```
## Warning: package 'rmdformats' was built under R version 3.3.2
```

Volatility Factor In China Euqity Market

A simple portfolio

Data

All stocks(3000+) listed in China A share mkt on 2016-12-30. Each has OHLC and `Volume` and `Amount` (Cash Volume).

Universe

We thinning the world.

Criteria:

1. Select stocks that have full history (traded from 2014-01-01 to 2016-12-30)

- Then there are (2400+) stocks remaining. We build a mkt equal weight index `INDX_EQW` at this level, and use it as hedge if any.

2. Cut the whole database into *train* and *test*.

- *train*: 2014-01-01 – 2015-12-31
- *test* : 2016-01-01 – 2016-12-30

3. We play over the liquid world. Based on the *train* dataset, screen out stocks that have median of daily `AMOUNT` less than half billion CNY.

Now there are 104 stocks remaining. Welcome to the liquid playgroud!

Assumption

Volatility has positive return over US/EURO equity market. One explanation is that institutions prefer low vol stock due to tight risk budget. So a *long low vol/short high vol* portfolio has unexplained (by mkt) positive return.

Does it apply to China Equity Mkt?

Portfolio

Due to the short ban, it is hard to short single stocks. The portfolio implements the long and hedge out mkt beta by `INDX_EQW`

Portfolio P1:

- long: stocks in the first quartile of return std deviation (computation bases on *train* dataset)
- short: stocks in the last quartile of return std deviation
- hedge: neutralize the mkt beta by `INDX_EQW` (beta estimation is based on *train* dataset)

The *long* names and weights (equal weight):

```r
w <- c(a, -b)
print(w)
```

```
  2VDCFVAX3   3Z31CWVT5   53RPAYFP9   7H7P6RZ50   7TUNF66J1   8HVTD1952 
 0.04000000  0.04000000  0.04000000  0.04000000  0.04000000  0.04000000 
  9WGVKL7H4   9X4NJVF92   AR8494CW4   G4UKSDGK9   HUNZXC9B9   K929LXWK4 
 0.04000000  0.04000000  0.04000000  0.04000000  0.04000000  0.04000000 
  L3TJAH2Y5   LCQXJZGF4   MHDLH21A4   MPU87MWZ4   RQWBZMP85   SW9KXL211 
 0.04000000  0.04000000  0.04000000  0.04000000  0.04000000  0.04000000 
  T66Z63FH8   TZY7VMLW5   VLAS166R9   WR26YUBR7   X9W1FV4P5   XKTFPCJ52 
 0.04000000  0.04000000  0.04000000  0.04000000  0.04000000  0.04000000 
  YC8QN17L2   463XM2F38   4Q1JUN3G6   4V4YAGH89   6W2VYUR85   89NY5DPL8 
 0.04000000 -0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 
  964RUUNW1   9QYLAVFU0   ARH5QY892   FK7DJWZ98   FYB9JPPW7   GT3J87VA7 
-0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 
  GT7JYN9Z9   LU9NRN284   MFAF84VG4   NTHW5PGV9   PLKKVAUC6   Q7HKL8S54 
-0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 
  QFYLK57K2   RJZGHWKN7   RVMFRAS13   SR5VXVMA8   SSGUPS396   V5XUKKMR4 
-0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 -0.03846154 
  VQSP1LDP8   XT5ANDJD3   ZGXCL2SG3 
-0.03846154 -0.03846154 -0.03846154 
```

Beta of *long* and *short*:

```r
portf_a <- portfConst(UniverseNames = names(U_train_liq), a)
portfBeta_a <- betaExposure(portf_a, UTrainLiqBeta)

print(portfBeta_a)
```

```
[1] 0.4393381
```

```r
portf_b <- portfConst(UniverseNames = names(U_train_liq), b)
portfBeta_b <- betaExposure(portf_b, UTrainLiqBeta)

print(portfBeta_b)
```

```
[1] 1.229711
```

One can neutralize the beta by `INDX_EQW`

Performance

For stake of simplicity, we hold staic portfilio.

The in-sample performance.

