HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY

FACULTY OF COMPUTER SCIENCE & ENGINEERING

DISCRETE STRUCTURE FOR COMPUTING (CO1007)

Assignment Report – Group 10

VIETNAMESE STOCK MARKET ANALYSIS

Lecturer: Mr. Trần Tuấn Anh, PhD

Student Name	Student ID.	Work Distribution
Nguyễn Đức Bảo Huy	2152089	25%
Nguyễn Vũ Khánh	2153444	25%
Đặng Hoàng Gia	2153312	25%
Hoàng Tiến Đức	2152520	25%

HO CHI MINH CITY, DECEMBER, 2022

Contents

I. Introduction

Vietnam finished 2021 with a 2.58% GDP growth rate, despite witnessing one of the harshest COVID lockdowns in the world during the second half of 2021. Yet, Viet Nam is also one of the rare economies to post two consecutive years of growth since the start of COVID-19 globally. However, 2022 is a difficult time for economic growth. Recently, we can follow the news in the press and see that the financial situation of Vietnam, particularly, and the world in general, has become chaotic and complex. We can see that the economic crisis is approaching, and state banks continuously increase interest rates. This leads to a significant impact on the stock market. These are tough times for all businesses, but it's also an opportunity for those who can. Who will be the millionaire? Is this the game for us?

In this assignment, we conducted several research on stock market indexes during the past 1 year, which was negatively affected by the Covid-19 Pandemic. Applying some Statistics and Discrete Mathematics learning, we implemented several calculations based on the data imported, and used R code to visualize and compute the result more precisely.

II. Problem Solving

1. Collecting Data

Our task is to analyze the stock indexes in group B, which is HNX (Hanoi Stock Exchange) during the past 1 year (from July 2021 to July 2022). Therefore, it is necessary to collect the data of HNX and other related stock prices in the equivalent indexes.

The task has to be conducted by choosing the trading market to clone the stock prices. In this project, we collect the data from investing.com and vietstock.vn so as to provide sufficient and in-depth information with our data storage.

From vietstock.vn, we collect the data of the HNX index, and 341 stock tickers in that index.

THỐNG KÊ GIAO DỊCH

Ngày xuất: 11/23/2022 4:32 PM

-0,49

STT	Ngày	Mã	Tham chiếu	Mở cửa		Cao nhất	Thấp nhất	Trung bình	Thay đổi giá		GD khớp lệnh		GD thỏa thuận		Tổng giao dịch		Vốn hóa thị trường
									+/-	%	KL	GT	KL	GT	KL	GT	
1	04/01/2022	L43	7,1	7,1	7,1	7,2	6,9	7,1	0	0,00	29.300	208	0	0	29.300	208	24.850
2	05/01/2022	L43	7,1	7,1	7,1	7,1	7,0	7,0	0	0,00	32.700	230	0	0	32.700	230	24.850
3	06/01/2022	L43	7,1	7,1	7,2	7,3	6,9	7,1	100	1,41	60.000	425	0	0	60.000	425	25.200
4	07/01/2022	L43	7,2	7,2	7,6	7,8	6,9	7,2	400	5,56	99.100	714	0	0	99.100	714	26.600
5	10/01/2022	L43	7,6	7,8	8,0	8,3	7,6	8,1	400	5,26	183.800	1.489	0	0	183.800	1.489	28.000
6	11/01/2022	L43	8,0	8,2	8,1	8,3	7,9	8,1	100	1,25	63.500	511	0	0	63.500	511	28.350
7	12/01/2022	L43	8,1	7,3	7,4	7,8	7,3	7,5	-700	-8,64	74.100	556	0	0	74.100	556	25.900
8	13/01/2022	L43	7,4	7,3	7,1	7,5	7,0	7,2	-300	-4,05	67.300	486	0	0	67.300	486	24.850
9	14/01/2022	L43	7,1	6,8	7,2	7,3	6,5	6,9	100	1,41	18.400	127	0	0	18.400	127	25.200
10	17/01/2022	L43	7,2	7,3	7,0	7,3	6,8	7,0	-200	-2,78	29.700	209	0	0	29.700	209	24.500
11	18/01/2022	L43	7,0	7,0	6,3	7,0	6,3	6,5	-700	-10,00	8.200	53	0	0	8.200	53	22.050
12	19/01/2022	L43	6,3	5,8	6,9	6,9	5,8	6,8	600	9,52	12.400	84	0	0	12.400	84	24.150
13	20/01/2022	L43	6,9	6,9	7,4	7,5	6,9	7,3	500	7,25	29.500	216	0	0	29.500	216	25.900
14	21/01/2022	L43	7,4	7,3	7,4	7,4	7,3	7,4	0	0,00	20.600	152	0	0	20.600	152	25.900
15	24/01/2022	L43	7,4	7,0	6,9	7,3	6,9	7,1	-500	-6,76	3.800	27	0	0	3.800	27	24.150
16	25/01/2022	L43	6,9	7,3	7,0	7,3	7,0	7,3	100	1,45	47.500	346	0	0	47.500	346	24.500
17	26/01/2022	L43	7,0	6,4	7,3	7,4	6,4	7,2	300	4,29	800	6	0	0	800	6	25.550
18	27/01/2022	L43	7,3	6,7	6,8	6,9	6,7	6,8	-500	-6,85	4.400	30	0	0	4.400	30	23.800
19	28/01/2022	L43	6,8	6,8	7,0	7,0	6,5	6,6	200	2,94	35.200	231	0	0	35.200	231	24.500
20	07/02/2022	L43	7,0	6,9	7,4	7,4	6,9	7,1	400	5,71	1.400	10	0	0	1.400	10	25.900
21	08/02/2022	L43	7,4	7,0	7,4	7,4	7,0	7,0	0	0,00	12.600	90	0	0	12.600	90	25.900
22	09/02/2022	L43	7,4	7,5	7,7	7,7	7,2	7,5	300	4,05	39.300	296	0	0	39.300	296	26.950
23	10/02/2022	L43	7,7	7,6	7,7	7,7	7,5	7,7	0	0,00	99.300	761	0	0	99.300	761	26.950
24	11/02/2022	L43	7,7	7,7	7,7	7,8	7,6	7,7	0	0,00	35.200	272	0	0	35.200	272	26.950
25	14/02/2022	L43	7,7	7,6	7,7	7,7	7,6	7,6	0	0,00	4.000	30	0	0	4.000	30	26.950
26	15/02/2022	L43	7,7	7,7	7,9	7,9	7,7	7,7	200	2,60	3.300	25	0	0	3.300	25	27.650
27	16/02/2022	L43	7,9	7,5	8,2	8,3	7,5	8,0	300	3,80	96.500	771	0	0	96.500	771	28.700
28	17/02/2022	L43	8,2	7,8	8,2	8,2	7,8	8,2	0	0,00	29.000	237	0	0	29.000	237	28.700
29	18/02/2022	L43	8,2	7,5	7,5	8,0	7,5	7,8	-700	-8,54	1.300	10	0	0	1.300	10	26.250
30	21/02/2022	L43	7,5	7,4	7,6	8,0	7,4	7,5	100	1,33	11.700	88	0	0	11.700	88	26.600
31	22/02/2022	L43	7,6	7,4	7,6	7,6	7,0	7,2	0	0,00	6.100	44	0	0	6.100	44	26.600
32	23/02/2022	L43	7,6	7,4	7,5	7,6	7,4	7,6	-100	-1,32	4.400	33	0	0	4.400	33	26.250
33	24/02/2022	L43	7,5	7,5	7,6	7,9	7,5	7,6	100	1,33	20.500	156	0	0	20.500	156	26.600
34	25/02/2022	L43	7,6	7,3	7,5	7,5	7,3	7,4	-100	-1,32	400	3	0	0	400	3	26.250
35	28/02/2022	L43	7,5	7,5	8,2	8,2	7,5	8,2	700	9,33	106.700	872	0	0	106.700	872	28.700
36	01/03/2022	L43	8,2	8,5	8,2	8,5	8,1	8,2	0	0,00	11.400	94	0	0	11.400	94	28.700
37	02/03/2022	L43	8,2	7,6	7,4	7,6	7,4	7,6	-800	-9,76	2.300	17	0	0	2.300	17	25.900
38	03/03/2022	L43	7,4	7,1	7,5	7,5	6,7	6,7	100	1,35	109.800	738	0	0	109.800	738	26.250
39	04/03/2022	L43	7,5	7,5	7,5	7,5	7,5	7,5	0	0,00	2.300	17	0	0	2.300	17	26.250
40	07/03/2022	L43	7,5	7,2	7,7	7,7	7,0	7,2	200	2,67	13.400	96	0	0	13.400	96	26.950
41	08/03/2022	L43	7,7	7,6	7,6	8,4	7,6	8,3	-100	-1,30	109.000	901	0	0	109.000	901	26.600
42	09/03/2022	L43	7,6	7,7	8,3	8,3	7,4	8,3	700	9,21	17.900	148	0	0	17.900	148	29.050
43	10/03/2022	L43	8,3	8,2	7,7	8,2	7,6	7,7	-600	-7,23	10.300	79	0	0	10.300	79	26.950
44	11/03/2022	L43	7,7	8,1	8,0	8,1	7,6	7,7	300	3,90	500	4	0	0	500	4	28.000
45	14/03/2022	L43	8,0	7,5	7,6	7,6	7,5	7,6	-400	-5,00	7.900	60	0	0	7.900	60	26.600
46	15/03/2022	L43	7,6	7,7	7,3	7,9	7,1	7,3	-300	-3,95	4.300	31	0	0	4.300	31	25.550
47	16/03/2022	L43	7,3	7,3	7,3	8,0	7,3	8,0	0	0,00	43.000	341	0	0	43.000	341	25.550
48	17/03/2022	L43	7,3	7,8	8,0	8,0	7,8	8,0	700	9,59	79.100	633	0	0	79.100	633	28.000
49	18/03/2022	L43	8,0	8,3	7,6	8,3	7,6	7,9	-400	-5,00	2.800	22	0	0	2.800	22	26.600

