

Name: Nguyen Thi Cam Hoang
Student ID: 20165327

Date: April 27th, 2017

Assignment 2

1. Preparing the TAQ data

- a) Get information
- Find the prior day's closing day by Yahoo Finance: **June 5, 2013**

```
The SAS System                               1
                                           19:01 Wednesday, April 26, 2017

Obs      CUSIP      SYMBOL_      SYMBOL_
          CUSIP      ROOT       SUFFIX
1         626717102    MUR
2         703395103    PDCO
```

- I have information as below:

Cusip number (8 digit)	Cusip number (9 digit)	Stock Market	Symbol	Company name
62671710	626717102	N	MUR	MURPHY OIL CORP
70339510	703395103	Q	PDCO	PATTERSON COMPANIES INC

b) Export data.

All the data are retrieved by following these rules:

- Only quotes/trades during normal market hours are considered
 - Quotes : 9:00 am ~ 4:00 pm
 - Trades : 9:30 am ~ 4:00 pm
- Negative values of size or prices
- Normal quote conditions
 - Qu_cond : other than (A, B, H, O, R, and W) delete
- Canceled quote
 - Qu_cancel : B(delete)
- Normal Trade conditions
 - Tr_corr : other than "00" delete

After cleaning and screening data, 3 different data; trade, quote, nbbo; are exported into one excel file by using:

```
proc export data=ct
outfile='C:\Users\Hoang\Google
Drive\UNIST\3rd_Courses\finance\assignment\AS2\data\data.xlsx'
dbms=xlsx;
sheet='trade';
```

```
proc export data=cq
outfile='C:\Users\Hoang\Google
Drive\UNIST\3rd_Courses\finance\assignment\AS2\data\data.xlsx'
```

```
dbms=xlsx replace;
sheet='quote';
```

```
proc export data=nbbo
outfile='C:\Users\Hoang\Google
Drive\UNIST\3rd_Courses\finance\assignment\AS2\data\data.xlsx'
dbms=xlsx replace;
sheet='nbbo';
run;
```

2. Trade data analysis

After preprocessing, the trade data will be like:

	A	B	C	D	E	F	G	H
1	DATE	TIME_M	EX	SYM_ROOT	SIZE	PRICE	TR_SEQNUM	type
2	06/06/2013	09:30	N	MUR	6523	62.99	2140	T
3	06/06/2013	09:30	B	MUR	100	62.98	5188	T
4	06/06/2013	09:30	J	MUR	100	62.8	5273	T
5	06/06/2013	09:30	N	MUR	100	62.96	5274	T
6	06/06/2013	09:30	T	MUR	10	62.8	5319	T
7	06/06/2013	09:30	D	MUR	100	62.891	5320	T
8	06/06/2013	09:30	P	MUR	100	62.83	6783	T
9	06/06/2013	09:30	Y	MUR	100	62.83	6784	T
10	06/06/2013	09:30	P	MUR	100	62.83	6785	T
11	06/06/2013	09:30	T	MUR	100	62.83	6786	T
12	06/06/2013	09:30	B	MUR	100	62.83	6787	T
13	06/06/2013	09:30	N	MUR	100	62.81	6788	T
14	06/06/2013	09:30	N	MUR	100	63	6789	T
15	06/06/2013	09:30	K	MUR	100	62.83	6790	T

- a) What is the largest single trade of the day? When was it executed? at what price?

It can be done by using **max** function in excel.

Symbol	Time	Price	Size
MUR	09:30:31	62.99	6,523
PDCO	09:50:57	38.42	10,000

- b) How much volume (in shares) was done on each exchange? Which exchange has the largest volume? What is the market share of each reporting exchange?