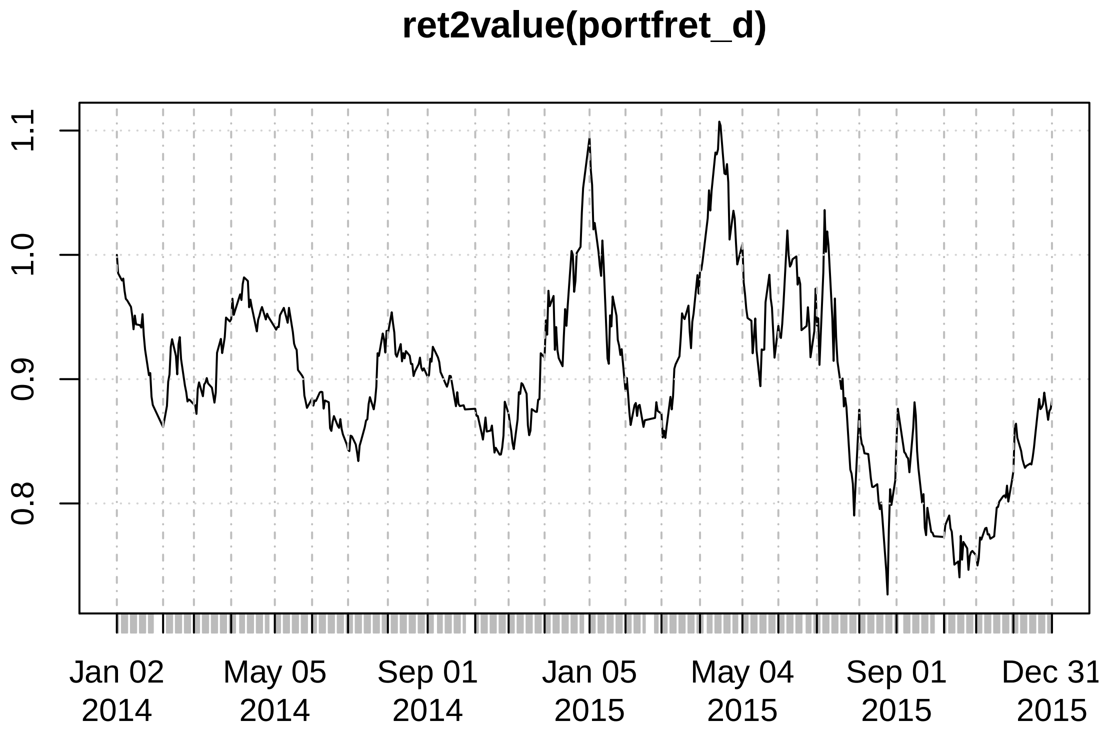
```r
portf_d <- portfConst(UniverseNames = names(U_train_liq), w)
portfBeta_d <- betaExposure(portf_d, UTrainLiqBeta)

portfret_d <- portfRet(Universe = U_train_liq, portf = portf_d, betahedge = T,
    INDX = INDX_EQW_train, UniverseBeta = UTrainLiqBeta)
portfValue_d <- ret2value(portfret_d)
summary(portfret_d)
```

```
     Index              portfret_d        
 Min.   :2014-01-02   Min.   :-0.0760531  
 1st Qu.:2014-07-04   1st Qu.:-0.0099715  
 Median :2014-12-31   Median :-0.0014764  
 Mean   :2015-01-02   Mean   :-0.0000867  
 3rd Qu.:2015-07-03   3rd Qu.: 0.0085668  
```

```
 Max.   :2015-12-31   Max.   : 0.0813963
                      NA's   :1
```

```r
plot(ret2value(portfret_d))
```




```r
plot(INDX_EQW_train$VALUE)
```

P1 fails to match the index, especially in the period 2014-11 – 2015- 07. Timing is necessary.

Volatility and Risk Appetite

As metioned above, volatility facotr return comes from risk aversion. But the mkt is not always risk averse. Timing should be applied.

Intuition

When the mkt is a safe heaven, investors loosens risk budget and tends to play risk. Then low vol premium (low vol/high vol portfolio return) is negative.

When the mkt is tight, safety, ie low volatility, has highest priority.

