Are High Frequency Traders Prudent and Temperate?

R/Finance Chicago 2015

Jerzy Pawlowski *jp3900@nyu.edu*

NYU Polytechnic School of Engineering

May 28, 2015

Prudence and *Temperance*

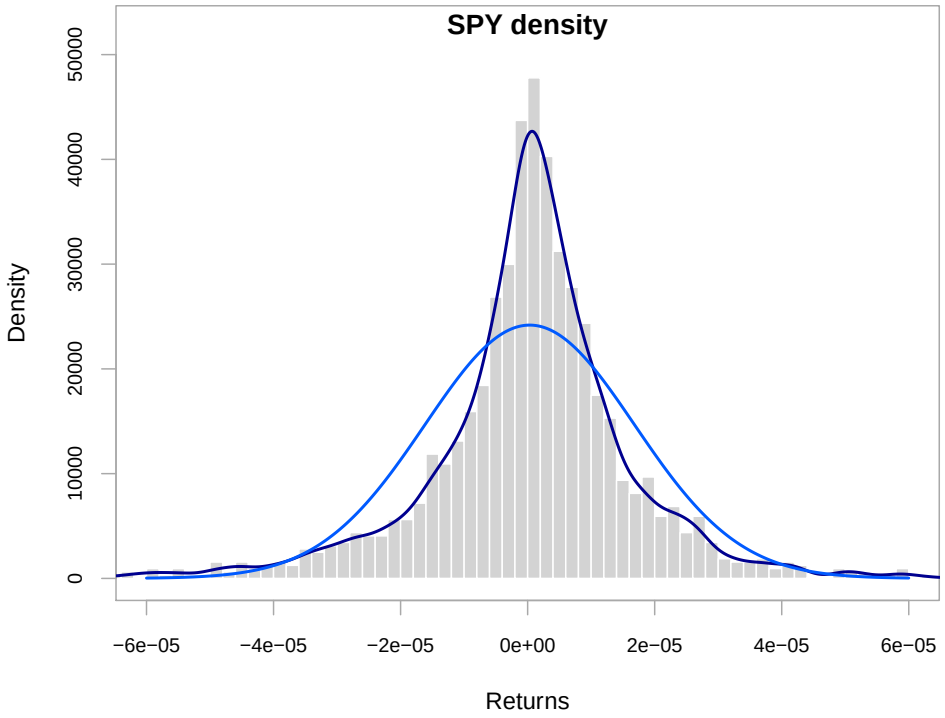
Risk averse investors prefer positive *means* (first moment) and small *variance* (second moment),

Investors may also have preferences for higher *moments* - they may prefer large positive odd moments and small even moments,

Prudence refers to the preference for large positive *skewness* (third moment),

Temperance refers to the preference for small *kurtosis* (fourth moment),

Amaya et al., Does Realized Skewness Predict the Cross-Section of Equity Returns? papers.ssrn.com/sol3/papers.cfm?abstract_id=1898735




```
> sym_bol <- load(file.path(output_dir,
+       paste0(sym_bol, ".RData")))
> re_turns <- calc_rets(xts_data=to.daily(get(sym_bol)))
> len_rets <- nrow(re_turns)  # number of obser
> mean_rets <- mean(re_turns[, 1])  # calculate
> sd_rets <- sd(re_turns[, 1])  # calculate sta
> # calculate skew and kurtosis
> (sum(((re_turns[, 1] - mean_rets)/sd_rets)^3)
[1] -0.3773046
> (sum(((re_turns[, 1] - mean_rets)/sd_rets)^4)
[1] 7.448475
```

```
> library(PerformanceAnalytics)
> chart.Histogram(re_turns[, 1], main="",
+   xlim=c(-6e-5, 6e-5),
+   methods = c("add.density", "add.normal"))
> # add title
> title(main=paste(sym_bol,
+    "density"), line=-1)
```

Package HighFreq for Managing High Frequency Data

Package HighFreq contains functions for managing high frequency *TAQ* and *OHLC* market data:

- reading and writing data from files,
- managing time zones and alligning indices,
- chaining and joining time series,
- scrubbing bad data points,
- converting *TAQ* data to *OHLC* format,
- aggregating data to lower frequency,

HighFreq is inspired by the package highfrequency, and follows many of its conventions,

HighFreq depends on packages xts, quantmod, lubridate, and caTools,

The function scrub_agg() scrubs a single day of *TAQ* data, aggregates it, and converts it to *OHLC* format,

