**Name:** Nguyen Thi Cam Hoang **Date:** June 7th, 2017

**Student ID:** 20165327

# Assignment 3

1. **Merging Trade and Quote Data**

* Date of analysis: **August 26, 2015**
* Information of stock for analysis:

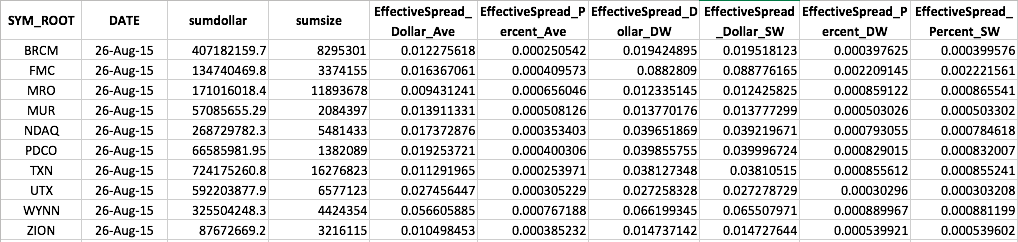
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Cusip number (8 digit)** | **Cusip number (9 digit)** | **Stock Market** | **Symbol** | **Company name** |
| 1 | 62671710 | 302491303 | N | FMC | FMC Corp |
| 2 | 56584910 | 565849106 | MRO | Marathon Oil Corporation |
| 3 | 62671710 | 626717102 | MUR | Murphy Oil Corporation |
| 4 | 88250810 | 882508104 | TXN | Texas Instruments Incorporated |
| 5 | 91301710 | 913017109 | UTX | United Technologies Corporation |
| 6 | 11132010 | 111320107 | Q | BRCM | Broadcom Ltd |
| 7 | 63110310 | 631103108 | NDAQ | Nasdaq Inc |
| 8 | 70339510 | 703395103 | PDCO | Patterson Companies, Inc. |
| 9 | 98313410 | 983134107 | WYNN | Wynn Resorts, Limited |
| 10 | 98970110 | 989701107 | ZION | Zions Bancorp |

To compute the liquidity variables, these steps were performed:

* **Step 1**: Clean the NBBO, Trade, Quote data.
* **Step 2**: Create the merged data: OfficialCompleteNBBO, TradesandCorrespondingNBBO, BuySellIndicators.
* **Step 3**: Compute the spreads

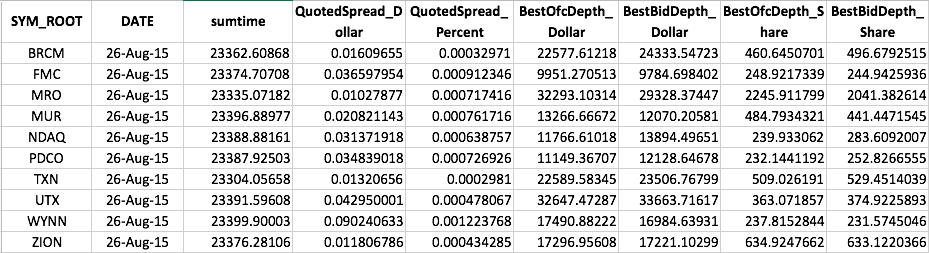
1. Report the effective spreads of selected stocks.

* The effective spreads were derived from **step 11** in the provided code.
* The data was used: **BuySellIndicators**.



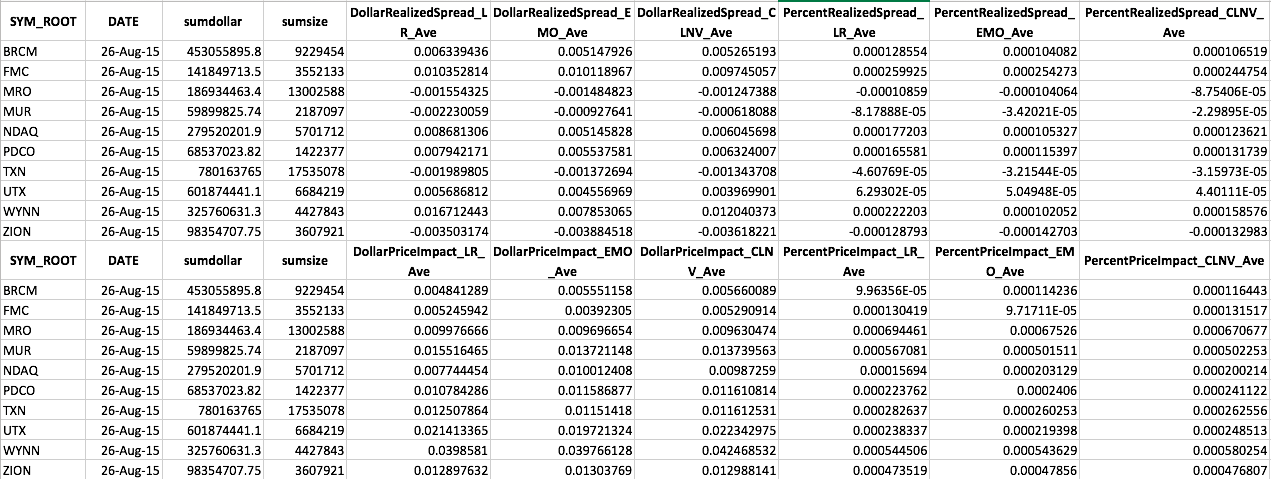
1. Report the quoted spreads of selected stocks.