Volatility: Timing is the Key

Vol Timing Factor: Amount/Volume and Volatility

Intuition indicates 2 factors: Volume and Volatility

Note:

1. The dataset doesnot have a mktwise volume entry. I use the `INDX_EQW` hypothetical volume (AMOUNT/VALUE) instead.
2. Chinese mkt does not have an indicator like `VIX`. The proxy I use is `INDX_EQW` 50d rolling std div.

```
lm.1 <- lm(portfret_d ~ INDX_EQW_train$RET.CC.1 + log(INDX_EQW_train$VOLUME) +
    INDX_EQW_train$sd50)

summary(lm.1)
```

```
Call:
lm(formula = portfret_d ~ INDX_EQW_train$RET.CC.1 + log(INDX_EQW_train$VOLUME) +
    INDX_EQW_train$sd50)

Residuals:
      Min        1Q    Median        3Q       Max
-0.077982 -0.010139 -0.000719  0.009375  0.080174

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                -0.057029   0.038840  -1.468    0.143
INDX_EQW_train$RET.CC.1    -0.006398   0.039967  -0.160    0.873
log(INDX_EQW_train$VOLUME)  0.003042   0.002012   1.512    0.131
INDX_EQW_train$sd50        -0.113214   0.083726  -1.352    0.177

Residual standard error: 0.01895 on 435 degrees of freedom
  (50 observations deleted due to missingness)
Multiple R-squared:  0.0072,    Adjusted R-squared:  0.0003534
F-statistic: 1.052 on 3 and 435 DF,  p-value: 0.3695
```

Seems like the hedge is effective. `RET.CC.1` is not relevant to the low vol premium.

IDEA: Volume and Volatility should have double effect– explosion can happen both when mkt is overheaded or in panic– so direction should be introduced.

```r
lm.2 <- lm(portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$sd50)

summary(lm.2)
```

```
## 
## Call:
## lm(formula = portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$sd50)
## 
## Residuals:
##       Min        1Q    Median        3Q       Max 
## -0.076458 -0.009920 -0.001276  0.008923  0.082337 
## 
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.591e-03  1.818e-03   0.875    0.382
## INDX_EQW_train$signedlogVolume  1.716e-06  4.771e-05   0.036    0.971
## INDX_EQW_train$sd50            -7.297e-02  7.973e-02  -0.915    0.361
## 
## Residual standard error: 0.01898 on 436 degrees of freedom
##   (50 observations deleted due to missingness)
## Multiple R-squared:  0.00194,   Adjusted R-squared:  -0.002638 
## F-statistic: 0.4238 on 2 and 436 DF,  p-value: 0.6548
```

Here is the magic

```r
lm.3 <- lm(portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$signedsd50 +
    INDX_EQW_train$signedsd50logVolume)

summary(lm.3)
```

```
Call:
lm(formula = portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$signedsd50 +
    INDX_EQW_train$signedsd50logVolume)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.074307 -0.010307 -0.001459  0.009906  0.079463 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                        -0.0000943  0.0009223  -0.102   0.9186
INDX_EQW_train$signedlogVolume      0.0002948  0.0000957   3.080   0.0022
INDX_EQW_train$signedsd50          -1.7898141  2.2985912  -0.779   0.4366
INDX_EQW_train$signedsd50logVolume  0.0765160  0.1159184   0.660   0.5095
                                     
(Intercept)                          
INDX_EQW_train$signedlogVolume     **
INDX_EQW_train$signedsd50            
INDX_EQW_train$signedsd50logVolume   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01876 on 435 degrees of freedom
  (50 observations deleted due to missingness)
```

```
Multiple R-squared:  0.02726,   Adjusted R-squared:  0.02056
F-statistic: 4.064 on 3 and 435 DF,  p-value: 0.007253
```

Seems like `signedVolume` dominates.