The function save_scrub_agg() loads, scrubs, aggregates, and binds multiple days of *TAQ* data for a single symbol, and saves the *OHLC*

```
> # install package "HighFreq" from github
> install.packages("devtools")
> library(devtools)
> install_github(repo="algoquant/HighFreq")
> # load package "HighFreq"
> library(HighFreq)
> # set data directories
> data_dir <- "C:/Develop/data/hfreq/src/"
> output_dir <- "C:/Develop/data/hfreq/scrub/"
> # define sym_bol
> sym_bol <- "SPY"
> # load a single day of TAQ data
> sym_bol <- load(
+   file.path(data_dir,
+       paste0(sym_bol, "/2014.05.02.",
+              sym_bol, ".RData")))
> # scrub, aggregate single day of TAQ data to O
> ohlc_data <- scrub_agg(taq_data=get(sym_bol))
> # aggregate TAQ data for symbol, save to file
> save_scrub_agg(sym_bol,
+          data_dir=data_dir,
+          output_dir=output_dir,
+          period="minutes")
```

High Frequency *OHLC* Data

Aggregating high frequency *TAQ* data into *OHLC* format with lower periodicity allows for data compression while maintaining some information about volatility,

```
> # load package "HighFreq"
> library(HighFreq)
> # define sym_bol
> sym_bol <- "SPY"
> # load OHLC data
> output_dir <- "C:/Develop/data/hfreq/scrub/"
> sym_bol <- load(
+   file.path(output_dir,
+       paste0(sym_bol, ".RData")))
> ran_ge <- "2013-11-11 09:30:00/
+       2013-11-11 10:30:00"
> chartSeries(get(sym_bol)[ran_ge],
+       name=sym_bol,
+       theme=chartTheme("white"))
```

Estimating Volatility From *OHLC* Data

Package `TTR` contains statistical estimators and technical indicators implemented in fast `C` code,

The function `volatility()` from package `TTR` estimates the volatility from *OHLC* data,

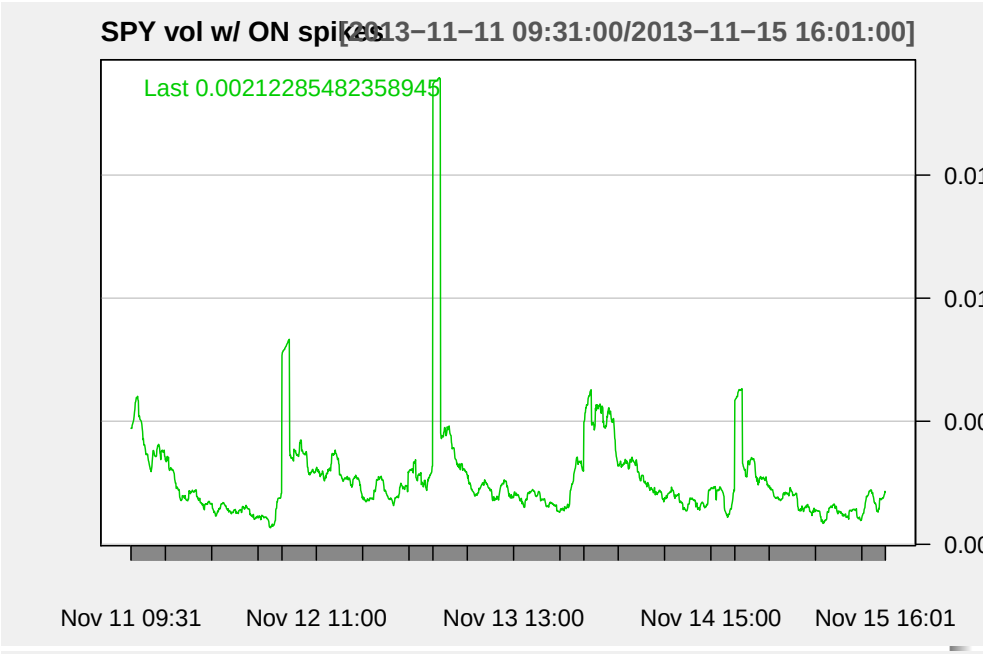
`volatility()` includes the *Garman-Klass* estimator:

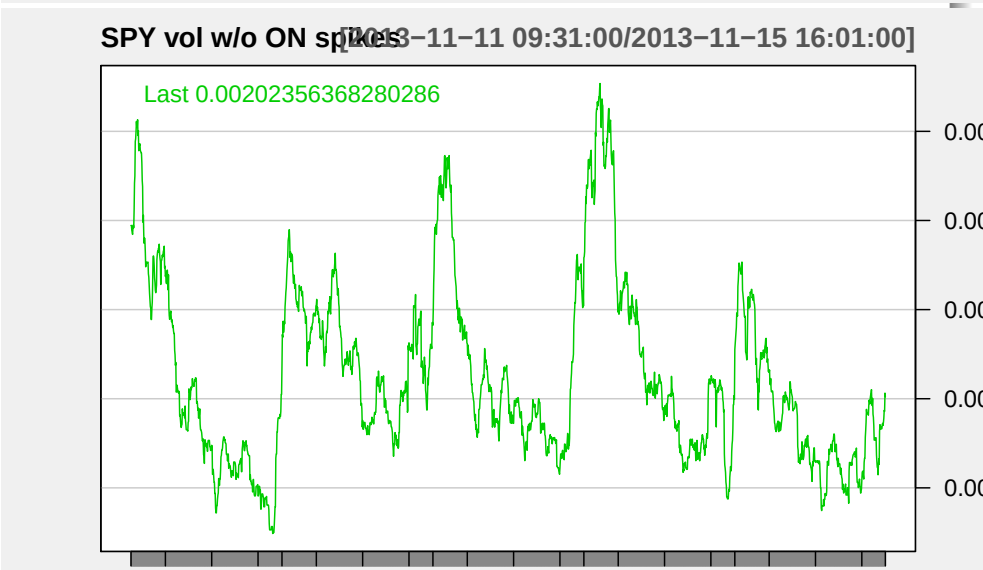
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(0.5(H_i - L_i)^2 - (2\log 2 - 1)(C_i - O_i)^2)$$

and the *Rogers-Satchell* estimator:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}((H_i - O_i)(H_i - C_i) + (L_i - O_i)(L_i - C_i))$$

```
> library(TTR)
> ran_ge <- "2013-11-11/2013-11-15"
> vol_at <- volatility(OHLC=get(sym_bol),
+                      calc="yang.zhang", n=20)
> chartSeries(vol_at[ran_ge],
+       name=paste(sym_bol, "vol w/ ON spikes"),
+       theme=chartTheme("white"))
> vol_at <- volatility(OHLC=get(sym_bol),
+                      calc="rogers.satchell", n=20)
> chartSeries(vol_at[ran_ge],
+       name=paste(sym_bol, "vol w/o ON spikes")
+       theme=chartTheme("white"))
```