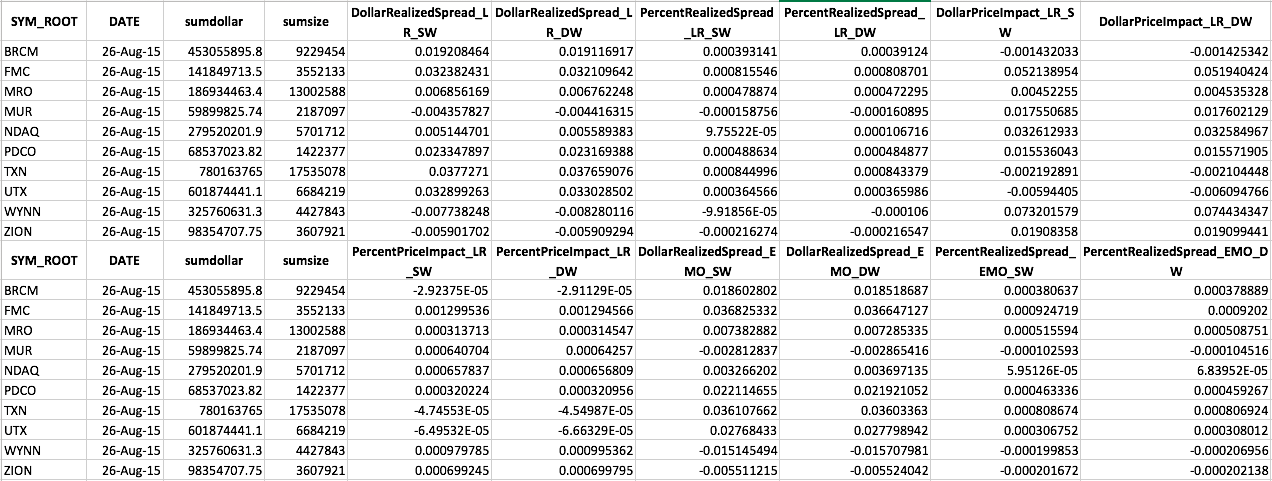
* The quoted spreads were derived from **step 12** in the provided code.
* The data was used: **OfficialCompleteNBBO** and **QSpread**.

