi 20`)

ls1 <- list(stat.size_d.Lo20,stat.size_d.Qnt2,stat.size_d.Qnt3,stat.size_d.Qnt4,stat.size_d.Hi20)
names(ls1) <- c("Lo 20","Qnt 2","Qnt 3","Qnt 4","Hi 20")
print(ls1)
```

```
## $`Lo 20`
## $`Lo 20`$mean
## [1] 0.04938659
##
## $`Lo 20`$sd
## [1] 1.27372
##
## $`Lo 20`$auto_corr
## [1] 1.00000000 0.15307237 0.03796883 0.05145236
## [5] 0.04590589 0.05560256
##
## $`Lo 20`$BP_Q
## [1] 831.033
##
##
## $`Qnt 2`
## $`Qnt 2`$mean
## [1] 0.04986299
##
## $`Qnt 2`$sd
## [1] 1.237926
##
## $`Qnt 2`$auto_corr
## [1] 1.000000000 0.133836035 0.006931077
## [4] 0.035407042 0.028569565 0.022899115
##
## $`Qnt 2`$BP_Q
## [1] 522.0605
##
##
## $`Qnt 3`
```

```
## $`Qnt 3`$mean
## [1] 0.04868971
##
## $`Qnt 3`$sd
## [1] 1.184538
##
## $`Qnt 3`$auto_corr
## [1]  1.000000000  0.126744141 -0.007370692
## [4]  0.023752563  0.024385977  0.024114071
##
## $`Qnt 3`$BP_Q
## [1] 453.5945
##
##
## $`Qnt 4`
## $`Qnt 4`$mean
## [1] 0.04663884
##
## $`Qnt 4`$sd
## [1] 1.140958
##
## $`Qnt 4`$auto_corr
## [1]  1.000000000  0.105821564 -0.020256938
## [4]  0.017341507  0.024565822  0.008464187
##
## $`Qnt 4`$BP_Q
## [1] 319.6311
##
##
## $`Hi 20`
## $`Hi 20`$mean
## [1] 0.04109414
##
## $`Hi 20`$sd
## [1] 1.08692
##
## $`Hi 20`$auto_corr
## [1]  1.000000000  0.023698097 -0.031461329
## [4] -0.004889072  0.010271554  0.004295067
##
## $`Hi 20`$BP_Q
## [1] 43.15977
```

```
# 5 size portfolios, monthly
stat.size_m.Lo20 <- get_stat(size_m$`Lo 20`)
stat.size_m.Qnt2 <- get_stat(size_m$`Qnt 2`)
stat.size_m.Qnt3 <- get_stat(size_m$`Qnt 3`)
stat.size_m.Qnt4 <- get_stat(size_m$`Qnt 4`)
stat.size_m.Hi20 <- get_stat(size_m$`Hi 20`)

ls2 <- list(stat.size_m.Lo20,stat.size_m.Qnt2,stat.size_m.Qnt3,stat.size_m.Qnt4,stat.size_m.Hi20)
names(ls2) <- c("Lo 20","Qnt 2","Qnt 3","Qnt 4","Hi 20")
print(ls2)
```

```
## $`Lo 20`
```

```
## $`Lo 20`$mean
## [1] 1.096097
##
## $`Lo 20`$sd
## [1] 8.888405
##
## $`Lo 20`$auto_corr
## [1]  1.00000000  0.19371750  0.01882251
## [4] -0.06776915 -0.07780082 -0.04105286
##
## $`Lo 20`$BP_Q
## [1] 58.14517
##
##
## $`Qnt 2`
## $`Qnt 2`$mean
## [1] 0.961943
##
## $`Qnt 2`$sd
## [1] 7.580598
##
## $`Qnt 2`$auto_corr
## [1]  1.000000000  0.182276180 -0.002934314
## [4] -0.094975950 -0.073526678 -0.021256815
##
## $`Qnt 2`$BP_Q
## [1] 55.71332
##
##
## $`Qnt 3`
## $`Qnt 3`$mean
## [1] 0.8734542
##
## $`Qnt 3`$sd
## [1] 6.845182
##
## $`Qnt 3`$auto_corr
## [1]  1.00000000  0.15079838 -0.01434129
## [4] -0.09190037 -0.03710258  0.01158412
##
## $`Qnt 3`$BP_Q
## [1] 38.10086
##
##
## $`Qnt 4`
## $`Qnt 4`$mean
## [1] 0.7883592
##
## $`Qnt 4`$sd
## [1] 6.188394
##
## $`Qnt 4`$auto_corr
## [1]  1.00000000  0.12020764 -0.01101773
## [4] -0.08230086 -0.01042303  0.03160633
```