49	18/03/2022	L43	8,0	8,3	7,6	8,3	7,6	7,9	-400	-5,00	2.800	22	0	0	2.800	22	26.600
50	21/03/2022	L43	7,6	7,6	7,5	7,6	7,4	7,5	-100	-1,32	5.447	41	0	0	5.447	41	26.250
51	22/03/2022	L43	7,5	7,4	7,5	7,5	7,4	7,4	0	0,00	2.900	21	0	0	2.900	21	26.250
52	23/03/2022	L43	7,5	7,5	7,7	7,7	7,4	7,4	200	2,67	7.800	58	0	0	7.800	58	26.950
53	24/03/2022	L43	7,7	7,7	7,8	7,8	7,5	7,6	100	1,30	1.100	8	0	0	1.100	8	27.300
54	25/03/2022	L43	7,8	7,8	7,8	7,8	7,8	7,8	0	0,00	0	0	0	0	0	0	27.300
55	28/03/2022	L43	7,8	7,5	7,2	7,7	7,2	7,3	-600	-7,69	15.400	112	0	0	15.400	112	25.200
56	29/03/2022	L43	7,2	7,4	7,9	7,9	7,4	7,8	700	9,72	23.100	181	0	0	23.100	181	27.650
57	30/03/2022	L43	7,9	7,9	8,0	8,0	7,9	8,0	100	1,27	5.400	43	0	0	5.400	43	28.000
58	31/03/2022	L43	8,0	7,4	7,4	7,4	7,4	7,4	-600	-7,50	100	1	0	0	100	1	25.900
59	01/04/2022	L43	7,4	7,4	7,4	7,4	7,4	7,4	0	0,00	0	0	0	0	0	0	25.900
60	04/04/2022	L43	7,4	7,4	7,5	7,7	7,4	7,5	100	1,35	6.200	47	0	0	6.200	47	26.250
61	05/04/2022	L43	7,5	7,5	7,5	7,5	7,5	7,5	0	0,00	0	0	0	0	0	0	26.250
62	06/04/2022	L43	7,5	7,3	7,4	7,4	7,2	7,3	-100	-1,33	2.200	16	0	0	2.200	16	25.900
63	07/04/2022	L43	7,4	7,2	8,1	8,1	6,9	7,3	700	9,46	144.700	1.058	0	0	144.700	1.058	28.350
64	08/04/2022	L43	8,1	7,8	7,8	7,8	7,8	7,8	-300	-3,70	100	1	0	0	100	1	27.300
65	12/04/2022	L43	7,8	7,3	7,3	7,3	7,3	7,3	-500	-6,41	800	6	0	0	800	6	25.550
66	13/04/2022	L43	7,3	6,9	6,9	7,0	6,6	6,6	-400	-5,48	10.400	69	0	0	10.400	69	24.150
67	14/04/2022	L43	6,9	6,8	6,8	6,8	6,8	6,8	-100	-1,45	230	2	0	0	230	2	23.800
68	15/04/2022	L43	6,8	7,2	7,4	7,4	7,2	7,4	600	8,82	13.600	101	0	0	13.600	101	25.900
69	18/04/2022	L43	7,4	7,3	7,3	7,3	6,9	7,3	-100	-1,35	5.317	39	0	0	5.317	39	25.550
70	19/04/2022	L43	7,3	7,2	6,8	7,2	6,7	6,9	-500	-6,85	1.800	12	0	0	1.800	12	23.800
71	20/04/2022	L43	6,8	6,8	6,8	6,8	6,8	6,8	0	0,00	500	3	0	0	500	3	23.800
72	21/04/2022	L43	6,8	6,8	6,3	6,8	6,2	6,5	-500	-7,35	400	3	0	0	400	3	22.050
73	22/04/2022	L43	6,3	6,3	5,7	6,3	5,7	6,1	-600	-9,52	2.500	15	0	0	2.500	15	19.950
74	25/04/2022	L43	5,7	5,2	5,2	5,2	5,2	5,2	-500	-8,77	9.400	49	0	0	9.400	49	18.200
75	26/04/2022	L43	5,2	5,2	5,2	5,2	5,2	5,2	0	0,00	300	2	0	0	300	2	18.200
76	27/04/2022	L43	5,2	5,0	5,5	5,5	5,0	5,0	300	5,77	4.600	23	0	0	4.600	23	19.250
77	28/04/2022	L43	5,5	5,5	5,7	5,8	5,5	5,7	200	3,64	1.900	11	0	0	1.900	11	19.950
78	29/04/2022	L43	5,7	5,3	5,5	5,5	5,3	5,3	-200	-3,51	200	1	0	0	200	1	19.250
79	04/05/2022	L43	5,5	5,5	5,9	5,9	5,5	5,5	400	7,27	1.000	6	0	0	1.000	6	20.650
80	05/05/2022	L43	5,9	6,2	6,0	6,2	6,0	6,2	100	1,69	1.700	11	0	0	1.700	11	21.000
81	06/05/2022	L43	6,0	6,0	6,0	6,0	6,0	6,0	0	0,00	0	0	0	0	0	0	21.000
82	09/05/2022	L43	6,0	5,5	5,5	5,5	5,5	5,5	-500	-8,33	100	1	0	0	100	1	19.250
83	10/05/2022	L43	5,5	5,1	5,1	5,1	5,1	5,1	-400	-7,27	100	1	0	0	100	1	17.850
84	11/05/2022	L43	5,1	5,1	5,1	5,1	5,1	5,1	0	0,00	0	0	0	0	0	0	17.850
85	12/05/2022	L43	5,1	4,7	4,7	4,7	4,7	4,7	-400	-7,84	200	1	0	0	200	1	16.450
86	13/05/2022	L43	4,7	4,7	4,7	4,7	4,7	4,7	0	0,00	0	0	0	0	0	0	16.450
87	16/05/2022	L43	4,7	4,7	4,7	4,7	4,7	4,7	0	0,00	0	0	0	0	0	0	16.450
88	17/05/2022	L43	4,7	4,7	4,7	4,7	4,7	4,7	0	0,00	0	0	0	0	0	0	16.450
89	18/05/2022	L43	4,7	5,0	5,1	5,1	5,0	5,1	400	8,51	400	2	0	0	400	2	17.850
90	19/05/2022	L43	5,1	5,5	5,5	5,5	5,5	5,5	400	7,84	100	1	0	0	100	1	19.250
91	20/05/2022	L43	5,5	5,5	5,5	5,5	5,5	5,5	0	0,00	0	0	0	0	0	0	19.250
92	23/05/2022	L43	5,5	5,5	5,5	5,5	5,5	5,5	0	0,00	100	1	0	0	100	1	19.250
93	24/05/2022	L43	5,5	5,0	5,0	5,0	5,0	5,0	-500	-9,09	100	1	0	0	100	1	17.500
94	25/05/2022	L43	5,0	5,0	4,6	5,0	4,5	4,7	-400	-8,00	400	2	0	0	400	2	16.100
95	26/05/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	69.427	319	0	0	69.427	319	16.100
96	27/05/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	0	0	0	0	0	0	16.100
97	30/05/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	0	0	0	0	0	0	16.100
98	31/05/2022	L43	4,6	4,6	4,5	4,6	4,5	4,6	-100	-2,17	400	2	0	0	400	2	15.750
99	01/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
100	02/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
101	03/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	100	0	0	0	100	0	15.750
102	06/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
103	07/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
104	08/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
105	09/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
106	10/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	100	0	0	0	100	0	15.750