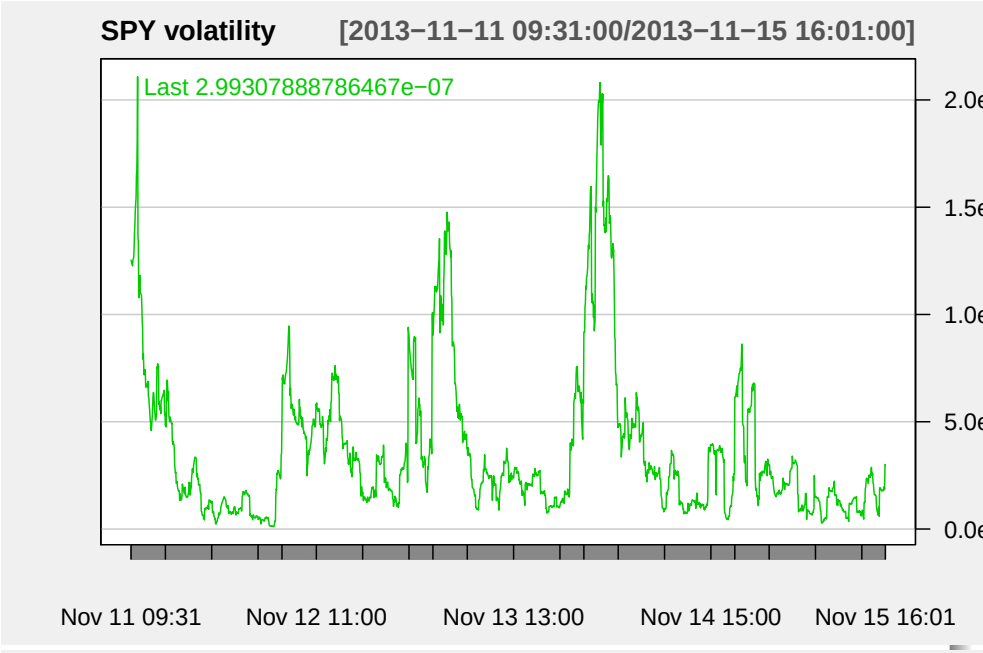
Estimating Skew From *OHLC* Data

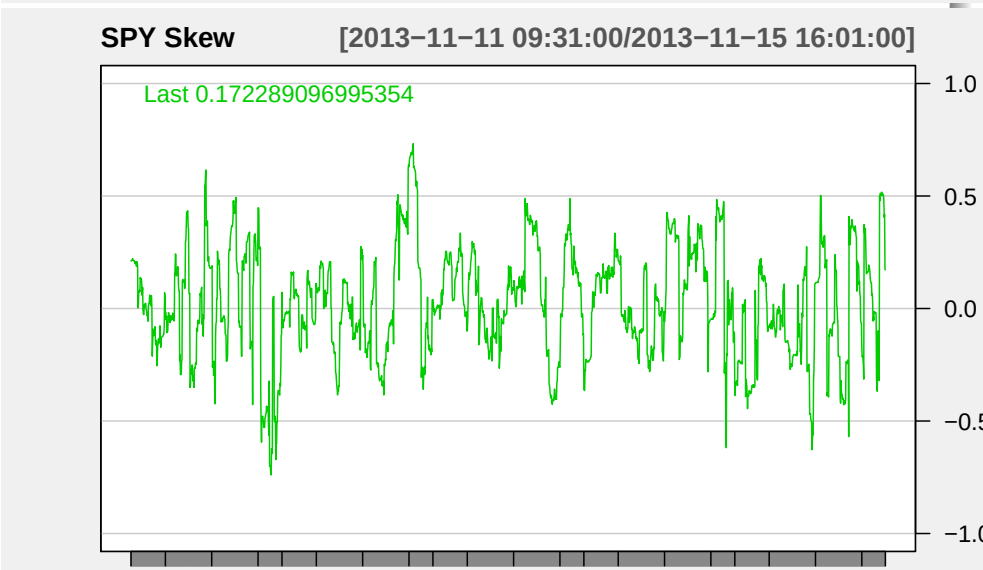
The function `skew_ohlc()` from package `HighFreq` calculates a skew-like indicator:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}((H_i - O_i)(H_i - C_i)(H_i - 0.5(O_i + C_i)) + (L_i - O_i)(L_i - C_i)(L_i - 0.5(O_i + C_i))$$

The function `run_moment_ohlc()` calculates running, volume weighted moment estimators,

```
> library(HighFreq)  # load package "HighFreq"
> # running volatility
> vol_at <- run_moment_ohlc(ohlc=get(sym_bol))
> # running skew
> sk_ew <- run_moment_ohlc(ohlc=get(sym_bol),
+                          mom_fun="skew_ohlc")
> sk_ew <- sk_ew/(vol_at)^(1.5)
> sk_ew[1, ] <- 0
> sk_ew <- na.locf(sk_ew)
> ran_ge <- "2013-11-11/2013-11-15"
> chartSeries(vol_at[ran_ge],
+       name=paste(sym_bol, "volatility"),
+       theme=chartTheme("white"))
> chartSeries(sk_ew[ran_ge],
+       name=paste(sym_bol, "Skew"),
+       theme=chartTheme("white"),
+       yrange=c(-1, 1))
```

