FE570 - Homework #3

I pledge my honor that I have abided by the Stevens Honor System.

Sid Bhatia

2023-10-23

Use the same dataset as in Assignment #2 and keep only the trades taking place in the time period 10:00 to 14:00.

```r
# Load necessary packages.
library(xts)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## ###################################### WARNING ####################################
## # We noticed you have dplyr installed. The dplyr lag() function breaks how    #
## # base R's lag() function is supposed to work, which breaks lag(my_xts).      #
## #                                                                              #
## # If you call library(dplyr) later in this session, then calls to lag(my_xts) #
## # that you enter or source() into this session won't work correctly.          #
## #                                                                              #
## # All package code is unaffected because it is protected by the R namespace   #
## # mechanism.                                                                   #
## #                                                                              #
## # Set `options(xts.warn_dplyr_breaks_lag = FALSE)` to suppress this warning.  #
## #                                                                              #
## # You can use stats::lag() to make sure you're not using dplyr::lag(), or you #
## # can add conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop   #
## # dplyr from breaking base R's lag() function.                                 #
## ###################################### WARNING ####################################
```

```r
library(highfrequency)

# Load in data set.
options(digits.secs=3)
absolute_path <- 'C:/Users/sbhatia2/My Drive/University/Academics/Semester V/FE570 - Market Microstructu
load(paste(absolute_path, "sampleTQdata.RData", sep = ""))

# Added to remove warnings about time zone mismatch.
Sys.setenv(TZ='GMT')
```

```r
head(tqdata)
```

```
##                     SYMBOL EX     BID BIDSIZ     OFR OFRSIZ MODE   PRICE SIZE
## 2008-01-04 09:30:27    XXX  N 193.340    4.5 193.890   11.5   12 193.710 9100
## 2008-01-04 09:30:28    XXX  N 193.340    4.5 193.890   11.5   12 193.590  200
## 2008-01-04 09:30:29    XXX  N 193.250   12.5 193.810    8.5   12 193.445  200
## 2008-01-04 09:30:30    XXX  N 193.470    0.5 193.630    0.5   12 193.380  250
## 2008-01-04 09:30:31    XXX  N 193.470    0.5 193.630    0.5   12 193.340  300
## 2008-01-04 09:30:33    XXX  N 193.300    2.5 193.640    0.5   12 193.520  400
```

```r
tail(tqdata)
```

```
##                     SYMBOL EX     BID BIDSIZ     OFR OFRSIZ MODE   PRICE SIZE
## 2008-01-04 15:59:52    XXX  N 191.600   60.5 191.670    3.5   12 191.695  550
## 2008-01-04 15:59:55    XXX  N 191.620    0.5 191.790    1.5   12 191.620 1600
## 2008-01-04 15:59:57    XXX  N 191.600    180 191.690   27.5   12 191.690  350
## 2008-01-04 15:59:58    XXX  N 191.600    180 191.690   27.5   12 191.650  150
## 2008-01-04 15:59:59    XXX  N 191.600    180 191.690   27.5   12 191.620   50
## 2008-01-04 16:00:00    XXX  N 191.600    180 191.690   27.5   12 191.670   50
```

```r
# Filter data for trades between 10:00 and 14:00
tqdata_filtered <- tqdata["T10:00/T14:00"]

head(tqdata_filtered)
```

```
##                     SYMBOL EX     BID BIDSIZ     OFR OFRSIZ MODE   PRICE SIZE
## 2008-01-04 10:00:01    XXX  N 190.420   11.5 190.530   69.5   12 190.420  950
## 2008-01-04 10:00:02    XXX  N 190.420   11.5 190.530   69.5   12 190.400  150
## 2008-01-04 10:00:03    XXX  N 190.420   11.5 190.530   69.5   12 190.420  100
## 2008-01-04 10:00:04    XXX  N 190.420   11.5 190.530   69.5   12 190.470   50
## 2008-01-04 10:00:10    XXX  N 190.420    3.5 190.490      5   12 190.395  200
## 2008-01-04 10:00:11    XXX  N 190.395      2 190.420      8   12 190.395  200
```

```r
tail(tqdata_filtered)
```

```
##                     SYMBOL EX     BID BIDSIZ     OFR OFRSIZ MODE   PRICE SIZE
## 2008-01-04 14:00:42    XXX  N 191.140      3 191.200    4.5   12 191.130  100
## 2008-01-04 14:00:45    XXX  N 191.140      3 191.200    4.5   12 191.130   50
## 2008-01-04 14:00:50    XXX  N 191.130    0.5 191.180      1   12 191.180   50
## 2008-01-04 14:00:55    XXX  N 191.130    0.5 191.180      1   12 191.110  550
## 2008-01-04 14:00:56    XXX  N 191.110      5 191.200    3.5   12 191.080   50
## 2008-01-04 14:00:57    XXX  N 191.100    0.5 191.150    0.5   12 191.080   50
```