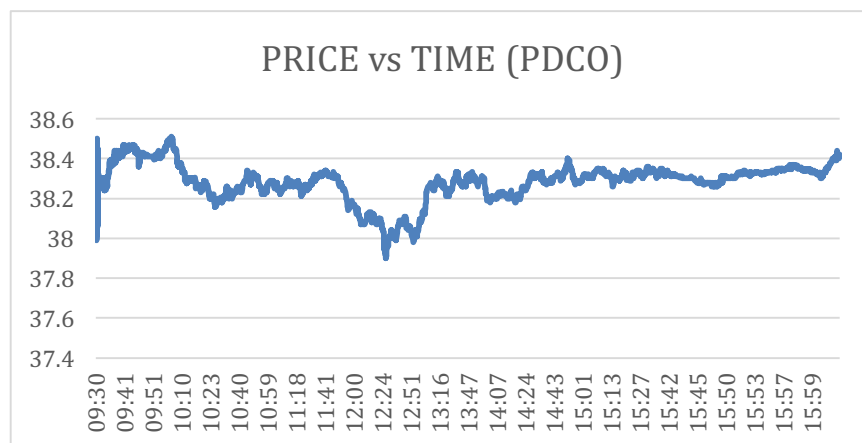
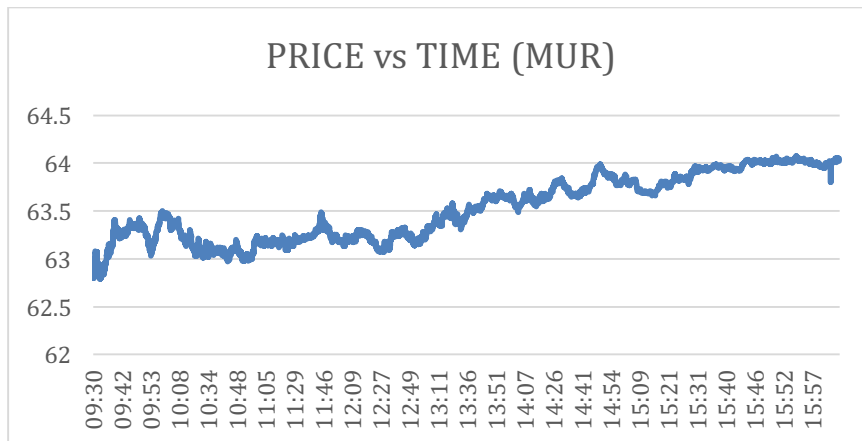
It can be done by using **Pivot** table

Sum of SIZE	Column Labels				
Row Labels	MUR	PDCO	(blank)	Grand Total	Market shares
B	16229	20455		36684	1.95%
C	3000	435		3435	0.18%
D	230719	164671		395390	20.97%
J	64970	15363		80333	4.26%
K	69805	108990		178795	9.48%
M	5000			5000	0.27%
N	322045			322045	17.08%
P	132161	66708		198869	10.55%
Q		287085		287085	15.23%
T	221785			221785	11.76%
W	9733	2600		12333	0.65%
X	900	2292		3192	0.17%
Y	12600	5755		18355	0.97%
Z	68826	53487		122313	6.49%
(blank)					
Grand Total	1157773	727841		1885614	100%

- The exchange with the largest volume is:

- MUR: N
- PDCO: Q
- Total: D

c) Between 9:00 and 16:00, plot the trades. That is, plot price versus time. Price vs time plotting



d) Construct 30-minutes time buckets, i.e. 9:30-10:00, 10:01-10:30, ... In each interval, compute the total number of trades. Again, compute the market share of each reporting exchange within time interval.

- To construct 30-min time bucket, the trading time are rounded to the lower limit of the interval. For example, trading time 9:37 will be rounded to 9:30. By taking advantage of Pivot, the below result is achieved.