91	20/05/2022	L43	5,5	5,5	5,5	5,5	5,5	5,5	0	0,00	0	0	0	0	0	0	19.250
92	23/05/2022	L43	5,5	5,5	5,5	5,5	5,5	5,5	0	0,00	100	1	0	0	100	1	19.250
93	24/05/2022	L43	5,5	5,0	5,0	5,0	5,0	5,0	-500	-9,09	100	1	0	0	100	1	17.500
94	25/05/2022	L43	5,0	5,0	4,6	5,0	4,5	4,7	-400	-8,00	400	2	0	0	400	2	16.100
95	26/05/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	69.427	319	0	0	69.427	319	16.100
96	27/05/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	0	0	0	0	0	0	16.100
97	30/05/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	0	0	0	0	0	0	16.100
98	31/05/2022	L43	4,6	4,6	4,5	4,6	4,5	4,6	-100	-2,17	400	2	0	0	400	2	15.750
99	01/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
100	02/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
101	03/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	100	0	0	0	100	0	15.750
102	06/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
103	07/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
104	08/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
105	09/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
106	10/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	100	0	0	0	100	0	15.750
107	13/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	0	0	0	0	0	0	15.750
108	14/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	700	3	0	0	700	3	15.750
109	15/06/2022	L43	4,5	4,5	4,5	4,5	4,5	4,5	0	0,00	1	0	0	0	1	0	15.750
110	16/06/2022	L43	4,5	4,5	4,9	4,9	4,5	4,7	400	8,89	6.300	30	0	0	6.300	30	17.150
111	17/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
112	20/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
113	21/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
114	22/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
115	23/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
116	24/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
117	27/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	1.000	5	0	0	1.000	5	17.150
118	28/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
119	29/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
120	30/06/2022	L43	4,9	4,9	4,9	4,9	4,9	4,9	0	0,00	0	0	0	0	0	0	17.150
121	01/07/2022	L43	4,9	4,5	4,6	4,6	4,5	4,5	-300	-6,12	400	2	0	0	400	2	16.100
122	04/07/2022	L43	4,6	4,6	4,6	4,6	4,6	4,6	0	0,00	0	0	0	0	0	0	16.100
123	05/07/2022	L43	4,6	4,3	4,3	4,3	4,3	4,3	-300	-6,52	1.000	4	0	0	1.000	4	15.050
124	06/07/2022	L43	4,3	4,3	4,3	4,3	4,3	4,3	0	0,00	400	2	0	0	400	2	15.050
125	07/07/2022	L43	4,3	4,3	4,3	4,3	4,3	4,3	0	0,00	0	0	0	0	0	0	15.050
126	08/07/2022	L43	4,3	4,3	4,3	4,3	4,3	4,3	0	0,00	0	0	0	0	0	0	15.050
127	11/07/2022	L43	4,3	4,3	4,3	4,3	4,3	4,3	0	0,00	0	0	0	0	0	0	15.050
128	12/07/2022	L43	4,3	4,3	3,9	4,3	3,9	4,0	-400	-9,30	800	3	0	0	800	3	13.650
129	13/07/2022	L43	3,9	3,9	3,9	3,9	3,9	3,9	0	0,00	100	0	0	0	100	0	13.650
130	14/07/2022	L43	3,9	3,9	3,9	3,9	3,9	3,9	0	0,00	0	0	0	0	0	0	13.650
131	15/07/2022	L43	3,9	4,0	4,0	4,0	4,0	4,0	100	2,56	200	1	0	0	200	1	14.000
132	18/07/2022	L43	4,0	4,0	4,0	4,0	4,0	4,0	0	0,00	0	0	0	0	0	0	14.000
133	19/07/2022	L43	4,0	4,0	4,0	4,0	4,0	4,0	0	0,00	0	0	0	0	0	0	14.000
134	20/07/2022	L43	4,0	4,0	4,0	4,0	4,0	4,0	0	0,00	0	0	0	0	0	0	14.000
135	21/07/2022	L43	4,0	4,0	4,0	4,0	4,0	4,0	0	0,00	0	0	0	0	0	0	14.000
136	22/07/2022	L43	4,0	4,0	3,6	4,0	3,6	3,7	-400	-10,00	7.200	26	0	0	7.200	26	12.600
137	25/07/2022	L43	3,6	3,6	3,6	3,6	3,6	3,6	0	0,00	0	0	0	0	0	0	12.600
138	26/07/2022	L43	3,6	3,6	3,6	3,6	3,6	3,6	0	0,00	0	0	0	0	0	0	12.600
139	27/07/2022	L43	3,6	3,6	3,6	3,6	3,6	3,6	0	0,00	0	0	0	0	0	0	12.600
140	28/07/2022	L43	3,6	3,7	3,9	3,9	3,7	3,9	300	8,33	6.500	25	0	0	6.500	25	13.650
141	29/07/2022	L43	3,9	3,6	3,6	3,6	3,6	3,6	-300	-7,69	100	0	0	0	100	0	12.600