```
##
## $`Qnt 4`$BP_Q
## [1] 25.99976
##
##
## $`Hi 20`
## $`Hi 20`$mean
## [1] 0.601019
##
## $`Hi 20`$sd
## [1] 5.132905
##
## $`Hi 20`$auto_corr
## [1]  1.00000000  0.07046618 -0.02727821
## [4] -0.07677942  0.03095657  0.07181753
##
## $`Hi 20`$BP_Q
## [1] 20.5206
```

(c) Subsample Results

```
# split sample period
T_d = 12700 # half num. of daily  obs.
T_m = 579 # half num. of monthly  obs.
Rm_d1 <- Rm_d[1:T_d]
Rm_d2 <- Rm_d[-1:-T_d]
Rm_m1 <- Rm_m[1:T_m]
Rm_m2 <- Rm_m[-1:-T_m]

size_d1 <- size_d[1:T_d,]
size_d2 <- size_d[-1:-T_d,]
size_m1 <- size_d[1:T_m,]
size_m2 <- size_d[-1:-T_m,]

stat.Rm_d1 <- get_stat(Rm_d1)
stat.Rm_d2 <- get_stat(Rm_d2)
stat.Rm_m1 <- get_stat(Rm_m1)
stat.Rm_m2 <- get_stat(Rm_m2)
```

Market Returns

```
print(c("CRSP market returns subsample 1, daily",stat.Rm_d1))
```

Daily

```
## [[1]]
## [1] "CRSP market returns subsample 1, daily"
##
## $mean
## [1] 0.03877756
##
## $sd
## [1] 1.08517
##
```

```
## $auto_corr
## [1]  1.000000000  0.094720868 -0.048028667
## [4]  0.002481194  0.053069974  0.022474865
##
## $BP_Q
## [1] 185.5025
```

```
print(c("CRSP market returns subsample 2, daily",stat.Rm_d2))
```

```
## [[1]]
## [1] "CRSP market returns subsample 2, daily"
##
## $mean
## [1] 0.04489983
##
## $sd
## [1] 1.080117
##
## $auto_corr
## [1]  1.0000000000  0.0006613419 -0.0062979821
## [4] -0.0014786375 -0.0206237821 -0.0099102273
##
## $BP_Q
## [1] 7.185619
```

```
print(c("CRSP market returns subsample 1, monthly",stat.Rm_m1))
```

Monthly

```
## [[1]]
## [1] "CRSP market returns subsample 1, monthly"
##
## $mean
## [1] 0.8013299
##
## $sd
## [1] 6.014858
##
## $auto_corr
## [1]  1.000000000  0.138794579  0.002589833
## [4] -0.146328588  0.029751807  0.072144303
##
## $BP_Q
## [1] 27.08138
```

```
print(c("CRSP market returns subsample 2, monthly",stat.Rm_m2))
```

```
## [[1]]
## [1] "CRSP market returns subsample 2, monthly"
##
## $mean
## [1] 1.066701
##
## $sd
## [1] 4.573728
```

```
##
## $auto_corr
## [1]  1.00000000  0.02499583 -0.06743373
## [4]  0.02294084  0.00854866  0.05040574
##
## $BP_Q
## [1] 4.812764
```

Size Portfolio Returns