Sum of SIZE	Column Labels														Grand Total
	B	C	D	J	K	M	N	P	Q	T	W	X	Y	Z	
MUR	16229	3000	230719	64970	69805	5000	322045	132161		221785	9733	900	12600	68826	1157773
09:30	2800	1100	16003	2800	10955		34991	10039		13173	400		2800	6800	101961
10:00	1900	400	11187	2100	3000		14718	8808		10941	700		400	3200	57354
10:30	2100	200	29540	4900	1300		18388	3601		18786	1100		700	5018	85633
11:00	400		10567	3764	2000		11186	10336		11408	300	100	800	2700	53561
11:30	4580		5953	3100	3400		8900	15648		15218	500		500	1700	59499
12:00	400		16200	8300	10100		15171	6545		17875	500		500	4800	80391
12:30	300		13013	4827	6160		18798	5600		7743	800	100	300	3200	60841
13:00	300		6071	3468	2800		16691	9323		12694	200		600	2603	54750
13:30	700		15035	5150	4289		23147	8574		13850	200		700	5200	76845
14:00	400		6366	4200	3750		27673	9050		15824	400		700	5200	73563
14:30	200		14949	4900	6600		22489	12504		15592	600		1500	6700	86034
15:00	400		21332	5261	7651		28694	13786		22980	833		600	5000	106537
15:30	1749	1300	64503	12200	7800	5000	81199	18347		45701	3200	600	2500	16705	260804
PDCO	20455	435	164671	15363	108990			66708	287085		2600	2292	5755	53487	727841
09:30	2100		27000	1100	7197			5407	29957		100		300	4800	77961
10:00	1100		20193	1500	5800			3838	22334				100	3500	58365
10:30	800		6600	800	7506			5700	13709		400		200	3700	39415
11:00	500		6010	1000	6498			3834	11387		300			2300	31829
11:30	300		6078	400	13116			3205	12928		200	500	100	2800	39627
12:00	400		8375	100	5300			5466	11370					2400	33411
12:30	500		1700	500	6699			2600	8900		300			1575	22774
13:00	200		5298	1100	4100			2300	13522		300			3300	30120
13:30	200		3449	500	3000			2200	12715		100	100		3100	25364
14:00	400		15777	700	6200			6300	27447		200		200	3000	60224
14:30	1200		7352	1500	5900			2700	14847		100		700	2700	36999
15:00	2098		15579	500	9958			7093	20314				900	4000	60442
15:30	10657	435	41260	5663	27716			16065	87655		600	1692	3255	16312	211310
Grand Total	36684	3435	395390	80333	178795	5000	322045	198869	287085	221785	12333	3192	18355	122313	1885614

3. Quote data analysis

After preprocessing, the Quote data will be:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DATE	TIME_M	EX	SYM_ROOT	BID	BIDSIZ	ASK	ASKSIZ	QU_CONC	QU_SEQNUM	NATBBO_IND	NASOBBO_IND	Spread
2	06/06/2013	09:00	Y	MUR	62.36	6	64.99	6	R	384086	0	2	2.63
3	06/06/2013	09:01	Y	MUR	62.36	5	64.99	6	R	389331	0	2	2.63
4	06/06/2013	09:02	Y	MUR	62.36	5	64.99	6	R	393982	0	2	2.63
5	06/06/2013	09:03	Y	MUR	62.36	5	64.99	5	R	399256	0	2	2.63
6	06/06/2013	09:04	Y	MUR	62.36	6	64.99	5	R	405082	0	2	2.63
7	06/06/2013	09:05	Y	MUR	62.36	5	64.99	6	R	410225	0	2	2.63
8	06/06/2013	09:06	Y	MUR	62.36	6	64.99	6	R	415407	0	2	2.63
9	06/06/2013	09:06	Z	MUR	62.36	9	64.87	2	R	417303	6	2	2.51
10	06/06/2013	09:06	Z	MUR	62.71	2	64.87	2	R	418338	1	2	2.16
11	06/06/2013	09:07	Y	MUR	62.36	5	64.99	5	R	420251	0	2	2.63
12	06/06/2013	09:08	Y	MUR	62.36	5	64.99	6	R	425720	0	2	2.63
13	06/06/2013	09:09	Y	MUR	62.36	5	64.99	6	R	431296	0	2	2.63
14	06/06/2013	09:10	Y	MUR	62.36	5	64.99	6	R	436569	0	2	2.63
15	06/06/2013	09:20	Z	MUR	62.36	9	64.87	2	R	475129	6	2	2.51

As can be seen, there is no information about best bid in the data. This information cannot be derived easily. To be more specified, in order to observe the liquidity as well as the direction of buy/sell, we can merge trade, quote and nbbo data for more detail analysis. This kind of data will look like:

DATE	TIME_M	SYM_ROOT	SYM_SUFFIX	best_asksize	shares	type	EX	SIZE	PRICE	TR_SEQNUM	Qtime	nbo	nbb	nboqty	nbbqty
06-Jun-13	09:06	MUR		200	Q						09:06	64.87	62.7	200	100
06-Jun-13	09:06	MUR		200	Q						09:06	64.87	62.71	200	200
06-Jun-13	09:20	MUR		200	Q						09:20	64.87	62.71	200	200
06-Jun-13	09:28	MUR		300	Q						09:28	64.86	62.72	300	300
06-Jun-13	09:28	MUR		300	Q						09:28	64.86	62.73	300	200
06-Jun-13	09:28	MUR		200	Q						09:28	64.85	62.73	200	200
06-Jun-13	09:30	MUR		100	Q						09:30	64.84	62.73	100	200
06-Jun-13	09:30	MUR		200	Q						09:30	64.2	62.73	200	200
06-Jun-13	09:30	MUR		300	Q						09:30	64.19	62.73	300	200
06-Jun-13	09:30	MUR		400	Q						09:30	64.19	62.73	400	200
06-Jun-13	09:30	MUR		200	Q						09:30	63.06	62.73	200	200
06-Jun-13	09:30	MUR			T	N		6523	62.99	2140	09:30	63.06	62.73	200	200
06-Jun-13	09:30	MUR		700	Q						09:30	63	62.79	700	200
06-Jun-13	09:30	MUR		700	Q						09:30	63	62.79	700	200

In this exercise, for simplicity, I assume that the bid is represented as best bid so that we can see the difference between problem 3 and 4.

- a) During the day, how much the number of quote updates was done one each exchange? Which exchange has frequently updated the quotes? For what proportion of the quotes is the primary listing exchange at the best bid? (Note that the best bid or the best ask do not have to be alone. For example, if the primary exchange is the NYSE, the NYSE is at the best bid if the exchange of the best bid is marked as "N", "XN", or "BNZ".)

Count of EX	Column Labels		Proportion				
Row Labels	MUR	PDCO	(blank)	Grand Total	MUR	PDCO	Grand Total
B	14683	16274		30957	4.27%	10.73%	6.25%
C	40547	2535		43082	11.79%	1.67%	8.69%
J	44680	16045		60725	12.99%	10.58%	12.25%
K	30063	10120		40183	8.74%	6.67%	8.11%
N	50157			50157	14.58%	0.00%	10.12%
P	25582	16453		42035	7.44%	10.84%	8.48%
T	58508	44564		103072	17.01%	29.37%	20.79%
W	27356	8910		36266	7.95%	5.87%	7.32%
X	4505	2485		6990	1.31%	1.64%	1.41%
Y	12673	14484		27157	3.68%	9.55%	5.48%
Z	35233	19848		55081	10.24%	13.08%	11.11%
(blank)							
Grand Total	343987	151718		495705	100%	100%	100%

From the above table, it can be seen that T and J exchange were high-frequent quoted in the market. Therefore, these stocks seem to be more liquid compared to the others.

b) In each of the 30-minute time buckets, do the same thing for the above question.

Count of EX	Column Labels												
Row Labels	B	C	J	K	N	P	T	W	X	Y	Z	(blank)	Grand Total
MUR	14683	40547	44680	30063	50157	25582	58508	27356	4505	12673	35233		343987
09:00:00							2	1		11	6		20
09:30:00	949	1666	2861	2822	3539	1313	2632	1585	446	1134	1655		20602
10:00:00	921	1812	3014	3120	2938	1355	2616	1328	377	986	1421		19888
10:30:00	1576	2484	2909	3228	3977	1448	3894	1229	449	1399	2375		24968
11:00:00	1505	2135	3715	2501	3590	1844	4126	3604	380	1098	2859		27357
11:30:00	1477	3279	2715	2694	3459	1789	4056	1455	230	1013	2671		24838
12:00:00	1423	4507	3931	2726	3444	1940	4415	2481	411	1011	2784		29073
12:30:00	1579	3532	2406	2520	3489	1985	4485	1548	306	1135	2682		25667
13:00:00	1514	4508	4106	2613	4188	2753	5391	4415	432	1131	3831		34882
13:30:00	881	3148	2860	1569	3294	2222	4729	1361	340	687	3024		24115
14:00:00	672	3063	3316	1751	3374	1893	4775	1720	348	626	2277		23815
14:30:00	625	2496	3050	1649	3284	1983	4148	1261	232	666	2494		21888
15:00:00	692	3276	3748	1340	4200	1952	4740	1891	265	738	2645		25487
15:30:00	869	4641	6049	1530	7381	3105	8499	3477	289	1038	4509		41387
PDCO	16274	2535	16045	10120		16453	44564	8910	2485	14484	19848		151718
09:00:00	1	1	1			156	31	1	1	10	32		234
09:30:00	1532	255	245	491		1168	2758	452	305	407	792		8405
10:00:00	1233	280	1291	737		1333	3177	683	267	309	1183		10493
10:30:00	1454	188	1372	692		1432	2827	968	246	1544	1333		12056
11:00:00	1011	102	996	502		1078	2420	523	112	1142	1008		8894
11:30:00	1081	72	1036	641		957	2582	734	137	1314	1268		9822
12:00:00	1377	151	1558	818		1256	3389	45	175	1104	1512		11385
12:30:00	1453	179	1528	926		1514	4181	813	212	1246	2027		14079
13:00:00	1137	186	1155	1147		1280	3416	782	196	1065	1662		12026
13:30:00	795	93	866	856		868	2527	662	117	787	1315		8886
14:00:00	1074	125	1029	706		1059	3677	754	168	1043	1626		11261
14:30:00	857	94	897	563		948	3012	672	116	1075	1032		9266
15:00:00	1010	296	1752	861		1219	3853	698	143	1327	2023		13182
15:30:00	2258	513	2319	1180		2185	6714	1123	290	2111	3035		21728
16:00:00	1												1
(blank)													
(blank)													
Grand Total	30957	43082	60725	40183	50157	42035	103072	36266	6990	27157	55081		495705