Figure 1. *The data of L43 is collected from vietstock.vn*

The file downloaded is in .xlsx form, which means it could be opened via excel or imported by reading the excel file from Rstudio.

2. Analyzing movement

2.1. Data Selection

The purpose is to analyze the HNX index distribution over the period of 1 year and build a predictive model, which is a linear regression model in the next section, we need to have our factors to be unrelated to each other. Therefore, in our analysis, we will analyze **closing prices** according to the following variables: the opening prices, the highest price, the lowest price, and the volume of the stock index.

In detail, these dependent variables can be defined:

- Opening price: The price from the first transaction of a business day.
- Highest price: The highest closing price of a stock over the past 52 weeks, adjusted for any stock splits, or the highest intraday price of a stock in the most recent (or current) trading session. (according to NASDAQ)..
- Lowest price: the lowest price at which the securities were traded on the exchange on which the securities are listed.
- Volume: the number of shares traded in a particular stock, index, or other investment over a specific period of time.

2.2. Data Import

First, we have to import the data using R. We choose the data of HNX_Index.xls which was already collected via the previous step. We only select the specific column in the data set, according to those aforementioned dependent and independent variables.

```
#========== IMPORTING DATA ===========#
library(readxl)
my_data <- read_excel("C:/Users/ASUS/Downloads/HNX_Index.xls",
                      range = "D10:S150", col_names = FALSE,
                      col_types = c("skip",
                                    "skip", "numeric", "numeric",
                                    "numeric", "numeric", "skip",
                                    "skip", "skip", "skip",
                                    "skip", "skip", "skip",
                                    "numeric", "skip", "skip"))

names(my_data) <- c("opening_price", "closing_price",
                    "highest_price", "lowest_price", "volume")
View(my_data)
```

Figure 2. *Import and view the data.*

We should check out the data and also make sure it is imported properly.

	opening_price	closing_price	highest_price	lowest_price	volume
1	289.84	288.61	291.82	288.37	63666672
2	284.52	289.84	291.97	284.52	89490405
3	282.88	284.52	284.57	280.72	47902805
4	285.38	282.88	286.54	282.75	55168196
5	288.83	285.38	288.89	284.00	63377611
6	288.09	288.83	290.53	288.09	63746139
7	288.87	288.09	289.61	287.44	64573315
8	284.43	288.87	289.86	284.43	88156025
9	284.63	284.43	286.84	280.10	63390095
10	284.40	284.63	287.32	283.98	59536484
11	284.75	284.40	287.74	284.40	76025124
12	281.36	284.75	285.03	280.03	76273630
13	281.99	281.36	284.24	280.50	65724611

Figure 3. *View the table of data.*

2.3. Data Preprocessing

2.3.1. Data Cleaning

a. Eliminating NA values

Since the data set is not large, we do not have to remove the Not Available (NA) data. However, it is necessary to replace all NA values with the means of all the values in the equivalent column.

```
# install.packages("tidyverse") # make sure to install
library(tidyverse) #library helps processing data

my_data[my_data == 0] <- NA # replace all 0 to NA
my_data <- my_data %>%
  mutate_if(is.numeric, ~replace_na(.,mean(., na.rm = TRUE)))
```

Figure 4. *Data Cleaning – Remove NA values.*

b. Excluding outliers

Similar to NA values, they will lead to some fallacies; hence, outliers need to be removed. In terms of methods that help eliminate outliers, we will apply *interquartile range calculation* [1].

```
# using interquartile range calculation to remove outliers

dim(my_data)
# output: [1] 141     5

quartiles <- quantile(my_data$closing_price, probs=c(.25, .75), na.rm = TRUE)
IQR <- IQR(my_data$closing_price)

Lower <- quartiles[1] - 1.5*IQR
Upper <- quartiles[2] + 1.5*IQR

data_no_outlier <- subset(my_data,
                          my_data$closing_price > Lower &
                            my_data$closing_price < Upper)

dim(data_no_outlier)
# output: [1] 141     5
```

Figure 5. *Exclude outliers.*

We can see from the output, reasoning that outliers do not exist (the output gives the same results).

It is clearly depicted that, the non-processed data does not have outliers.