```
# subsample 1
stat.size_d1.Lo20 <- get_stat(size_d1$`Lo 20`)
stat.size_d1.Qnt2 <- get_stat(size_d1$`Qnt 2`)
stat.size_d1.Qnt3 <- get_stat(size_d1$`Qnt 3`)
stat.size_d1.Qnt4 <- get_stat(size_d1$`Qnt 4`)
stat.size_d1.Hi20 <- get_stat(size_d1$`Hi 20`)

ls1.1 <- list(stat.size_d1.Lo20,stat.size_d1.Qnt2,stat.size_d1.Qnt3,stat.size_d1.Qnt4,stat.size_d1.Hi20)
names(ls1.1) <- c("Lo 20","Qnt 2","Qnt 3","Qnt 4","Hi 20")
print(ls1.1)
```

Daily

```
## $`Lo 20`
## $`Lo 20`$mean
## [1] 0.05227402
##
## $`Lo 20`$sd
## [1] 1.411161
##
## $`Lo 20`$auto_corr
## [1] 1.00000000 0.21805484 0.01685494 0.05948348
## [5] 0.05531137 0.07138192
##
## $`Lo 20`$BP_Q
## [1] 755.9676
##
##
## $`Qnt 2`
## $`Qnt 2`$mean
## [1] 0.04973543
##
## $`Qnt 2`$sd
## [1] 1.261917
##
## $`Qnt 2`$auto_corr
## [1]  1.00000000  0.22530677 -0.02111244
## [4]  0.05455087  0.05149490  0.03669843
##
## $`Qnt 2`$BP_Q
## [1] 738.9263
##
##
## $`Qnt 3`
```

```
## $`Qnt 3`$mean
## [1] 0.04710394
##
## $`Qnt 3`$sd
## [1] 1.206095
##
## $`Qnt 3`$auto_corr
## [1]  1.00000000  0.19699282 -0.03418685
## [4]  0.03774973  0.05149174  0.04233880
##
## $`Qnt 3`$BP_Q
## [1] 582.2179
##
##
## $`Qnt 4`
## $`Qnt 4`$mean
## [1] 0.04281654
##
## $`Qnt 4`$sd
## [1] 1.161467
##
## $`Qnt 4`$auto_corr
## [1]  1.00000000  0.14407602 -0.04351866
## [4]  0.02815814  0.05612905  0.01891526
##
## $`Qnt 4`$BP_Q
## [1] 342.302
##
##
## $`Hi 20`
## $`Hi 20`$mean
## [1] 0.03762362
##
## $`Hi 20`$sd
## [1] 1.074666
##
## $`Hi 20`$auto_corr
## [1]  1.000000000  0.072562391 -0.046411988
## [4] -0.003076713  0.050556807  0.021144263
##
## $`Hi 20`$BP_Q
## [1] 132.4853
```

```
# subsample 2
stat.size_d2.Lo20 <- get_stat(size_d2$`Lo 20`)
stat.size_d2.Qnt2 <- get_stat(size_d2$`Qnt 2`)
stat.size_d2.Qnt3 <- get_stat(size_d2$`Qnt 3`)
stat.size_d2.Qnt4 <- get_stat(size_d2$`Qnt 4`)
stat.size_d2.Hi20 <- get_stat(size_d2$`Hi 20`)

ls1.2 <- list(stat.size_d2.Lo20,stat.size_d2.Qnt2,stat.size_d2.Qnt3,stat.size_d2.Qnt4,stat.size_d2.Hi20)
names(ls1.2) <- c("Lo 20","Qnt 2","Qnt 3","Qnt 4","Hi 20")
print(ls1.2)
```