Here is the majestic:

```r
lm.4 <- lm(portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$ma30signedlogVolume +
    INDX_EQW_train$signedsd50)

summary(lm.4)
```

```

Call:
lm(formula = portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$ma30signedlogVolume +
    INDX_EQW_train$signedsd50)

Residuals:
      Min        1Q    Median        3Q       Max
-0.073832 -0.010361 -0.001561  0.009717  0.078291

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                        -9.188e-04  1.288e-03  -0.713  0.47604
INDX_EQW_train$signedlogVolume      2.624e-04  9.495e-05   2.763  0.00596
INDX_EQW_train$ma30signedlogVolume  1.986e-04  2.115e-04   0.939  0.34819
INDX_EQW_train$signedsd50          -2.656e-01  8.018e-02  -3.313  0.00100

(Intercept)
INDX_EQW_train$signedlogVolume     **
INDX_EQW_train$ma30signedlogVolume
INDX_EQW_train$signedsd50          **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01875 on 435 degrees of freedom
  (50 observations deleted due to missingness)
Multiple R-squared:  0.02826,   Adjusted R-squared:  0.02156
F-statistic: 4.217 on 3 and 435 DF,  p-value: 0.005895
```

```r
cor(cbind(INDX_EQW_train$signedlogVolume, INDX_EQW_train$ma30signedlogVolume,
    INDX_EQW_train$signedsd50), use = "complete.obs")
```

```
                    signedlogVolume ma30signedlogVolume signedsd50
signedlogVolume           1.0000000           0.2169901  0.8629652
ma30signedlogVolume       0.2169901           1.0000000  0.1360366
signedsd50                0.8629652           0.1360366  1.0000000
```

```r
lm.4 <- lm(portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$ma30signedlogVolume +
    INDX_EQW_train$sd50)

summary(lm.4)
```

```

Call:
lm(formula = portfret_d ~ INDX_EQW_train$signedlogVolume + INDX_EQW_train$ma30signedlogVolume +
    INDX_EQW_train$sd50)
```

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.075241 -0.010252 -0.001354  0.009264  0.082413 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                         5.802e-05  2.339e-03   0.025    0.980
INDX_EQW_train$signedlogVolume     -8.741e-06  4.875e-05  -0.179    0.858
INDX_EQW_train$ma30signedlogVolume  2.333e-04  2.240e-04   1.042    0.298
INDX_EQW_train$sd50                -4.581e-02  8.387e-02  -0.546    0.585

Residual standard error: 0.01898 on 435 degrees of freedom
  (50 observations deleted due to missingness)
Multiple R-squared:  0.004424,  Adjusted R-squared:  -0.002442 
F-statistic: 0.6443 on 3 and 435 DF,  p-value: 0.5869
```

```r
cor(cbind(INDX_EQW_train$signedlogVolume, INDX_EQW_train$ma30signedlogVolume,
    INDX_EQW_train$sd50), use = "complete.obs")
```

```
                    signedlogVolume ma30signedlogVolume        sd50
signedlogVolume          1.00000000           0.2169901 -0.07000239
ma30signedlogVolume      0.21699010           1.0000000 -0.31780776
sd50                    -0.07000239          -0.3178078  1.00000000
```

Consider the lag version:

```r
lm.5 <- lm(portfret_d ~ INDX_EQW_train$lag5signedlogVolume + INDX_EQW_train$lag5ma30signedlogVolume)

summary(lm.5)
```

```

Call:
lm(formula = portfret_d ~ INDX_EQW_train$lag5signedlogVolume + 
    INDX_EQW_train$lag5ma30signedlogVolume)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.076882 -0.009900 -0.001525  0.008913  0.078694 

Coefficients:
                                         Estimate Std. Error t value
(Intercept)                            -7.151e-04  1.270e-03  -0.563
INDX_EQW_train$lag5signedlogVolume     -7.861e-05  4.714e-05  -1.667
INDX_EQW_train$lag5ma30signedlogVolume  2.738e-04  2.096e-04   1.306
                                       Pr(>|t|)  
(Intercept)                              0.5736  
INDX_EQW_train$lag5signedlogVolume       0.0961 .
INDX_EQW_train$lag5ma30signedlogVolume   0.1921  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01867 on 451 degrees of freedom
  (35 observations deleted due to missingness)
Multiple R-squared:  0.008168,  Adjusted R-squared:  0.00377 
F-statistic: 1.857 on 2 and 451 DF,  p-value: 0.1573
```

```r
cor(INDX_EQW_train$lag5signedlogVolume, INDX_EQW_train$lag5ma30signedlogVolume,
    use = "complete.obs")
```