Problem 1

1. How many trades are taking place during this time?

```r
num_of_trades <- nrow(tqdata_filtered)

num_of_trades
```

```
## [1] 4691
```

There are **4691** trades taking place during this time.

2. Compute the Quoted Spread and Effective Spread, averaged over all trades.

$$\text{Quoted Spread} = \text{Best Ask} - \text{Best Bid}$$

$$\text{Effective Spread} = 2 * D_t * (\text{Price}_t - \frac{1}{2}(a_t + b_t))$$

```
colnames(tqdata_filtered)
```

```
## [1] "SYMBOL" "EX"     "BID"    "BIDSIZ" "OFR"    "OFRSIZ" "MODE"   "PRICE"
## [9] "SIZE"
```

```
# Retrieve liquidity measures using `highfrequency` package.
liqudity_measures <- getLiquidityMeasures(tqdata_filtered)

head(liqudity_measures$effectiveSpread)
```

```
##                     effectiveSpread
## 2008-01-04 10:00:01    1.100000e-01
## 2008-01-04 10:00:02    1.500000e-01
## 2008-01-04 10:00:03    1.100000e-01
## 2008-01-04 10:00:04    1.000000e-02
## 2008-01-04 10:00:10    1.200000e-01
## 2008-01-04 10:00:11    2.500000e-02
```

```
# Retrieve trade signs using Lee-Ready method.
trade_signs <- getTradeDirection(tqdata_filtered)

# Retrieve trading price, bid, ask, and mid prices, respectively.
price <- as.numeric(tqdata_filtered$PRICE)
bid <- as.numeric(tqdata_filtered$BID)
ask <- as.numeric(tqdata_filtered$OFR)

mids <- (bid + ask) / 2

# Compute Quoted Spread.
quoted_spread <- mean(as.numeric(tqdata_filtered$OFR) - as.numeric(tqdata_filtered$BID))
quoted_spread
```

```
## [1] 0.0840226
```

```
# Compute Effective Spread.
effective_spread <- mean(2 * trade_signs * (price - mids))
effective_spread
```