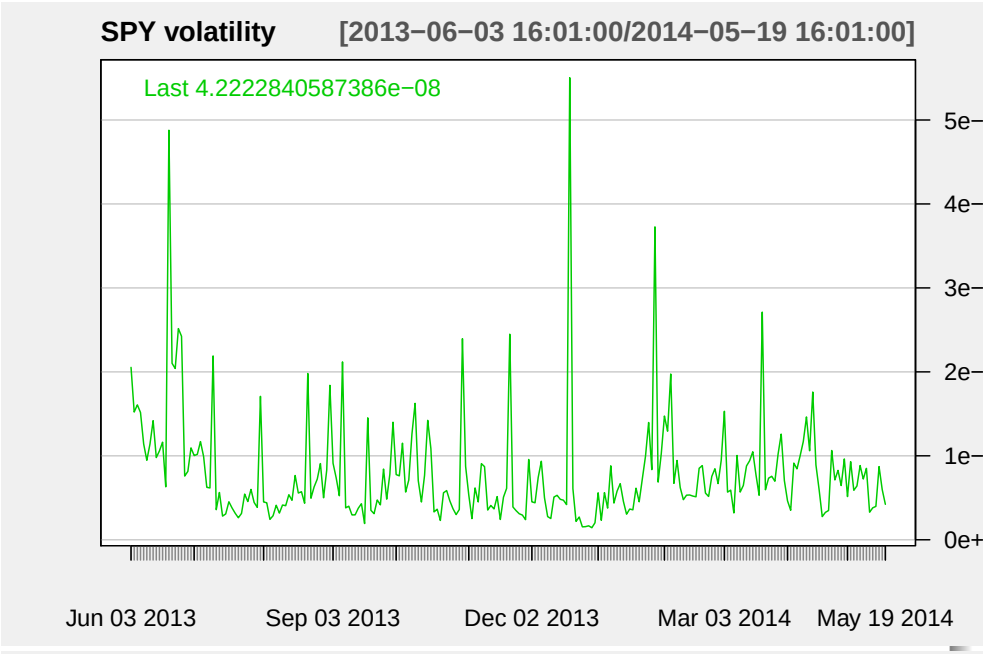


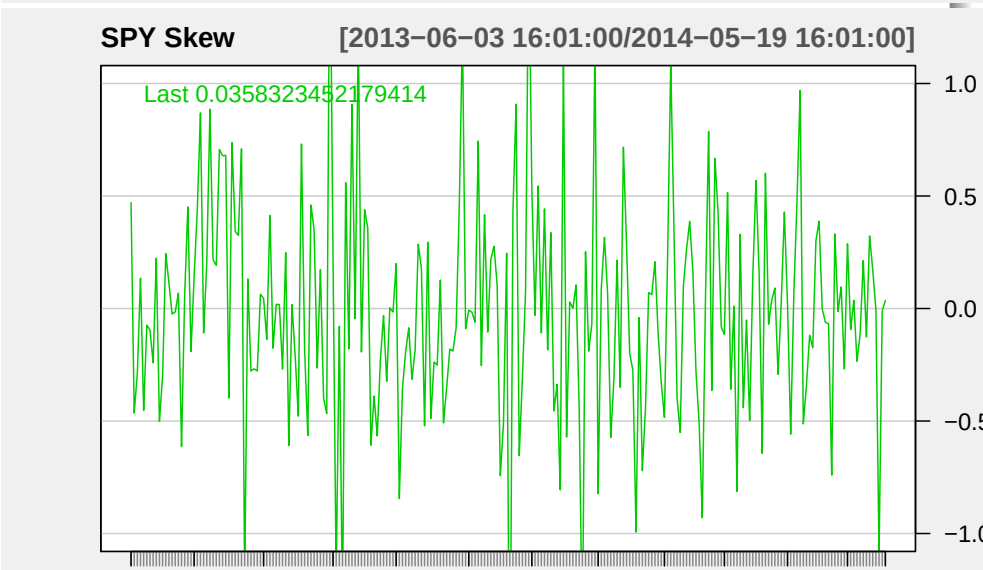


Daily Volatility and Skew From *OHLC* Data

The function `moment_ohlc()` calculates the volume weighted moment of a *OHLC* time series,

