Fid_Nov_18

Weiling Li

11/18/2019

```
## reads in functions that calculate daily log returns, log closed value, etc.
source(file = "./functions/fid_data_wrangler.R")
## reads in functions calculate goodness of fit
source(file = "./functions/fid_goodness_of_fit.R")
```

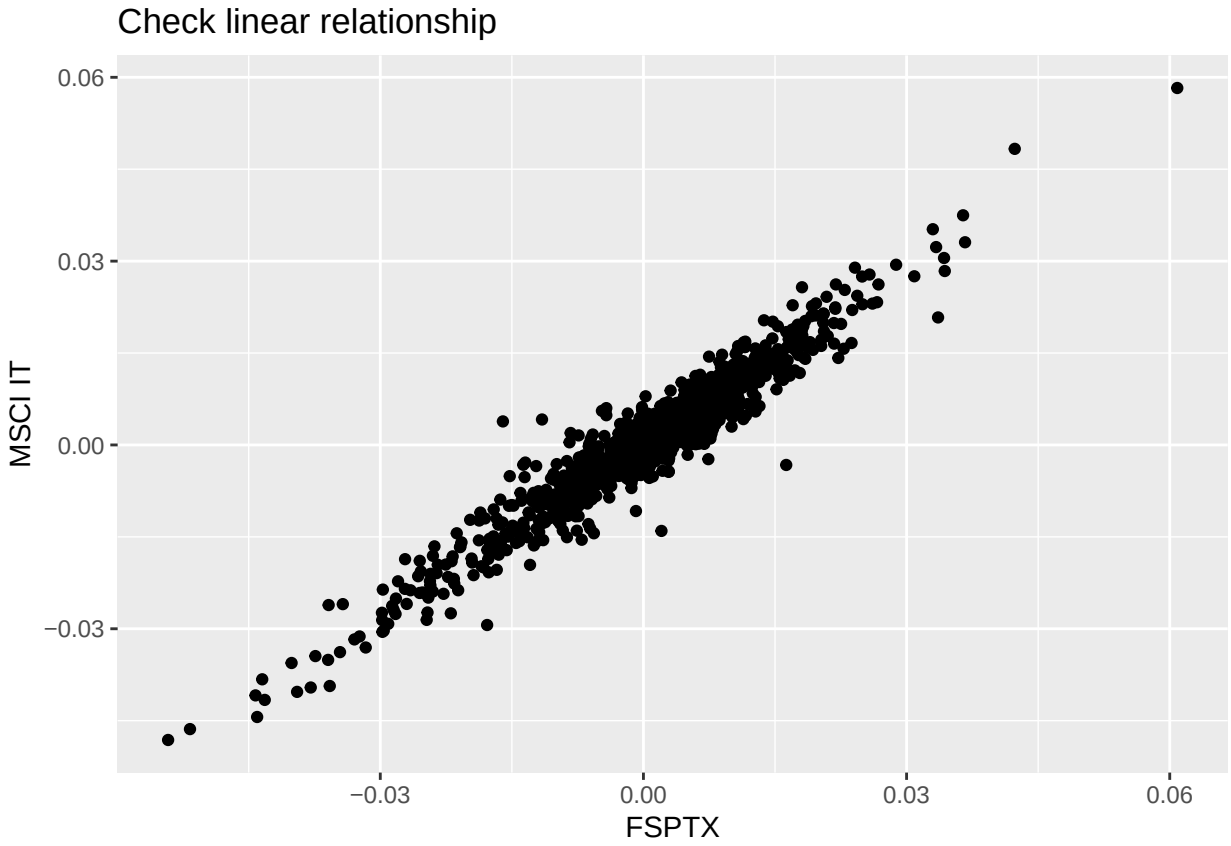
```
FSPTX = dailynlogReturn(StartDate,read_csv("./Index_Data/FSPTX.csv"))       ## Target Fund
NASDAQ = dailynlogReturn(StartDate,read_csv("./Index_Data/^IXIC.csv"))      ## NASDAQ
SnP500 = dailynlogReturn(StartDate,read_csv("./Index_Data/^GSPC.csv"))      ## S&P500
SnPMID = dailynlogReturn(StartDate,read_csv("./Index_Data/^MID.csv"))       ##
SnPSML = dailynlogReturn(StartDate,read_csv("./Index_Data/^SML.csv"))
RUSSELL2000 = dailynlogReturn(StartDate,read_csv("./Index_Data/^RUT.csv"))
VGT = dailynlogReturn(StartDate,read_csv("./Index_Data/VGT.csv"))
VIGAX = dailynlogReturn(StartDate,read_csv("./Index_Data/VIGAX.csv"))
VTSAX = dailynlogReturn(StartDate,read_csv("./Index_Data/VTSAX.csv"))
SnP500Info <- dailynlogReturn(StartDate,read_csv("./Index_Data/SnP500Info.csv"))
SnPNATech <- dailynlogReturn(StartDate,read_csv("./Index_Data/SnPNATECH_clean.csv"))
MSCI_INFO_TECH <- dailynlogReturn(StartDate,read_xls(path = "./Index_Data/MSCI_Us_IMI_information_Tech_1
Benchmark_industry_XOI <- dailynlogReturn(StartDate,read_csv("Index_Data/Benchmark_industry_XOI.csv"))
VIX <- dailynlogReturn(StartDate,read_csv("./Index_Data/^VIX.csv"))
```

fit benchmark

The logic goes:

- fit a `lm` model with `FSPTX` as the outcome and benchmark(`MSCI US IMI Info Tech 25/50`) as predictor(drop intercept)
- Check if fund and benchmark has linear relationship
- Check residuals
  - Does the residual plot show any trend?
  - Is residual follows a normal distribution?
  - Standardized residual Show follows a distribution of $\mathcal{N}(0,1)$. Does the residual plot support that?
- Check if we can regenerate the original data
  - Calculate y_{sim} and y_{obs}.
  - KS test: Check if we can reject the hypothesis they are from the same distribution.
  - Permutation test: Check if we can reject the hypothesis that they are independent.
  - T test: If they are dependent, check the mean of them using t test.(assume normality)
  - F test: If they are dependent, check the if the variance are different using f test.(assume normality)
  - Visually compare results using histogram.

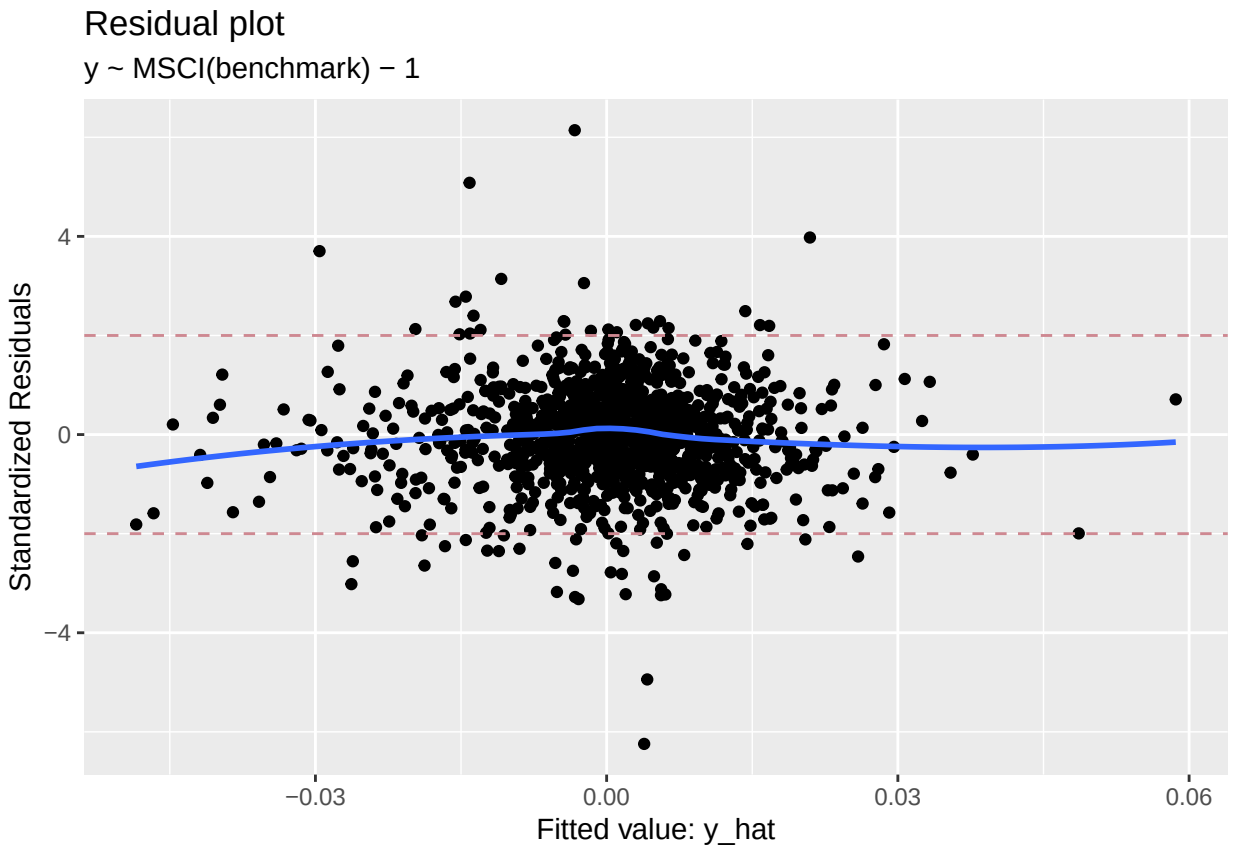
```
ggplot()+aes(x = FSPTX$log.Return, y = MSCI_INFO_TECH$log.Return)+
  geom_point()+xlab("FSPTX")+ylab("MSCI IT")+ggtitle(label = "Check linear relationship")
```