2.3.2. Data scaling (Feature scaling)

On examining, we can clearly see those data variables are not on the same scale. As a result, we will apply *feature scaling* [2] in order to boost the accuracy of the regression model. The following image shows how we scale the data according to the formula:

$$z = \frac{x - u}{\sigma}$$

```
# feature scaling
my_data <- as.data.frame(sapply(my_data,
                                function(x) (x-mean(x))/sd(x)))
```

Figure 6. *Feature scaling.*

2.3.3. Data visualization

a. Data Distribution

Using plot() function to examine the distribution of runtime. Below are the codes:

```
d <- density(my_data$closing_price)
plot(d, main="Runtime distribution")
polygon(d, col = "red", border="blue")

# install.packages("moments")
library(moments)
skewness(my_data$closing_price)
# [1] -0.04331536
```

Figure 6. *Use plot( ) to illustrate the distribution of runtime.*

After running these lines of codes, we have the following images:

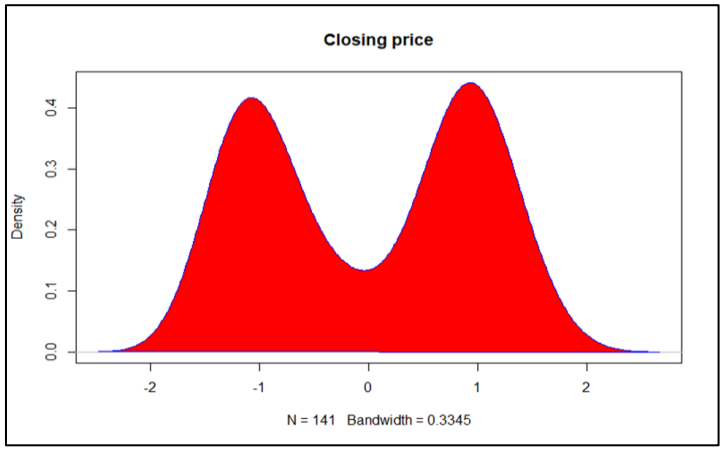

Figure 7. *The distribution of* ***closing_price****.*

The closing price apparently do **not** follow any distribution.

To further check how normal the distribution is, we will go through the qqnorm and qqline functions.

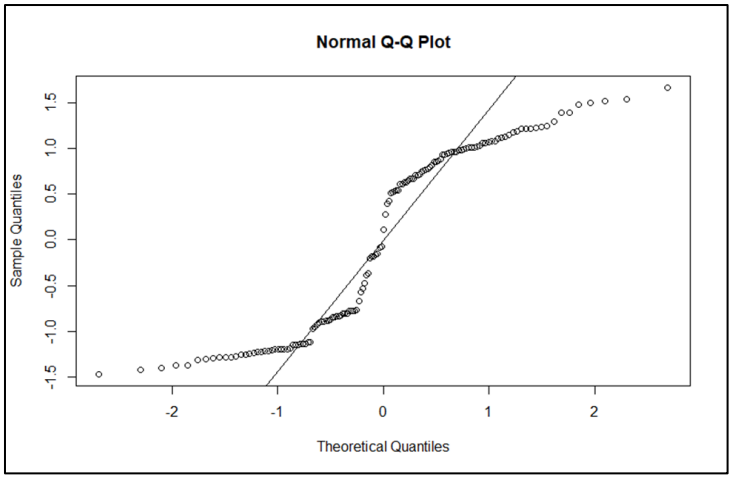

Figure 8. *Qqplot of* ***closing_price.***

This graph ensures the **abnormality** of the closing price.

b. Correlation Plot

We will use corrplot to see how each of the variables depends on the other with the following R implementation.

```
# install.packages("corrplot")
library(corrplot)
# correlation plot
M = cor(my_data)
corrplot(M, method = 'color')
```

Figure 9. *Install package to plot the correlation.*

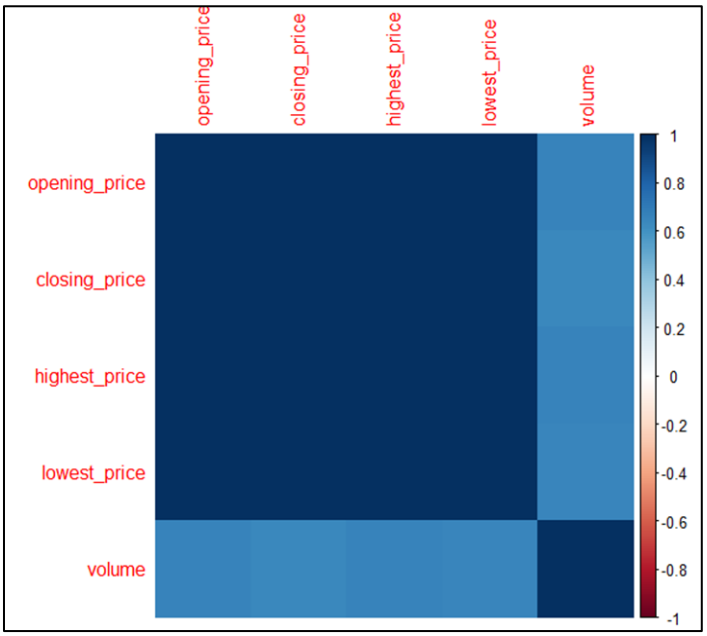

Figure 10. *Correlation plot.*

As we can see, there are strong connections between every two variables.

2.4. Build Predictive Model

2.4.1. Formula

In this part, we will use multiple linear regression models to examine the runtime.

- Dependent variable: closing_price
- Independent variable: opening_price, highest_price, lowest_price, volume

The model is performed by the following formula:

$$\text{closing price} = \alpha + \beta_1 * \text{opening\_price} + \beta_2 * \text{highest\_price} + \beta_3 * \text{lowest price} + \beta_4 * \text{volume} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$

Implement the following R code:

```
# linear regression model
linear_model <- lm(formula = closing_price ~ opening_price +
                     highest_price + lowest_price +
                     volume, data = my_data)
summary(linear_model)
```

Figure 11. *Fitting the linear regression model.*

After we call the summary of the model, we have the following results:

```
Call:
lm(formula = closing_price ~ opening_price + highest_price +
    lowest_price + volume, data = my_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.137198 -0.024290 -0.005937  0.019881  0.168652

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.840e-16  3.656e-03   0.000    1.000
opening_price -7.519e-01  9.311e-02  -8.075 3.20e-13 ***
highest_price  8.749e-01  1.010e-01   8.665 1.18e-14 ***
lowest_price   8.759e-01  5.402e-02  16.213  < 2e-16 ***
volume        -6.174e-04  5.086e-03  -0.121    0.904
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04342 on 136 degrees of freedom
Multiple R-squared:  0.9982,    Adjusted R-squared:  0.9981
F-statistic: 1.853e+04 on 4 and 136 DF,  p-value: < 2.2e-16
```

Figure 12. *The output of* ***summary(linear_model)***.

From the above results, the specific formula is:

closing_price= 3.840e–12 – 0.752* opening_price + 0.875* highest_price + 0.876* lowest_price – 0.00062* volume

2.4.2. Result Explanation

We will explain some significant points in the results:

The above equation follows the explanation that if **closing price** increases by 1, **opening price** will decrease by 0.752 (supposed that other values are fixed). This can be reasoned that if the opening **price** increases by 1, **the closing price** decreases by 1.33. Consequently, variables with *large coefficients* will have a more significant impact on dependent values.