```
## $`Lo 20`
```

```
## $`Lo 20`$mean
## [1] 0.04649894
##
## $`Lo 20`$sd
## [1] 1.119568
##
## $`Lo 20`$auto_corr
## [1] 1.00000000 0.04981595 0.07150739 0.03867837
## [5] 0.03096645 0.03051352
##
## $`Lo 20`$BP_Q
## [1] 139.4471
##
##
## $`Qnt 2`
## $`Qnt 2`$mean
## [1] 0.04999055
##
## $`Qnt 2`$sd
## [1] 1.213508
##
## $`Qnt 2`$auto_corr
## [1] 1.000000000 0.034913618 0.037272373
## [4] 0.014700952 0.003799523 0.007970352
##
## $`Qnt 2`$BP_Q
## [1] 36.85594
##
##
## $`Qnt 3`
## $`Qnt 3`$mean
## [1] 0.05027561
##
## $`Qnt 3`$sd
## [1] 1.162625
##
## $`Qnt 3`$auto_corr
## [1]  1.000000000  0.051132714  0.021504100
## [4]  0.008674398 -0.004770079  0.004502700
##
## $`Qnt 3`$BP_Q
## [1] 40.57653
##
##
## $`Qnt 4`
## $`Qnt 4`$mean
## [1] 0.05046145
##
## $`Qnt 4`$sd
## [1] 1.120105
##
## $`Qnt 4`$auto_corr
## [1]  1.000000000  0.064662890  0.004753788
## [4]  0.005690023 -0.009353791 -0.002801026
```

```
##
## $`Qnt 4`$BP_Q
## [1] 55.00703
##
##
## $`Hi 20`
## $`Hi 20`$mean
## [1] 0.04456493
##
## $`Hi 20`$sd
## [1] 1.099069
##
## $`Hi 20`$auto_corr
## [1]  1.000000000 -0.023034855 -0.017148364
## [4] -0.006633994 -0.028231239 -0.011864839
##
## $`Hi 20`$BP_Q
## [1] 22.94021
```

```
# subsample 1
stat.size_m1.Lo20 <- get_stat(size_m1$`Lo 20`)
stat.size_m1.Qnt2 <- get_stat(size_m1$`Qnt 2`)
stat.size_m1.Qnt3 <- get_stat(size_m1$`Qnt 3`)
stat.size_m1.Qnt4 <- get_stat(size_m1$`Qnt 4`)
stat.size_m1.Hi20 <- get_stat(size_m1$`Hi 20`)

ls2.1 <- list(stat.size_m1.Lo20,stat.size_m1.Qnt2,stat.size_m1.Qnt3,stat.size_m1.Qnt4,stat.size_m1.Hi20)
names(ls2.1) <- c("Lo 20","Qnt 2","Qnt 3","Qnt 4","Hi 20")
print(ls2.1)
```

Monthly

```
## $`Lo 20`
## $`Lo 20`$mean
## [1] 0.08196891
##
## $`Lo 20`$sd
## [1] 0.8503905
##
## $`Lo 20`$auto_corr
## [1]  1.000000000  0.043247735  0.015125417
## [4] -0.001506639  0.047152915  0.036274627
##
## $`Lo 20`$BP_Q
## [1] 3.265943
##
##
## $`Qnt 2`
## $`Qnt 2`$mean
## [1] 0.08314335
##
## $`Qnt 2`$sd
## [1] 0.6829451
##
```