```
                    lag5ma30signedlogVolume
lag5signedlogVolume               0.2178518
```

Conclusion:

Low Vol premium is highly related to `signedlogVolume` and `signedsd50`, even with respect to the lag of moving average smoothed version.

More to Go

Other Potential Factors:

Limit-up Ceiling

Limit-down Floor

Intraday Floor-Ceiling dynamics

Implied Vol forecasting (China VIX)

……

The Full Model

Just as many other factors, the factor return of low vol depends over the market regime. One systametic approach to dynamic factor rotation strategies is a *Market Regime Switch Model*.

Market Regime Switch: Probablity Graphic Approach

Markov Graph: Different Market status, Status may transfer. The transition is described by a transition matrix. One optimal factor portfolio should be held if one does not have a forecasting power to the forward mkt status. The optimal factor portfolio can be a start point of a multi factor rotation strategy.

Some adhoc ways

ML approach: SVM classification. RF ??

Way to Go

The problems:

1. Beta estimation

Dynamic beta hedge is not employed. The portfolio does have beta exposure though not significant.

2. Vol estimtion

Intuitively, Volatility should have extra info to the vol premium. An accurate estimation and forecasting of mkt realized vol may help (For how, check http://rpubs.com/ericwbzhang/217044)

Some info from the implied vol may boost portfolio performance.

3. More signal introduced to forecast vol premium.

eg. Celling and Floor.