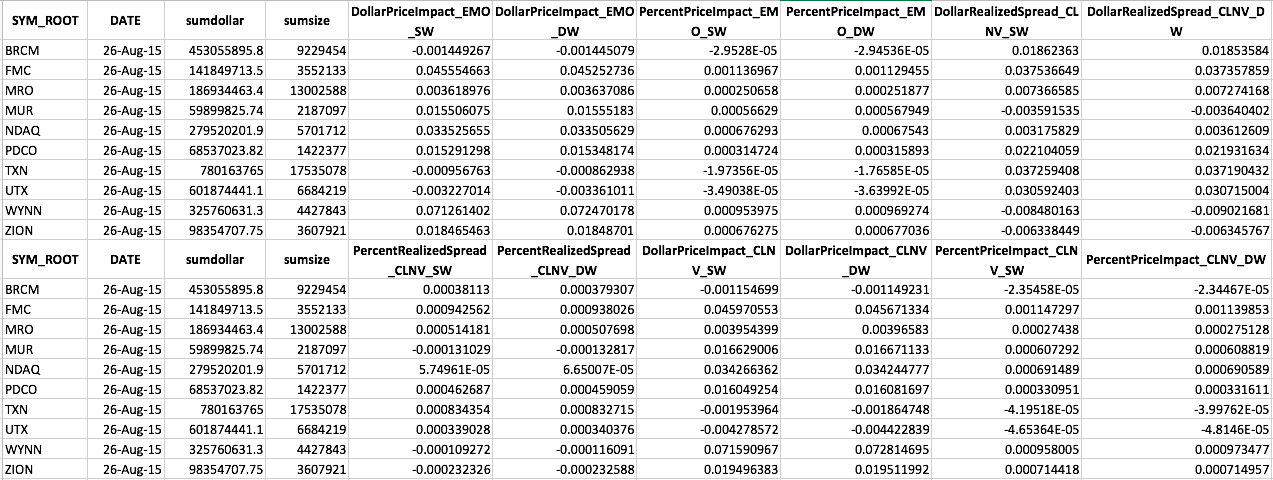


1. Report the realized spreads of selected stocks.

* The realized spreads were derived from **step 12** in the provided code.
* The data was used: **BuySellIndicators** and **QSpread**.







1. **Time-Series Analysis**
2. For each return series, test the null hypothesis that the first 12 lags of autocorrelations are zero at 5% level.

* This is done by using the following statement in sas:

proc arima data=trade;

by sym\_root;

identify var=ret;

run;

quit;

* Significant level:
* Hypothesis:
* Null hypothesis: lags of autocorrelations are zero
* Alternative hypothesis: at least one lag is not zero

Or:

* : = 0 for all l
* : 0 for at least one l

## Result

|  |  |  |
| --- | --- | --- |
|  | **Autocorrelation check** | **Correlation panel** |
| **BRCM** |  |  |
| **FMC** |  |  |
| **MRO** |  |  |
| **MUR** |  |  |
| **NDAQ** |  |  |
| **PDCO** |  |  |
| **TXN** |  |  |
| **UTX** |  |  |
| **WYNN** |  |  |
| **ZION** |  |  |

### Interpretation:

* We only consider to the lag from 1 to 12.
* As can be seen from all tables and figures above, there is a common trend across all stocks. The p-value is smaller than the significant level of 0.05. Hence, there is a strong evidence against the null hypothesis. In another words, we can reject the null hypothesis under the significant level of 0.05.
* Besides, from the figures, we can observe that the original variables are not stationary. Look at the ACF and PACF, the ACF tails off and PACF cuts off after lag 1, this may suggest us to use AR(1) model. However, the optimize p value for AR can be estimated later.

**For problem (b) and (c), I will analyse each stock one by one**

1. Choose the appropriate lags for building ARIMA model based on the SCAN and ESACF methods.

* Data trade used for this analysis is already processed from problem 1, this data satisfies all the conditions of trading.
* SCAN is done by using:

proc arima data=trade;

by sym\_root;

identify var=ret scan minic; \*option scan minic;

run;

quit;

* ESACF is done by using:

proc arima data=trade;

by sym\_root;

identify var=ret esacf minic; \*option esacf minic;

run;

quit;

1. Build an ARIMA model for the series:

* For simplicity, I hereby only consider the model to the difference of 1. In order to compare the performance of SCAN and ESACF method, the values of p and q smaller than 5 will be shown. This is because in some model, the the SCAN and ESACF report the optimal value of (p+d) is between the interval (5, 10).
* **ARIMA(p,0,q)**

proc arima data=trade;

by sym\_root;

identify var=ret;

estimate p=0 q=3 method=ml;

*\*using maximum likelihood method and*

*setting value for p and q depends on each instance;*

run;quit;

* **ARIMA(p,1,q)**

proc arima data=trade;

by sym\_root;

identify var=dret; *\*dret is the first order difference of ret*

estimate p=4 q=5 method=ml; \*maximum likelihood;

*\*using maximum likelihood method and*

*setting value for p and q depends on each instance;*

run;quit;

BRCM

|  |  |  |
| --- | --- | --- |
|  | SCAN does not output the ARMA result within the value of 5 | Use p+d = 0 and q=5 |
|  |  |

**ARIMA model**

|  |
| --- |
| **ARIMA(0,0,5)** |
|  |

**Final model:**

ARIMA(0,0,5)

****

## FMC

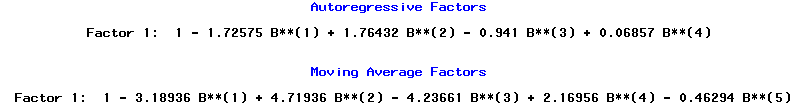
|  |  |  |
| --- | --- | --- |
|  |  | Use p+d = 5 and q=5 for further analysis |
|  |  |