```
## $`Qnt 2`$auto_corr
## [1]  1.000000000  0.100043822  0.008608057
## [4]  0.011348248 -0.014173440  0.008897249
##
## $`Qnt 2`$BP_Q
## [1] 6.074691
##
##
## $`Qnt 3`
## $`Qnt 3`$mean
## [1] 0.07666667
##
## $`Qnt 3`$sd
## [1] 0.6381646
##
## $`Qnt 3`$auto_corr
## [1]  1.000000000  0.078350812 -0.019041311
## [4] -0.011676083 -0.025224686 -0.006801915
##
## $`Qnt 3`$BP_Q
## [1] 4.238455
##
##
## $`Qnt 4`
## $`Qnt 4`$mean
## [1] 0.06958549
##
## $`Qnt 4`$sd
## [1] 0.6238269
##
## $`Qnt 4`$auto_corr
## [1]  1.000000000  0.058015193 -0.045968292
## [4]  0.004031675  0.027138767  0.018482406
##
## $`Qnt 4`$BP_Q
## [1] 3.80589
##
##
## $`Hi 20`
## $`Hi 20`$mean
## [1] 0.07908463
##
## $`Hi 20`$sd
## [1] 0.57882
##
## $`Hi 20`$auto_corr
## [1]  1.000000000  0.056679638 -0.044751749
## [4] -0.008760974  0.022456513  0.030788205
##
## $`Hi 20`$BP_Q
## [1] 3.904929
```

```
# subsample 2
stat.size_m2.Lo20 <- get_stat(size_m2$`Lo 20`)
```

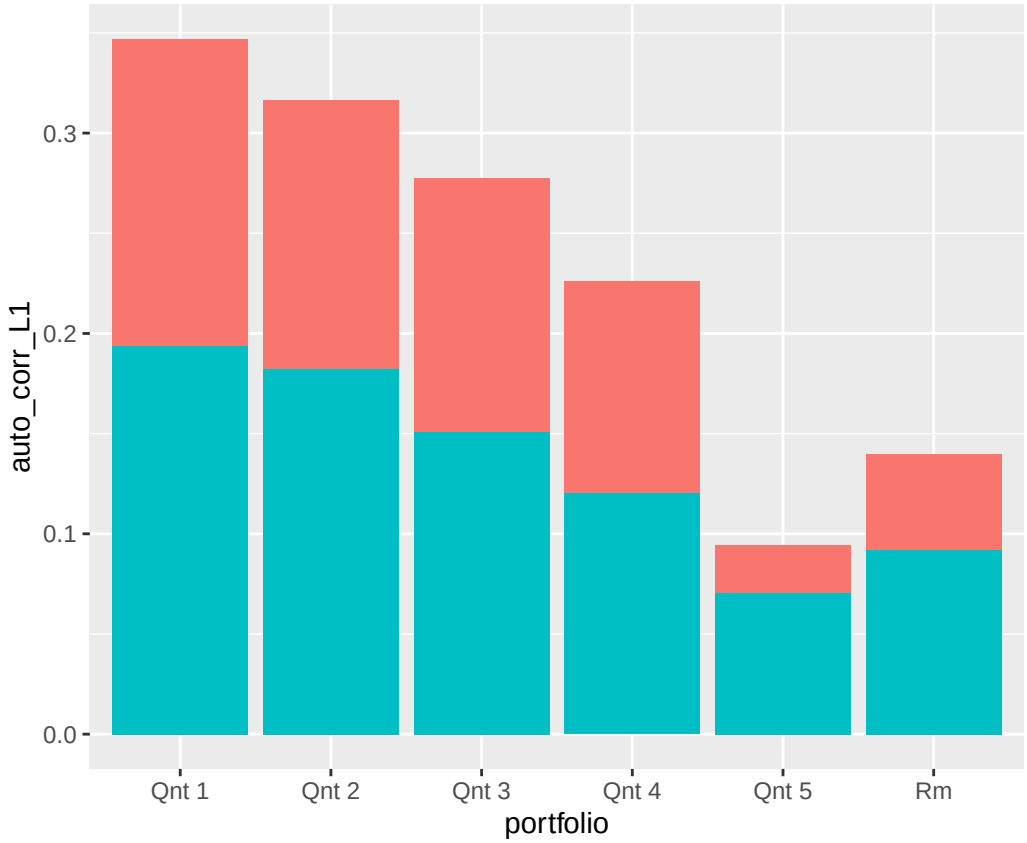
```r
stat.size_m2.Qnt2 <- get_stat(size_m2$`Qnt 2`)
stat.size_m2.Qnt3 <- get_stat(size_m2$`Qnt 3`)
stat.size_m2.Qnt4 <- get_stat(size_m2$`Qnt 4`)
stat.size_m2.Hi20 <- get_stat(size_m2$`Hi 20`)

ls2.2 <- list(stat.size_m2.Lo20,stat.size_m2.Qnt2,stat.size_m2.Qnt3,stat.size_m2.Qnt4,stat.size_m2.Hi20)
names(ls2.2) <- c("Lo 20","Qnt 2","Qnt 3","Qnt 4","Hi 20")
print(ls2.2)
```