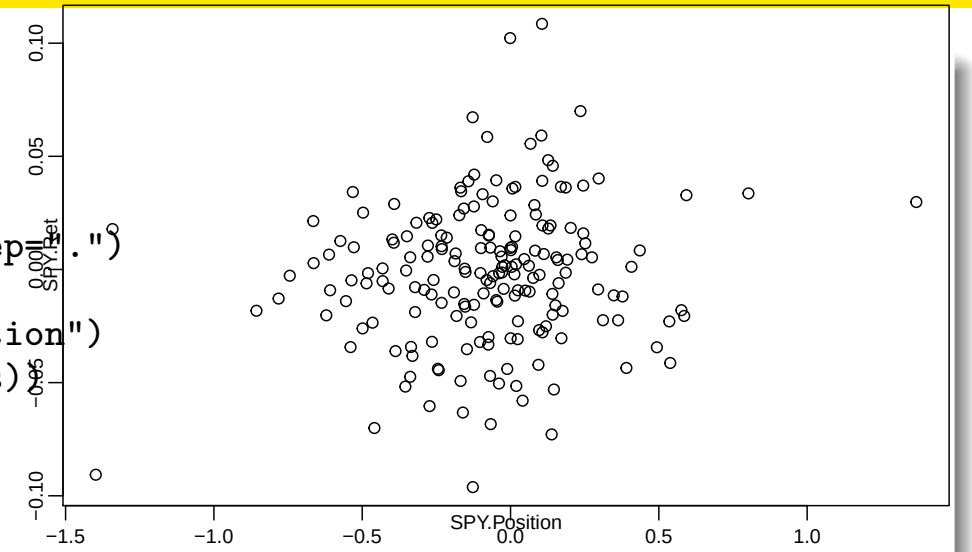
```
> library(HighFreq)  # load package "HighFreq"
> # daily volatility and skew
> vol_at <- apply.daily(x=get(sym_bol),
+                       FUN=moment_ohlc)
> colnames(vol_at) <- paste(
+   strsplit(colnames(get(sym_bol))[1],
+   split="[.]")[[1]][1], "Vol", sep=".")
> sk_ew <- apply.daily(x=get(sym_bol),
+           FUN=moment_ohlc, mom_fun="skew_ohlc")
> sk_ew <- sk_ew/(vol_at)^(1.5)
> colnames(sk_ew) <- paste(
+   strsplit(colnames(get(sym_bol))[1],
+   split="[.]")[[1]][1], "Skew", sep=".")
> ran_ge <- "2013-06-01/"
> chartSeries(vol_at[ran_ge],
+       name=paste(sym_bol, "volatility"),
+       theme=chartTheme("white"))
> chartSeries(sk_ew[ran_ge],
+       name=paste(sym_bol, "Skew"),
+       theme=chartTheme("white"),
+       yrange=c(-1, 1))
```