4. The Full Modell: Market Regime Switch

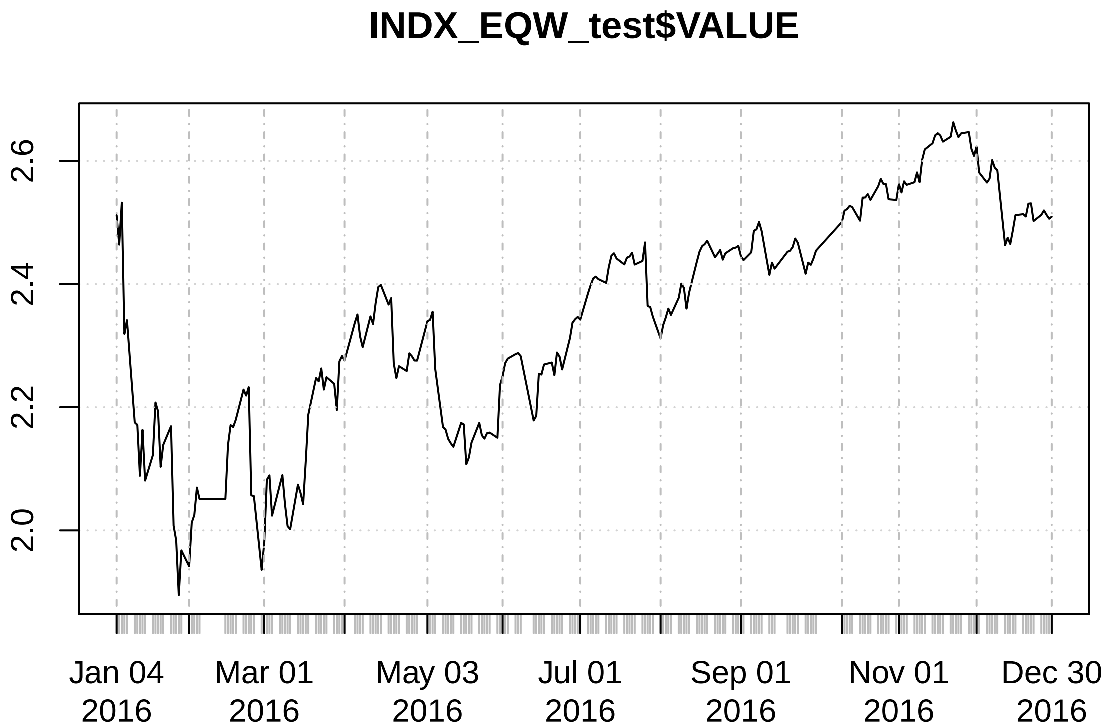
5. The Value of PM

Factor is employed by many professional investors since it is understandable, which means forecastable for seasoned practioners. PMs with alpha should have a forecasting power over the forward mkt status. The role a quant may play is to reveal what happens in a clear way.

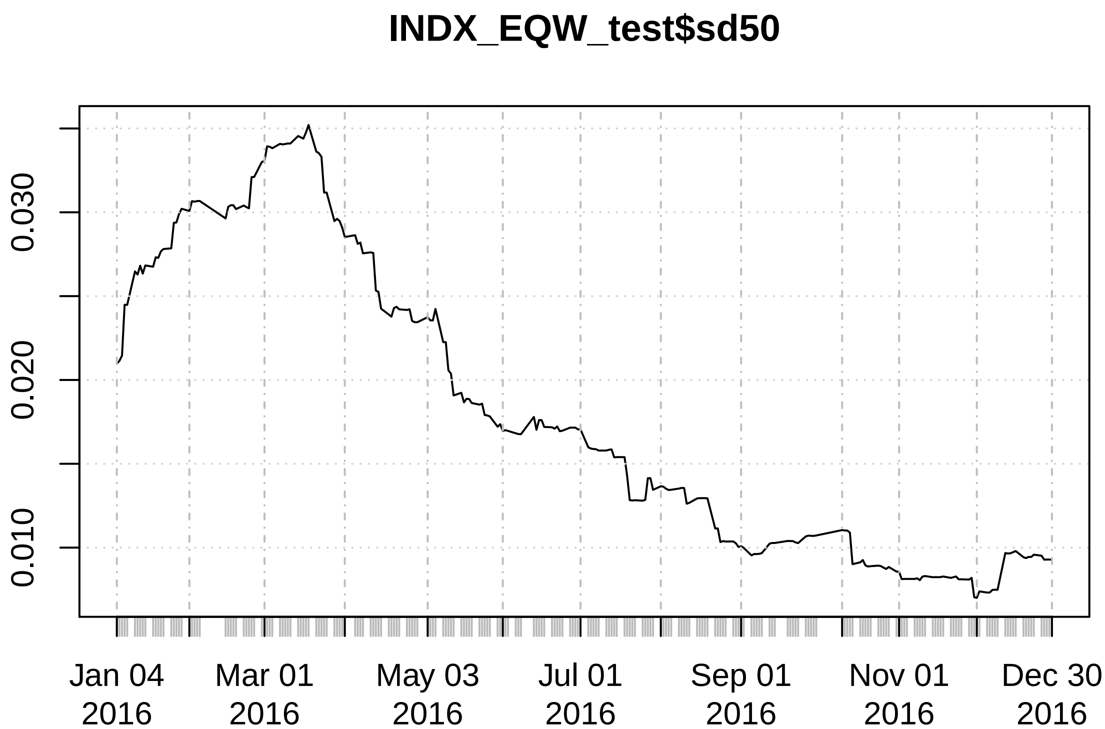
Show-off

I dont have much time to do a bar-by-bar out of sample backtesting. (Note that what I have done is purely over 2014-2015 dataset, the 2016 test set is not touched.) While a quick guess may be good enough.

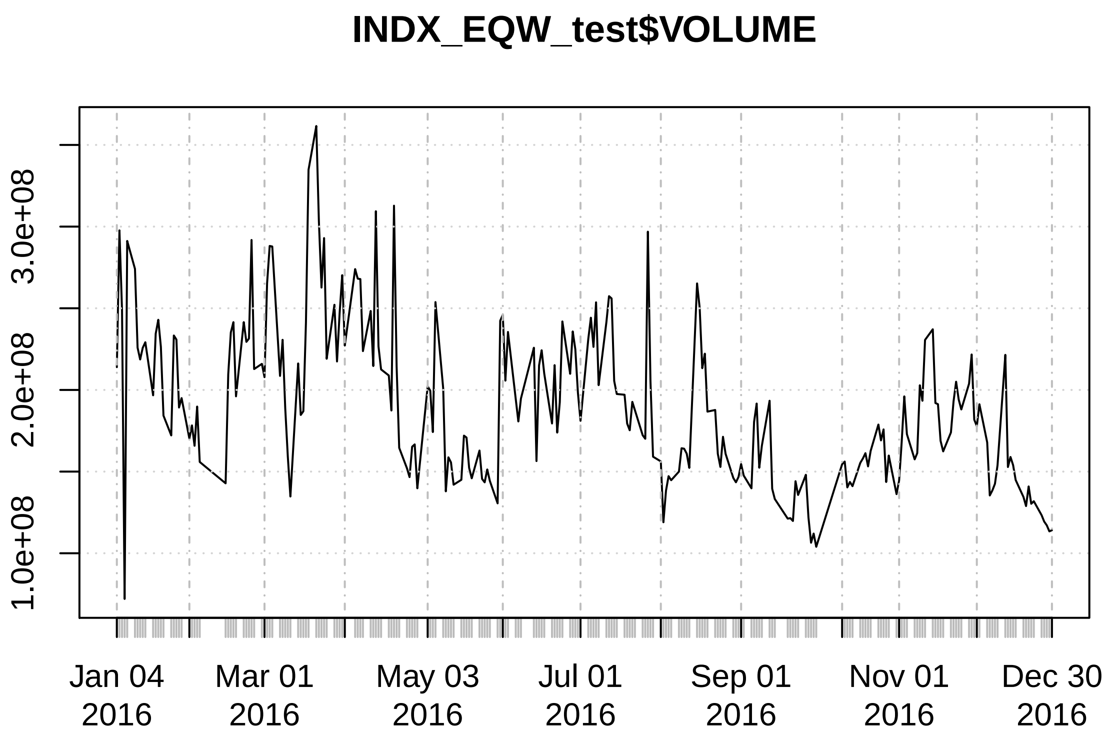
```r
plot(INDX_EQW_test$VALUE)
```



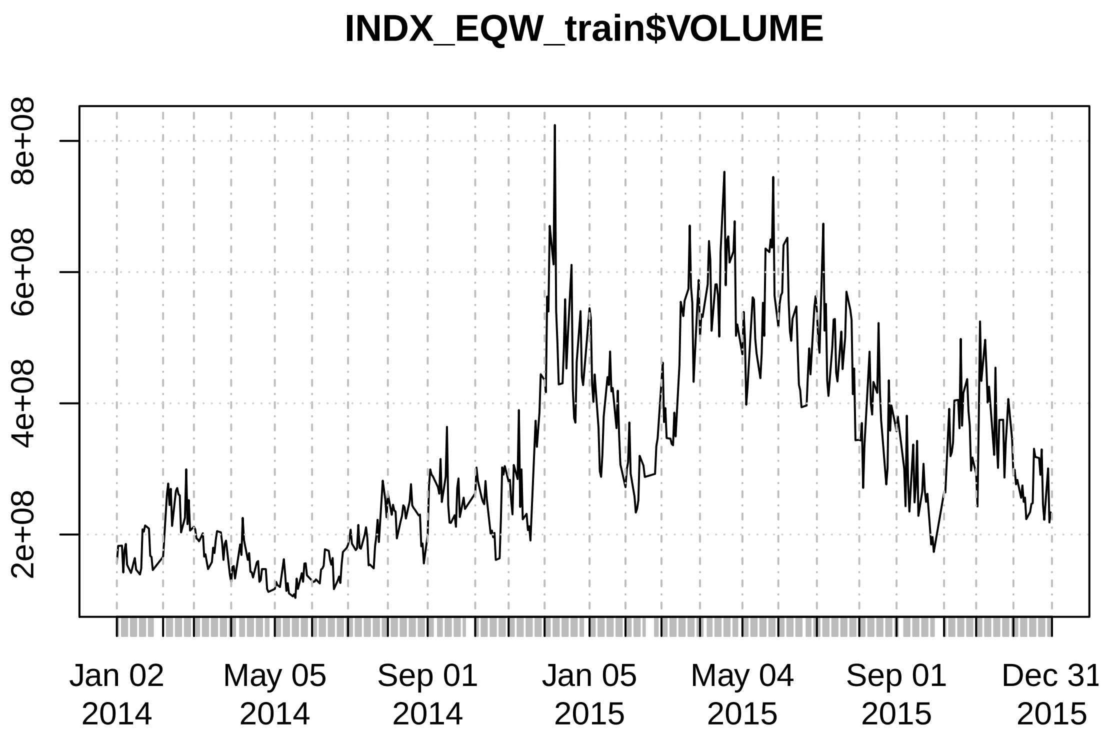