```
## $`Lo 20`
## $`Lo 20`$mean
## [1] 0.04862651
##
## $`Lo 20`$sd
## [1] 1.28193
##
## $`Lo 20`$auto_corr
## [1] 1.00000000 0.15352318 0.03832560 0.05212224
## [5] 0.04585450 0.05605517
##
## $`Lo 20`$BP_Q
## [1] 819.0541
##
##
## $`Qnt 2`
## $`Qnt 2`$mean
## [1] 0.04908662
##
## $`Qnt 2`$sd
## [1] 1.247927
##
## $`Qnt 2`$auto_corr
## [1] 1.000000000 0.133592672 0.006963557
## [4] 0.035711485 0.028768784 0.023041744
##
## $`Qnt 2`$BP_Q
## [1] 509.5389
##
##
## $`Qnt 3`
## $`Qnt 3`$mean
## [1] 0.04803707
##
## $`Qnt 3`$sd
## [1] 1.194304
##
## $`Qnt 3`$auto_corr
## [1]  1.000000000  0.126688969 -0.007214336
## [4]  0.024034152  0.024727219  0.024239021
##
## $`Qnt 3`$BP_Q
## [1] 443.7505
##
##
```

```
## $`Qnt 4`
## $`Qnt 4`$mean
## [1] 0.04610355
##
## $`Qnt 4`$sd
## [1] 1.150252
##
## $`Qnt 4`$auto_corr
## [1]  1.000000000  0.105840211 -0.019976035
## [4]  0.017568440  0.024557749  0.008375062
##
## $`Qnt 4`$BP_Q
## [1] 312.3117
##
##
## $`Hi 20`
## $`Hi 20`$mean
## [1] 0.0402079
##
## $`Hi 20`$sd
## [1] 1.095955
##
## $`Hi 20`$auto_corr
## [1]  1.000000000  0.023325980 -0.031276072
## [4] -0.004810460  0.010212301  0.004051982
##
## $`Hi 20`$BP_Q
## [1] 41.3537
```

Summary

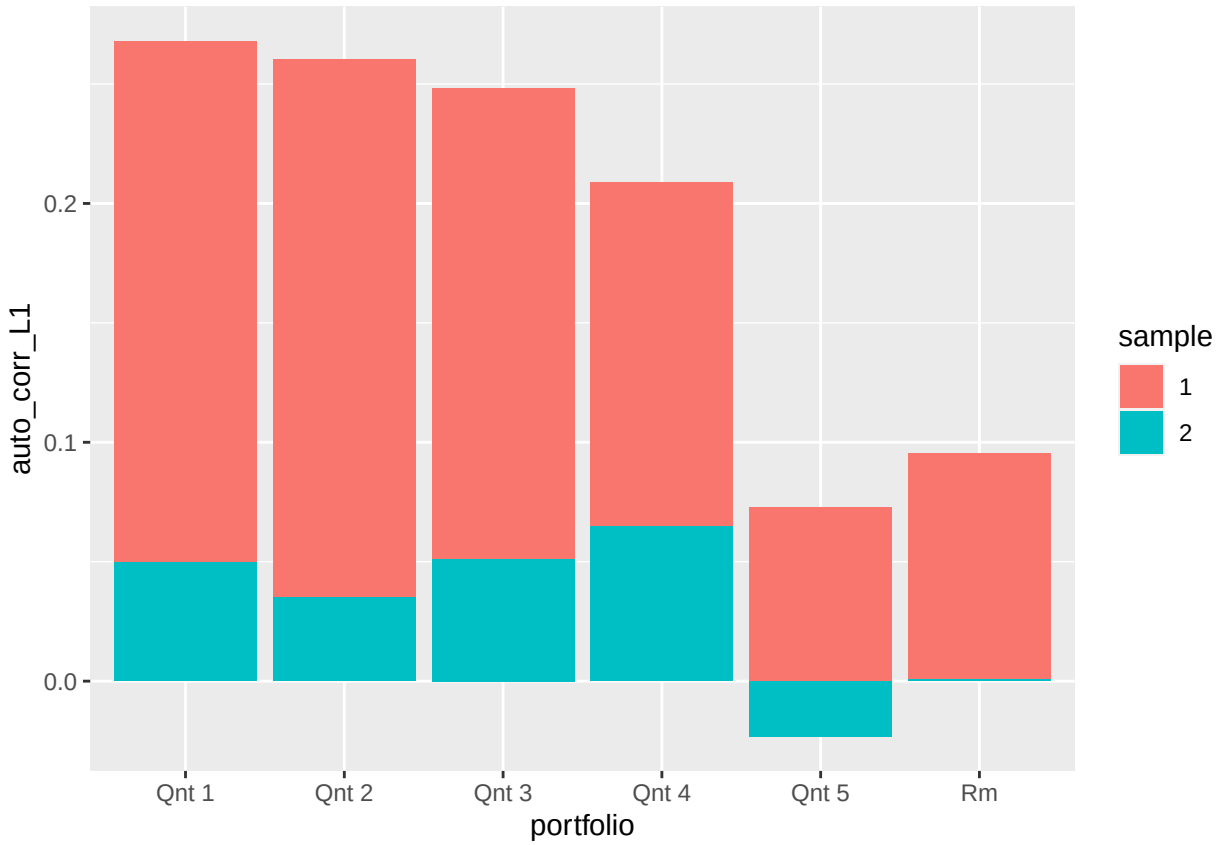
```
freq <- c(rep(c("daily"),6),rep(c("monthly"),6))
portfolio <- rep(c("Rm", "Qnt 1","Qnt 2","Qnt 3","Qnt 4","Qnt 5"),2)