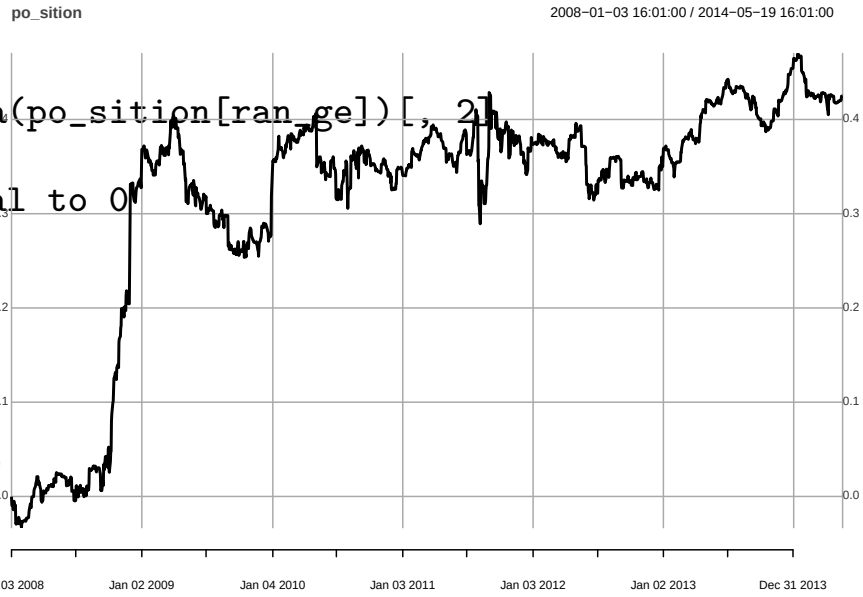




Daily Strategy Using Skew Oscillator

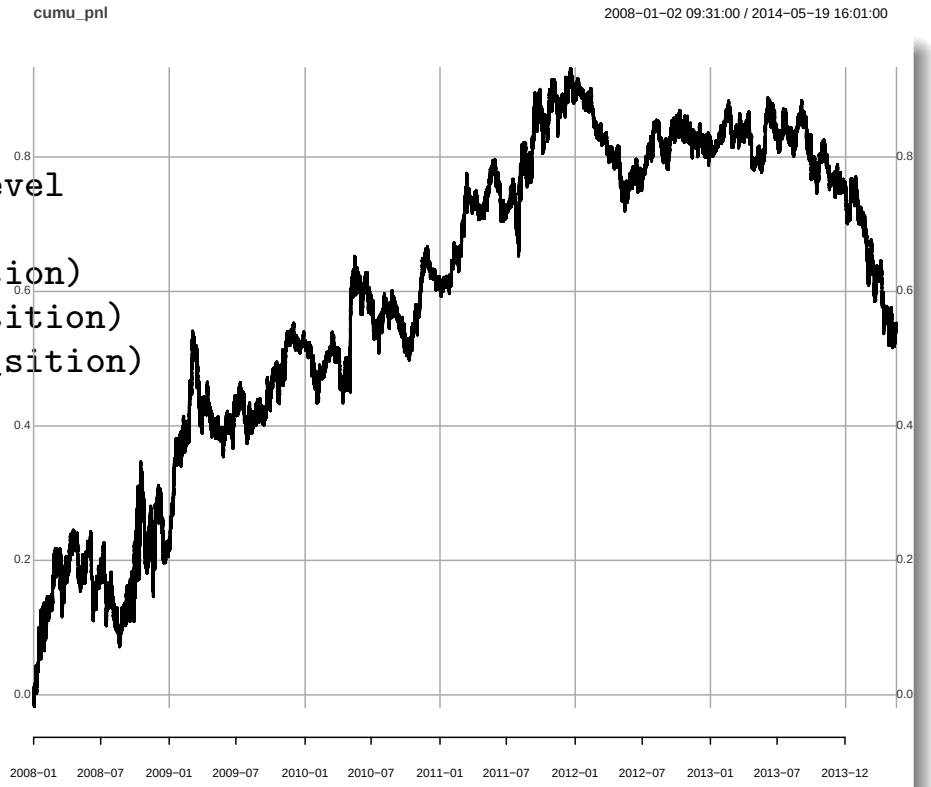
```
> # daily contrarian trading strategy
> re_turns <- get(sym_bol)[index(sk_ew), 4]
> re_turns <- diff(log(re_turns))
> colnames(re_turns) <- paste(sym_bol, "Ret", sep=".")
> po_sition <- -lag(sk_ew)
> colnames(po_sition) <- paste0(sym_bol, ".Position")
> po_sition <- na.omit(merge(po_sition, re_turns))
> # scatterplot of sk_ew and re_turns
> ran_ge <- "2008-09/2009-05"
> plot(coredata(po_sition[ran_ge]))
> cor.test(coredata(po_sition[ran_ge])[, 1],
+    coredata(po_sition[ran_ge])[, 2])

Pearson's product-moment correlation

data:  coredata(po_sition[ran_ge])[, 1] and coredata(po_sition[ran_ge])[, 2]
t = 1.9637, df = 185, p-value = 0.05107
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.0006168158  0.2806284438
sample estimates:
      cor
0.1428891
> po_sition <- cumsum(po_sition[, 1]*po_sition[, 2])
> plot(po_sition)
```





Intraday Strategy Using Skew Oscillator

```
> # intraday contrarian trading strategy
> re_turns <- calc_rets(xts_data=get(sym_bol))
> thresh_old <- 0.2  # signal threshold trading level
> po_sition <- NA*numeric(nrow(sk_ew))
> po_sition <- ifelse(sk_ew>thresh_old, -1, po_sition)
> po_sition <- ifelse(sk_ew<(-thresh_old), 1, po_sition)
> po_sition <- ifelse((sk_ew*lag(sk_ew))<0, 0, po_sition)
> # lag the po_sition
> po_sition <- c(0, po_sition)
> po_sition <- po_sition[-length(po_sition)]
> po_sition <- na.locf(po_sition)
> po_sition <- merge(sk_ew, po_sition)
> colnames(po_sition)[2] <-
+   paste0(sym_bol, ".Position")
> # cumulative PnL
> cumu_pnl <- cumsum(po_sition[, 2]*re_turns[, 1])
> plot(cumu_pnl, format.labels="%Y-%m")
```



Conclusion

Open questions:

- is there any interaction between volatility and skew?
- what is relationship between returns and cross-section of skew?
- does firm size have any effect?
- persistence of anomaly over time,

Acknowledgements:

- Brian Peterson for Thomson Reuters tick data,
- Jeffrey Ryan, Joshua Ulrich, and Brian Peterson for packages `xts`, `quantmod`, `PerformanceAnalytics`, and `TTR`,