4. NBBO data analysis

After preprocessing, the NBBO data looks like:

DATE	TIME_M	EX	SYM_ROOT	SYM_SUFFIX	QU_SEQNUM	BEST_BIDEX	BEST_BID	BEST_ASKEX	BEST_ASK	spread	best_bidsizeshares	best_asksizeshares
06-Jun-13	09:06	Z	MUR		417303	P	62.7	Z	64.87	2.17	100	200
06-Jun-13	09:28	T	MUR		507194	T	62.72	Z	64.87	2.15	300	200
06-Jun-13	09:28	Z	MUR		508304	Z	62.73	T	64.86	2.13	200	300
06-Jun-13	09:30	T	MUR		516810	Z	62.73	T	64.84	2.11	200	100
06-Jun-13	09:30	T	MUR		516822	Z	62.73	T	64.2	1.47	200	200
06-Jun-13	09:30	T	MUR		517083	Z	62.73	T	64.19	1.46	200	300
06-Jun-13	09:30	T	MUR		517189	Z	62.73	T	64.19	1.46	200	400
06-Jun-13	09:30	T	MUR		574962	T	62.8	N	63	0.2	100	800
06-Jun-13	09:30	T	MUR		574963	T	62.8	N	63	0.2	200	800
06-Jun-13	09:30	N	MUR		589938	T	62.8	N	63	0.2	200	900
06-Jun-13	09:30	N	MUR		594255	T	62.8	N	62.81	0.01	200	100
06-Jun-13	09:30	N	MUR		594259	T	62.8	N	62.97	0.17	200	100
06-Jun-13	09:30	N	MUR		594260	T	62.8	N	63	0.2	200	900
06-Jun-13	09:30	N	MUR		594261	T	62.8	N	63	0.2	200	800

NBBO provides the information on the highest bid as well as the lowest ask price, which means the best bid price in the data can be used for this problem.

a) Do the same analysis in question 3. Are there any differences from results in question 3.

Count of BEST_BIDEX				Column Labels			
Row Labels	MUR	PDCO	(blank)	Grand Total	MUR	PDCO	Total
B	350	409		759	0.69%	1.02%	0.84%
C	1146	1		1147	2.26%	0.00%	1.26%
J	6597	350		6947	13.01%	0.87%	7.65%
K	2018	2690		4708	3.98%	6.71%	5.18%
N	12604			12604	24.86%	0.00%	13.88%
P	5981	4594		10575	11.80%	11.45%	11.65%
Q		29408		29408	0.00%	73.30%	32.38%
T	15170			15170	29.93%	0.00%	16.71%
W	2166	640		2806	4.27%	1.60%	3.09%
X	47	33		80	0.09%	0.08%	0.09%
Y	828	71		899	1.63%	0.18%	0.99%
Z	3785	1922		5707	7.47%	4.79%	6.28%
(blank)							
Grand Total	50692	40118		90810	100.00%	100.00%	100.00%

Here the proportion of exchange Q and T are higher compared to the others, which yields different result from problem 3.