```r
plot(INDX_EQW_test$sd50)
```




```r
plot(INDX_EQW_test$VOLUME)
```




```r
plot(INDX_EQW_train$VOLUME)
```



Recall `lm.4`: Vol premium is positive when mkt is weak and mild, ie. the bar is short and volmue is gradually expanding– this is what happens during 2016.

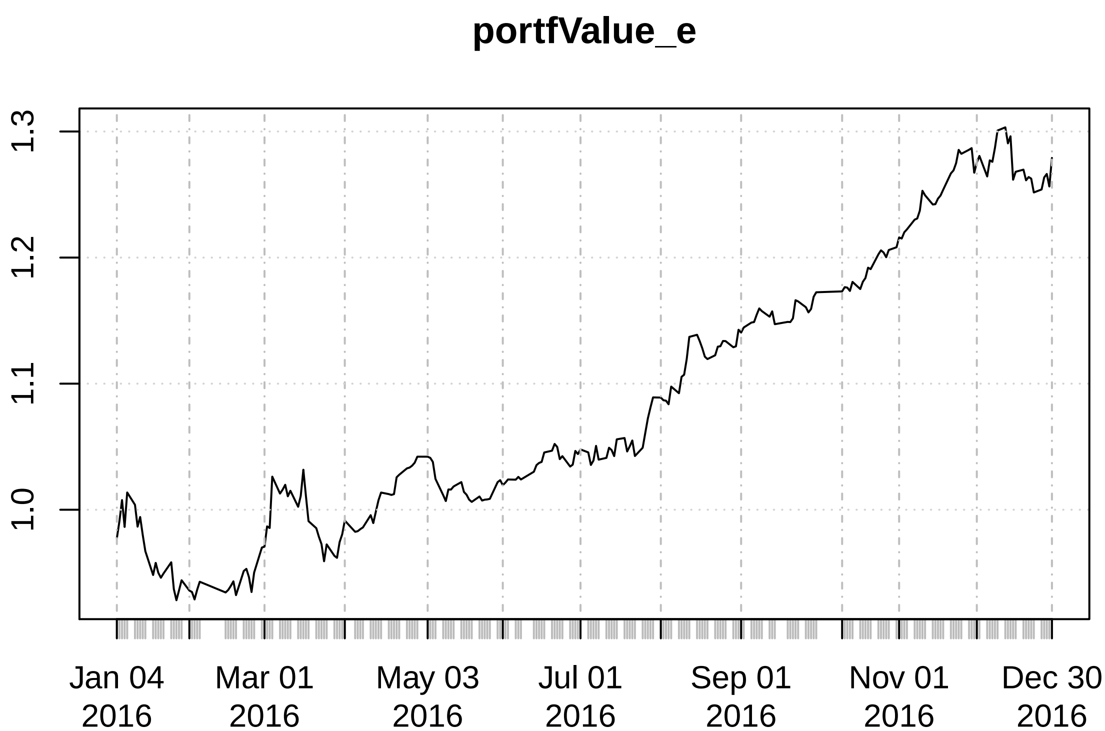
One could make a guess that the vol premium during 2016 should be decent (different from the trivial performance in 2014-2015), and the beginning may suffer a mild drawdown.

See what actually happens:

```
portf_e <- portfConst(UniverseNames = names(U_train_liq), c(a, -b))
portfret_e <- portfRet(Universe = U_test_liq, portf = portf_e, betahedge = T,
    INDX = INDX_EQW_test, UniverseBeta = UTrainLiqBeta)

portfValue_e <- ret2value(portfret_e)

plot(portfValue_e)
```




```r
summary(portfret_e)
```

```
     Index               portfret_e
 Min.   :2016-01-04   Min.   :-0.026592
 1st Qu.:2016-04-05   1st Qu.:-0.003146
 Median :2016-07-04   Median : 0.001517
 Mean   :2016-07-03   Mean   : 0.001047
 3rd Qu.:2016-09-29   3rd Qu.: 0.005366
 Max.   :2016-12-30   Max.   : 0.041375
```

```r
# Sharpe Ratio
mean(portfret_e, na.rm = T)/sd(portfret_e, na.rm = T) * 16
```

```
[1] 1.92887
```