In terms of Residuals, the average of the residuals is theoretically **zero**. The median in the results is -0.006 (not far from 0). Additionally, the 25% (1Q) and 75% (3Q) quantiles are also fairly well balanced between the median, which shows that the residual of this equation is *relatively balanced*.

The last three columns are standard error, t tests and p-values. Let's consider the hypothesis test for regression coefficient:

- H_0: Regression coefficient is not statistically significant ($\beta_i = 0$)
- H_1: Regression coefficient is statistically significant ($\beta_i \neq 0$)
- As the p-values of all factors < 0.05 → reject H_0 → All regression coefficients of these variables are *statistically significant* except **Volume**.
- This indicates that Volume has no impact on closing price.

Another thing that should be mentioned is the R-squared. This value indicates that the more it is near to 1, the tighter the connection between closing price and other variables. As the results have shown, the value is approximately 0.9982 (= 99.82%), which is an extremely high percentage.

2.4.3. Validity

Next, to test the **validity** of the equation, we will go through the following assumptions:

1. ***All independent variables have fixed value (fixed means these variables are supposed not to have any errors in measurement).***

This is true because our independent variables are either ordinal or binary, which indicates that those are fixed parameters and not deduced from any pre-calculation.

2. ***Error ε is normally distributed.***
3. ***ε has its mean equal to 0.***
4. ***ε has fixed variance.***
5. ***The error terms εi are mutually independent (which means ε_i and ε_j has no relation)***

To testify the last 3 assumptions, I will plot 4 different related graphs by calling plot (linear_model) in the R program.

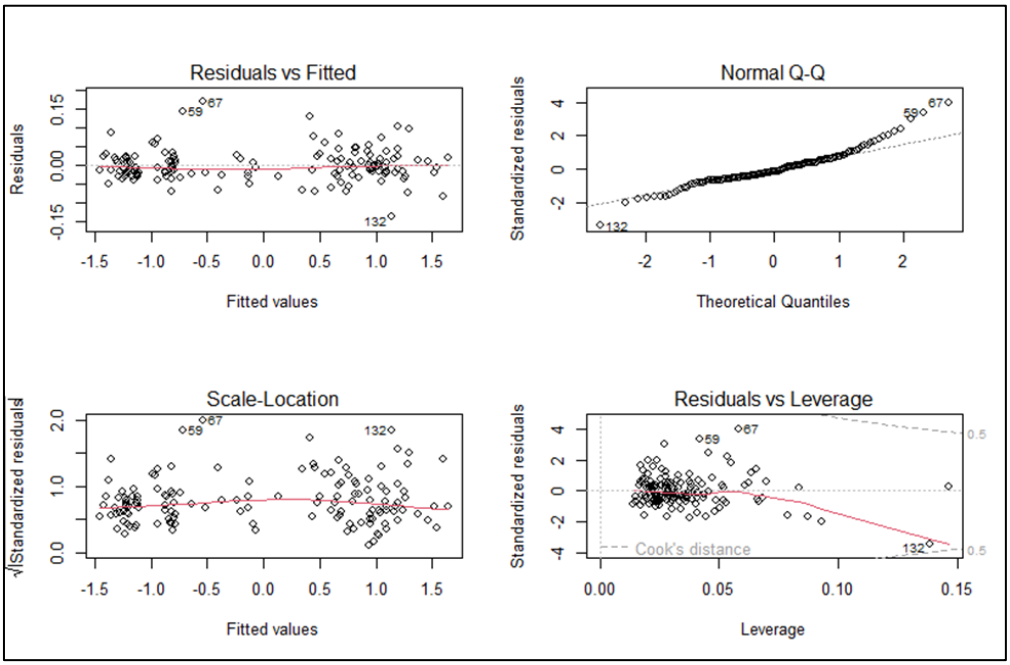

Figure 13. *4 related graphs of* ***linear_model.***

- The top left graph shows that residuals are clustered around the y = 0, so error ε has a mean value around 0 ((3) is acceptable).

- The graph at top right corner plots the residual and expected values based on normal distribution. We can see that the residuals are concentrated relatively close to the values on the standard curve, and therefore, assumption (2) can also be met.

- About the bottom left one, it is illustrated that there is not much difference between square root of standardized residuals and fitted values. This means (4) might be satisfied

In general, after analyzing the model, we can conclude that the model has demonstrated the connection between closing price and independent variables fairly well.

2.4.4. Prediction

Before we predict the stock index, let's have a look on how the fitted values plotted against the actual values.

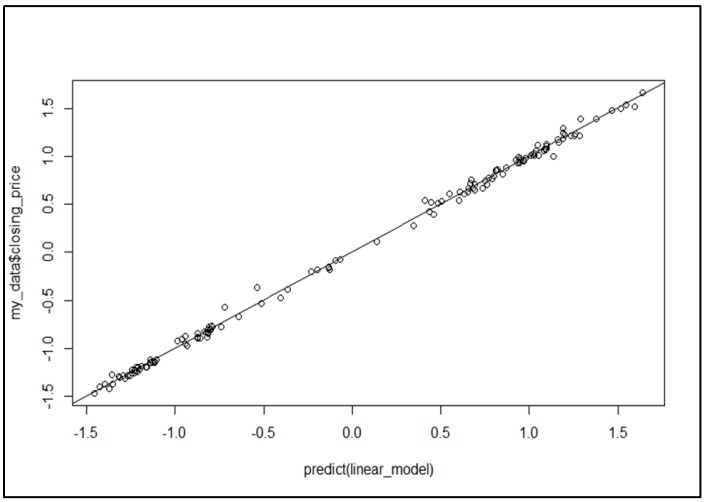

Figure 14. *Predict graph.*

It can be seen that it is relatively fit, which partly explains that the model is good.

Now let's predict the stock in the next 3 months to see if the model is actually good or not.

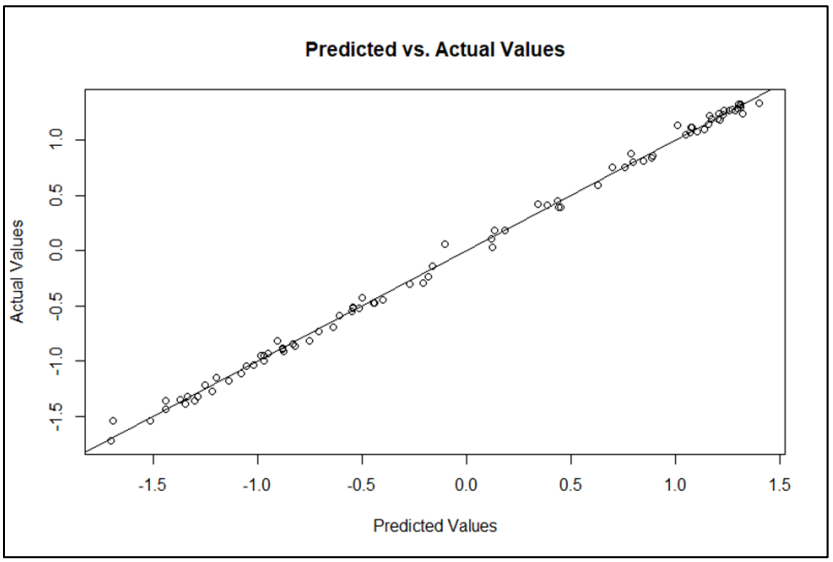

Figure 15. *Predict graph of the stock in the next 3 months.*

We can see that there is not much difference between the predicted values and the actual values in the next 3 months of the HNX index.