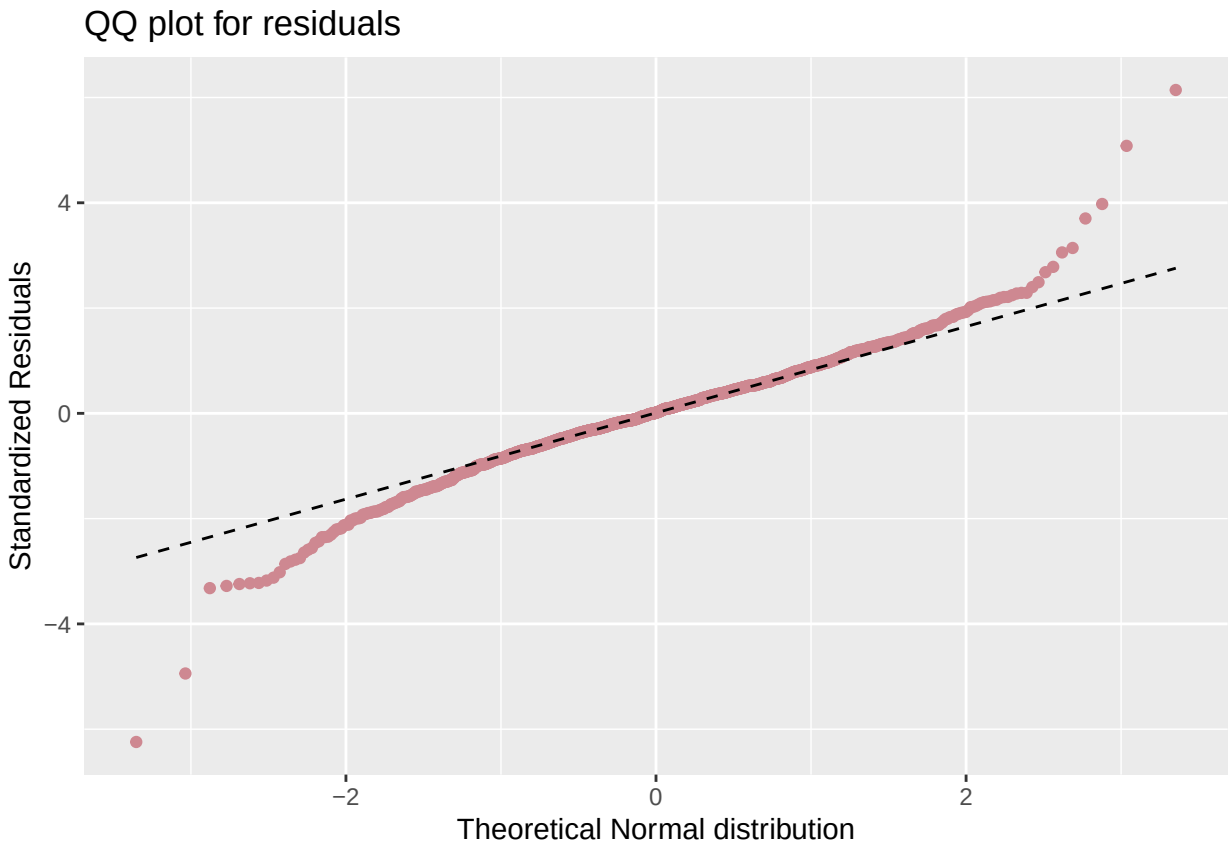

```
fit.benchmark <- lm(FSPTX$log.Return~MSCI_INFO_TECH$log.Return-1)
summary(fit.benchmark)
```

```
##
## Call:
## lm(formula = FSPTX$log.Return ~ MSCI_INFO_TECH$log.Return - 1)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.019885 -0.001731  0.000043  0.001789  0.019566
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## MSCI_INFO_TECH$log.Return 1.006367   0.008347   120.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003185 on 1248 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.9209, Adjusted R-squared:  0.9209
## F-statistic: 1.454e+04 on 1 and 1248 DF,  p-value: < 2.2e-16
```

```
p.benchmark <- plotmodel(fit.benchmark,"y ~ MSCI(benchmark) - 1","#CE8891")
## check residuals
p.benchmark$redisualplot
```




```
p.benchmark$qqplot
```