auto_corr_L1 <- c(stat.Rm_d$auto_corr[2],stat.size_d.Lo20$auto_corr[2],stat.size_d.Qnt2$auto_corr[2],sta
res <- data.frame(freq=freq,portfolio=portfolio,auto_corr_L1=auto_corr_L1)
ggplot(res, aes(x=portfolio, y=auto_corr_L1, fill=freq))+geom_col()
```



Sampling Frequency

The above plot shows the 1^{st} order correlation of size portfolios and market returns of daily and monthly frequency. The daily returns are more predictable than the monthly ones, having higher first-order correlation coefficients.

```
sample <- freq <- c(rep(c("1"),6),rep(c("2"),6))
auto_corr_L1.2 <- c(stat.Rm_d1$auto_corr[2],stat.size_d1.Lo20$auto_corr[2],stat.size_d1.Qnt2$auto_corr[
res2 <- data.frame(sample=sample,portfolio=portfolio,auto_corr_L1=auto_corr_L1.2)
ggplot(res2, aes(x=portfolio, y=auto_corr_L1, fill=sample))+geom_col()
```



Subsample
The above plot shows the 1st order correlation of size portfolios and market returns at the daily frequency of the two sub-samples. In the second sample, autocorrelation significantly drops across all portfolios.

Moreover, the effect of size becomes ambiguous in the second sample.

The effect of size As shown in previous plots, the effect of size plays a significant role in return predictability. Predictability decreases as market caps go up. This is true for different sampling frequencies. However, in the second half of the sample, the size effect becomes much less pronounced than the previous period.