3. Model Analysis, Influencing Factors and Cross-model Usability

3.1. Three stocks in three different sectors (HNX-INDEX): tourism, finance, F&B

3.1.1. Sources

We have chosen 3 stocks corresponding to 3 sectors:

- Tourism: OCH (One Capital Hospitality Joint Stock Company)

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=632

- Financial: EVS (Everest Securities Joint Stock Company)

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=13290

- F&B: SLS (Son La Sugar JSC)

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=921

3.1.2. Data Visualization

- Tourism (OCH):

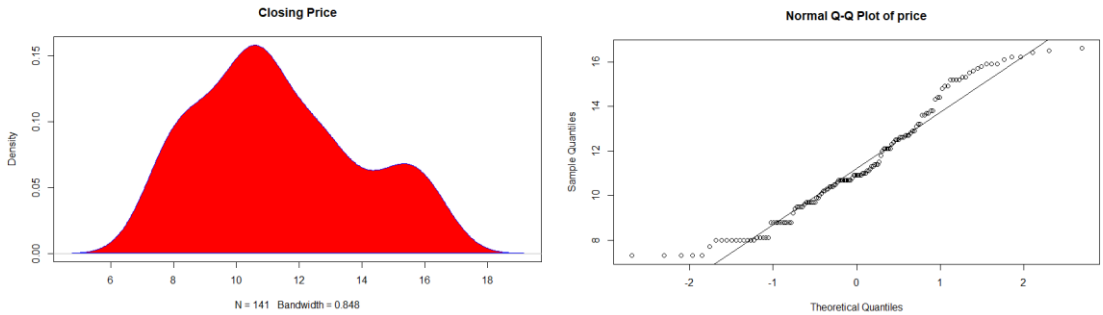

Figure 16. *The distribution of OCH.*

- Finance (EVS):

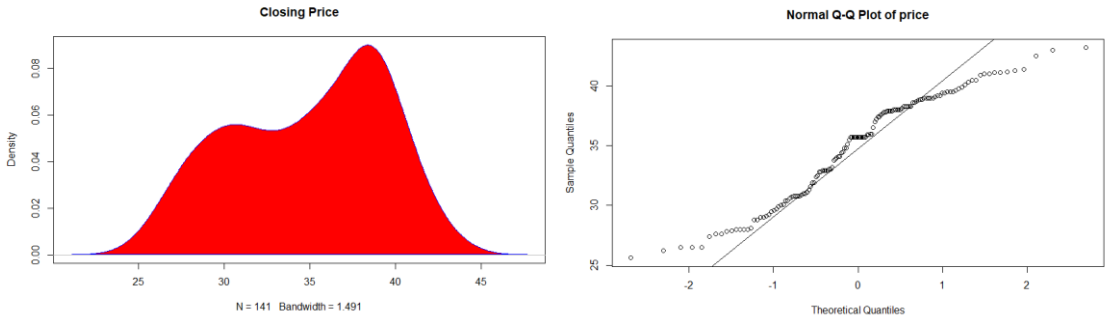

Figure 17. *The distribution of EVS.*

- F&B (SLS):

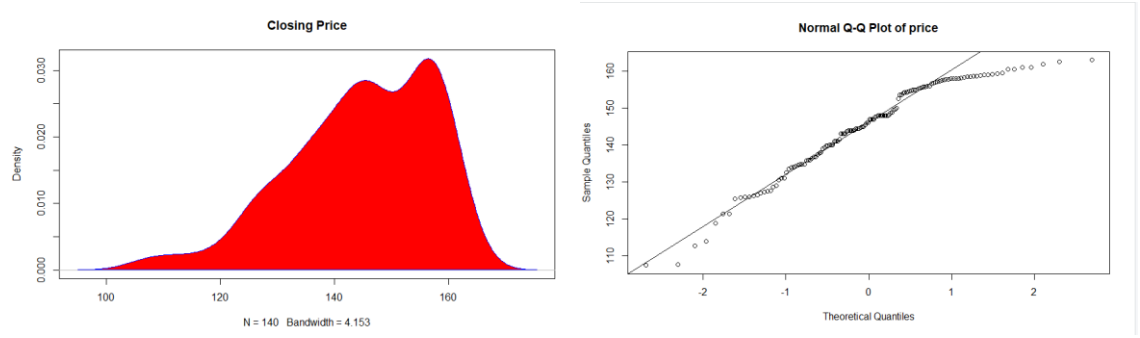

Figure 18. *The distribution of SLS.*

3.1.3. Correlation Plot

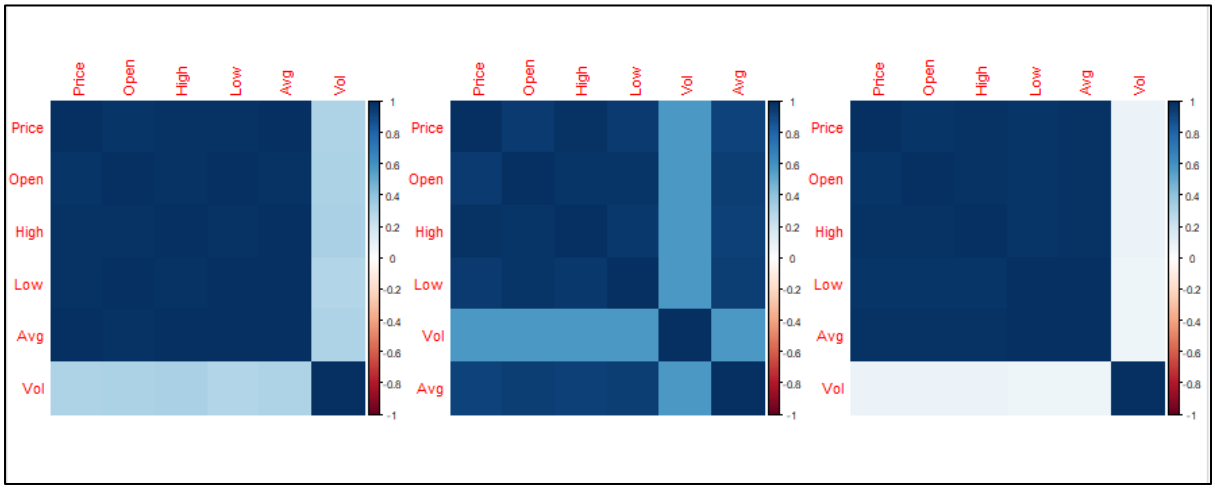

Figure 19. *Correlation plots of 3 stocks OCH, EVS, SLS respectively.*

3.1.4. Regression Model

Overall formula:

closing_price = $\alpha + \beta_1$* opening_price + β_2* highest_price + β_3* lowest_price + β_4* volume + β_5* avg + $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$