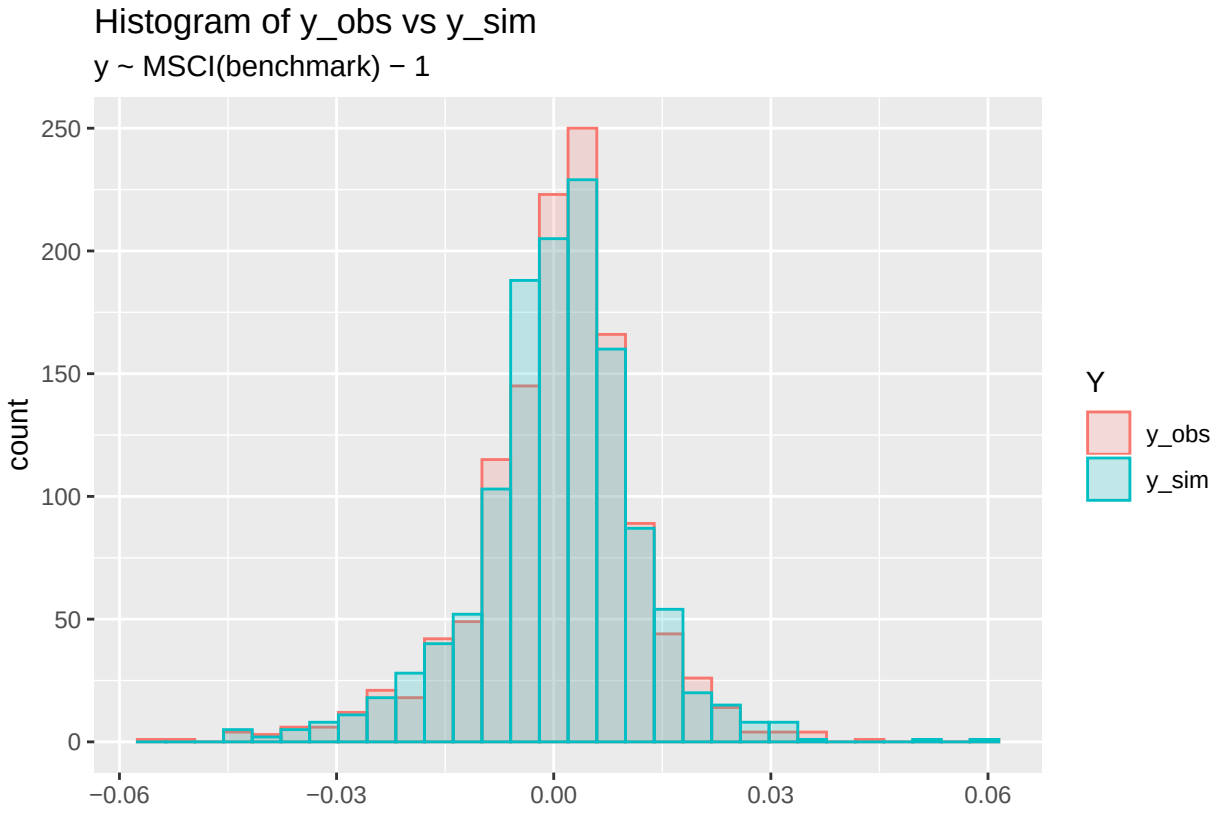

```
## check standardized residuals the 95% interval
quantile(scale(fit.benchmark$residuals),probs = c(.25,.95))
```

```
##        25%        95%
## -0.5436968  1.4779482
```

```
## simulate fake data and compare results
md_test(fit.benchmark,method = "pearson",modelequation ="y ~ MSCI(benchmark) - 1" )
```

```
## $KS_test
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  y_obs and y_sim
## D = 0.03763, p-value = 0.3395
## alternative hypothesis: two-sided
##
##
## $Permutation_test
##
##  Asymptotic General Independence Test
##
## data:  y_obs by y_sim
## Z = 32.365, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
##
## $t_test
```

```
##
##  Welch Two Sample t-test
##
## data:  y_obs and y_sim
## t = 0.080171, df = 2496, p-value = 0.9361
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.0008510090  0.0009235614
## sample estimates:
##    mean of x    mean of y
## 0.0005033058 0.0004670296
##
##
## $correlation_test
## [1] 0.9161674
##
## $variance_F_test
##
##  F test to compare two variances
##
## data:  y_sim and y_obs
## F = 0.99672, num df = 1248, denom df = 1248, p-value = 0.9537
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.8919933 1.1137436
## sample estimates:
## ratio of variances
##          0.9967206
##
##
## $histogram
```


Histogram of y_obs vs y_sim
y ~ MSCI(benchmark) − 1
count
250
200
150
100
50
0
−0.06
−0.03
0.00
0.03
0.06
Y
y_obs
y_sim