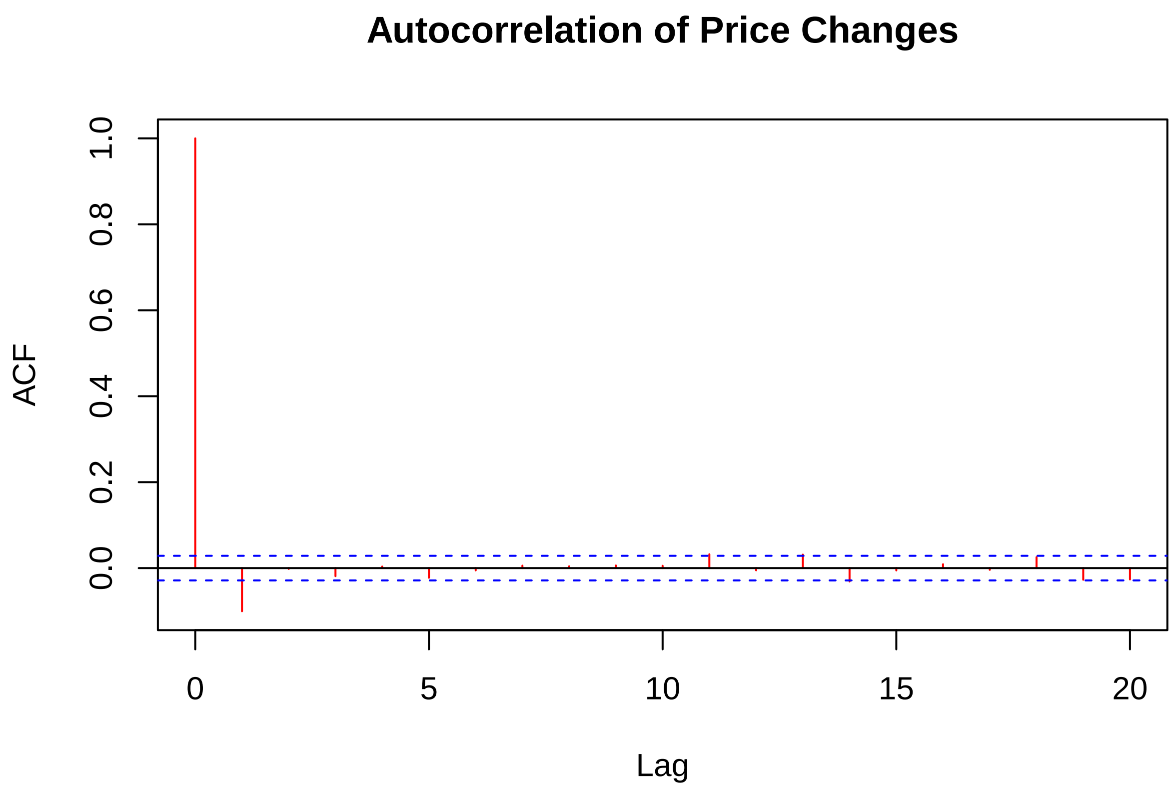
```
## [1] 0.07905137
```

As such, the average quoted spread is **0.084**$ and the average effective spread is **0.079**.

3. Calibrate the Roll model to this data, and find the parameters c (cost of trading) and σ_u (volatility of the efficient price).

```
# Calculate the difference in price changes.
dprice <- diff(price)

# Compute and plot the autocorrelation of price changes.
ac_pr <- acf(dprice, lag.max=20, type="correlation", plot=FALSE)
plot(ac_pr, col="red", main="Autocorrelation of Price Changes")
```




```
# Compute the covariances of the price changes.
covpr <- acf(dprice, lag.max=20, type="covariance", plot=FALSE)

# Retrieve gamma0 as the covariance at lag 0.
gamma0 <- covpr$acf[1]
gamma0
```

```
## [1] 0.002502923
```

```
# Retrieve gamma1 as the covariance at lag 1.
gamma1 <- covpr$acf[2]
gamma1
```

```
## [1] -0.0002512731
```

```
cparam <- sqrt(-gamma1)
cparam
```

```
## [1] 0.0158516
```

```
sig2u <- gamma0 + 2 * gamma1
sigu <- sqrt(sig2u)

cparam
```

```
## [1] 0.0158516
```

```
sigu
```

```
## [1] 0.04472557
```

As such, $c = 0.0159$ and $\sigma_u = 0.0447$.

4. Compute the Roll model estimate for the bid-ask spread, and compare with the spreads computed in point 2. Which one do you think is more accurate?

```
roll_spread <- cparam * 2
roll_spread
```

```
## [1] 0.03170319
```

As such, the Roll's model estimate of the bid-ask spread is **0.0317**.

I believe the quoted and effective spreads are more accurate since they are closer in difference to one another and the effective spread captures the real cost of trading, as it takes into consideration the price improvement or disimprovement a trader receives compared to the quoted spread.

The quoted spread is the difference between the best bid and the best ask prices in the market. It is an explicit measure and is easily observable in markets with transparent limit order books. The quoted spread gives a direct measure of the cost to trade immediately using market orders. However, it does not always capture the actual cost of trading, especially in markets where the best bid and ask prices might not represent substantial volume or where traders frequently transact inside the quoted spread.

Roll's measure is an implicit measure of the bid-ask spread derived from serial price changes in the absence of trade data. It is based on the covariance between consecutive price changes. While Roll's model provides an estimate of the spread in markets where direct spread measures are not observable, it assumes that prices follow a random walk and that there's no other noise in the price process besides the bid-ask bounce. In reality, other factors, such as volatility and market microstructure effects, can affect price changes, so Roll's measure can be imprecise in many real-world contexts.

As such, the effective spread, quoted spread, and Roll's model estimate for bid-ask spread would be the order from most to least accurate.