**ARIMA model**

|  |  |
| --- | --- |
| **ARIMA(5,0,5)** | **ARIMA(4,1,5)** |
|  |  |

**Final model:** The model is chosen based on the lowest AIC

ARIMA(4,1,5)

****

MRO

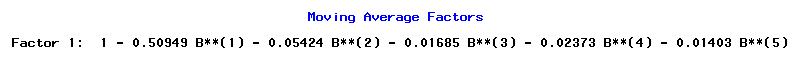
|  |  |  |
| --- | --- | --- |
|  |  | p+d=3 and q=1/p+d=0 and q=5 for further analysis (lowest BIC) |
|  |  |

**ARIMA model**

|  |  |
| --- | --- |
| **ARIMA(3,0,1)** | **ARIMA(2,1,1)** |
|  |  |
| **ARIMA(0,0,5)** |  |
|  |  |

**Final model:** The model is chosen based on the lowest AIC

ARIMA(0,0,5)

****

## MUR

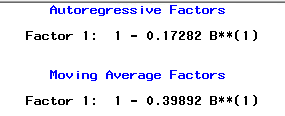
|  |  |  |
| --- | --- | --- |
|  |  | Use p+d = 1 and q=1 for further analysis (lowest BIC) |
|  |  |

**ARIMA model**

|  |  |
| --- | --- |
| **ARIMA(1,0,1)** | **ARIMA(0,1,1)** |
|  |  |

**Final model:** The model is chosen based on the lowest AIC

ARIMA(1,0,1)

****

## NDAQ

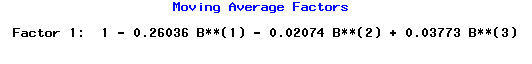
|  |  |  |
| --- | --- | --- |
|  | SCAN does not output the ARMA result within the value of 5 | Use p+d = 2 and q=3 for further analysis (lowest BIC) |
|  |  |

**ARIMA model**

|  |
| --- |
| **ARIMA(0,0,3)** |
|  |

**Final model:**

ARIMA(0,0,3)

****

## PDCO

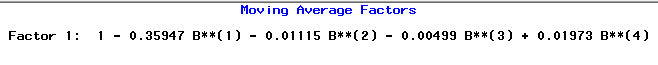
|  |  |  |
| --- | --- | --- |
|  | SCAN does not output the ARMA result within the value of 5 | Use p+d = 0 and q=4 for further analysis (lowest BIC) |
|  |  |

**ARIMA model**

|  |
| --- |
| **ARIMA(0,0,4)** |
|  |

**Final model:**

ARIMA(0,0,4)

****

## TXN

|  |  |  |
| --- | --- | --- |
|  |  | Use p+d = 0 and q=5 for further analysis (lowest BIC) |
|  |  |

**ARIMA model**

|  |
| --- |
| **ARIMA(0,0,5)** |
|  |

**Final model:**

ARIMA(0,0,5)



## UTX

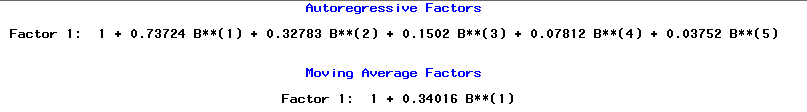
|  |  |  |
| --- | --- | --- |
|  |  | Use p+d = 5 and q=1 for further analysis (lowest BIC) |
|  | ESACF does not output the ARMA result within the value of 5 |

**ARIMA model**

|  |  |
| --- | --- |
| **ARIMA(5,0,1)** | **ARIMA(4,1,1)** |
|  |  |

**Final model:** The model is chosen based on the lowest AIC

ARIMA(5,0,1)



## WYNN

|  |  |  |
| --- | --- | --- |
|  |  | Use p+d = 5 and q=0 for further analysis (lowest BIC) |
|  |  |

**ARIMA model**

|  |  |
| --- | --- |
| **ARIMA(5,0,0)** | **ARIMA(4,1,0)** |
|  |  |

**Final model:** The model is chosen based on the lowest AIC

ARIMA(5,0,0)



ZION

|  |  |  |
| --- | --- | --- |
|  |  | Use p+d = 5 and q=5 for further analysis |
|  |  |

**ARIMA model**

|  |  |
| --- | --- |
| **ARIMA(5,0,5)** | **ARIMA(4,1,5)** |
|  |  |

**Final model:** The model is chosen based on the lowest AIC

ARIMA(5,0,5)