Count of SYM_ROOT		Column Labels												
Row Labels	B	C	J	K	N	P	Q	T	W	X	Y	Z	(blank)	Grand Total
MUR		350	1146	6597	2018	12604	5981		15170	2166	47	828	3785	50692
09:00:00							1		1				1	3
09:30:00		29	291	201	600	725	309		690	100		61	161	3167
10:00:00		37	35	416	67	713	309		699	106		100	135	2617
10:30:00		124	132	732	94	679	138		869	213		96	222	3299
11:00:00		5	81	747	134	662	387		1067	164	17	84	324	3672
11:30:00		48	46	433	120	923	440		959	125		67	213	3374
12:00:00		24	79	1018	295	626	157		1302	154	1	69	226	3951
12:30:00		20	160	614	124	621	383		1237	181	1	116	237	3694
13:00:00		21	51	829	87	1136	1022		1159	191		87	325	4908
13:30:00		2	65	363	111	846	623		1382	122		37	408	3959
14:00:00		15	18	260	45	1016	466		1379	69		31	385	3684
14:30:00		1	70	253	170	1106	636		887	140		41	323	3627
15:00:00		16	10	275	76	1579	560		1197	223		9	278	4223
15:30:00		8	108	456	95	1972	550		2342	378	28	30	547	6514
PDCO		409	1	350	2690		4594	29408		640	33	71	1922	40118
09:00:00							162					1	4	167
09:30:00		93		81	178		216	1720		49		25	204	2566
10:00:00		36		27	163		393	2254		100		7	119	3099
10:30:00		28		31	92		415	1841		49	3	5	130	2594
11:00:00		2		21	131		416	1418		12			66	2066
11:30:00		14		4	232		146	1577		59	8		94	2134
12:00:00		6		1	109		360	2381				7	133	2997
12:30:00		33		5	291		436	2519		145			140	3569
13:00:00		14		23	211		222	2214		47		3	287	3021
13:30:00		1		1	102		178	1732		22	6		195	2237
14:00:00		3		2	186		331	2755		23		1	113	3414
14:30:00				66	143		225	2120		100			87	2741
15:00:00		43		28	466		400	2430		22		6	111	3506
15:30:00		136	1	60	386		694	4447		12	16	16	239	6007
(blank)														
(blank)														
Grand Total		759	1147	6947	4708	12604	10575	29408	15170	2806	80	899	5707	90810

- b) In each interval, compute the average of the quoted bid-ask spread and plot it.

Row Labels	Sum of spread	Count of spread	Average of spread2
☐ MUR	1396.23	50692	0.02754
09:00:00	6.45	3	2.15000
09:30:00	177.1	3167	0.05592
10:00:00	93.52	2617	0.03574
10:30:00	118.59	3299	0.03595
11:00:00	113.44	3672	0.03089
11:30:00	104.65	3374	0.03102
12:00:00	117.79	3951	0.02981
12:30:00	108.24	3694	0.02930
13:00:00	140.85	4908	0.02870
13:30:00	91.7	3959	0.02316
14:00:00	73.36	3684	0.01991
14:30:00	71.62	3627	0.01975
15:00:00	69.48	4223	0.01645
15:30:00	109.44	6514	0.01680
☐ PDCO	1419.08	40118	0.03537
09:00:00	746.56	167	4.47042
09:30:00	108.94	2566	0.04246
10:00:00	62.85	3099	0.02028
10:30:00	40.65	2594	0.01567
11:00:00	31.72	2066	0.01535
11:30:00	24.91	2134	0.01167
12:00:00	55.29	2997	0.01845
12:30:00	61.4	3569	0.01720
13:00:00	54.33	3021	0.01798
13:30:00	30.38	2237	0.01358
14:00:00	50.22	3414	0.01471
14:30:00	36.12	2741	0.01318
15:00:00	48.38	3506	0.01380
15:30:00	67.33	6007	0.01121
☐ (blank)			
(blank)			
Grand Total	2815.31	90810	0.03100

Because the average of Spread within the first interval is outlier from the others, which was caused by the the amount of stock of the previous trading day. Therefore, it is not plotted in the figures.