- Toursim (OCH):

```
Call:
lm(formula = Price ~ Low + High + Vol + Open + Avg, data = scaling_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.22447 -0.05371  0.01043  0.02813  0.31820

Coefficients:
                      Estimate             Std. Error t value      Pr(>|t|)
(Intercept) -0.0000000000000002736  0.0073759775531798541   0.000       1.00000
Low          0.2498362116826042256  0.1008082251190073886   2.478       0.01443 *
High         0.2799412754408735493  0.1016711364240307919   2.753       0.00671 **
Vol          0.0019547588642826476  0.0079744557796413672   0.245       0.80673
Open        -0.4301085922658356986  0.0713131663835917945  -6.031 0.00000001470 ***
Avg          0.8923793511882762708  0.1364112872306430801   6.542 0.00000000116 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08758 on 135 degrees of freedom
Multiple R-squared:  0.9926,    Adjusted R-squared:  0.9923
F-statistic:  3623 on 5 and 135 DF,  p-value: < 0.00000000000000022
```

Figure 20. *Linear regression model of OCH.*

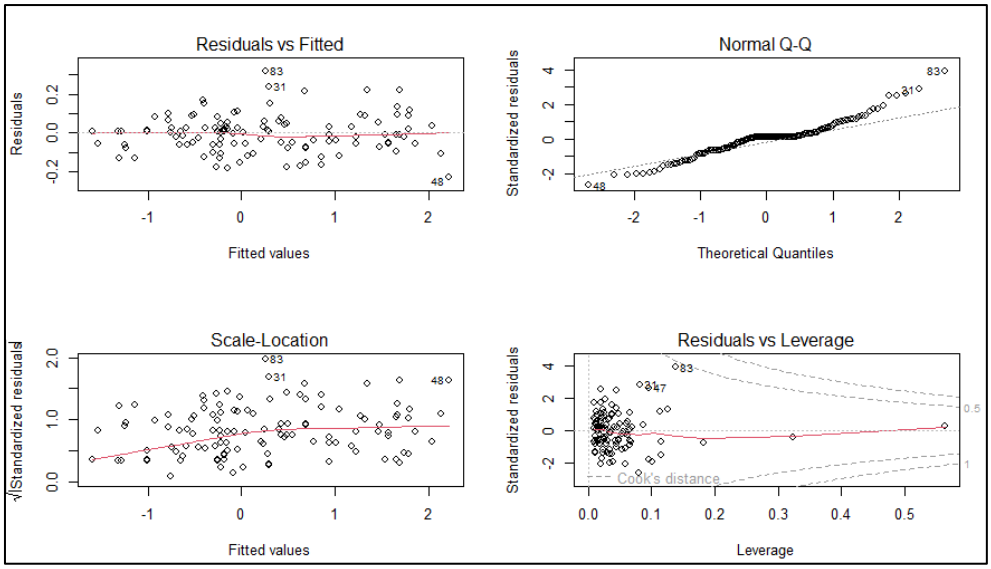

Figure 21. *4 related graphs.*

- Finance (EVS):

```
Call:
lm(formula = Price ~ Low + High + Vol + Open + Avg, data = scaling_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.85552 -0.07587  0.05571  0.10792  0.22588

Coefficients:
                        Estimate             Std. Error t value             Pr(>|t|)
(Intercept) -0.0000000000000004171  0.0150479364364516036   0.000               1.0000
Low          0.1869661995234728480  0.0746073783668983193   2.506               0.0134 *
High         1.0674470664586244162  0.0760405896747291810  14.038 <0.0000000000000002 ***
Vol          0.0021472895742555252  0.0187048074867217015   0.115               0.9088
Open        -0.2358586486650242620  0.0916450682048002147  -2.574               0.0111 *
Avg         -0.0367623894490373485  0.0493888262923014854  -0.744               0.4580
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1787 on 135 degrees of freedom
Multiple R-squared:  0.9692,    Adjusted R-squared:  0.9681
F-statistic:   850 on 5 and 135 DF,  p-value: < 0.00000000000000022
```

Figure 22. *Linear regression model of EVS.*

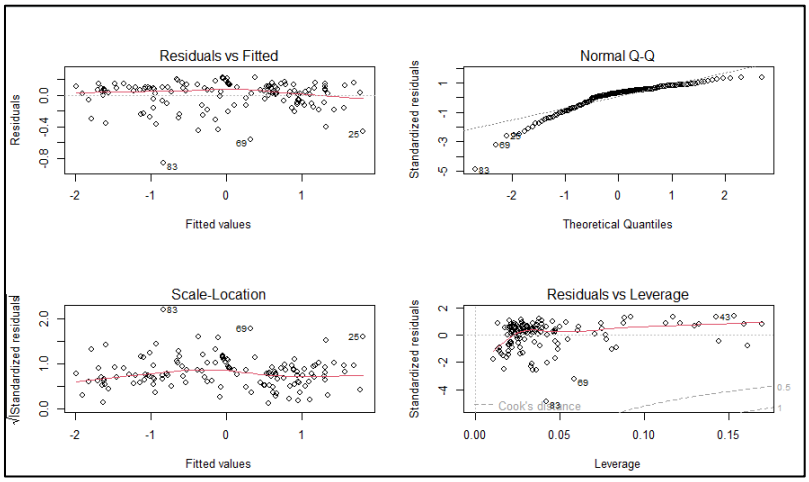

Figure 23. *4 related graphs.*

- F&B (SLS):

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.50355 -0.04728  0.01310  0.06573  0.51620

Coefficients:
                        Estimate             Std. Error t value     Pr(>|t|)
(Intercept) -0.0000000000000005214  0.0113938397177766381   0.000      1.00000
Low          0.2099643907757235417  0.0937197004283777391   2.240      0.02670 *
High         0.5673234968853140270  0.0956100799651119770   5.934 0.0000000236 ***
Vol          0.0028513732093552323  0.0115150217134696517   0.248      0.80480
Open        -0.2215690537847813113  0.0731637687937884557  -3.028      0.00295 **
Avg          0.4373720302447643982  0.1328648061613044162   3.292      0.00127 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1353 on 135 degrees of freedom
Multiple R-squared:  0.9823,    Adjusted R-squared:  0.9817
F-statistic:  1501 on 5 and 135 DF,  p-value: < 0.00000000000000022
```

Figure 24. *Linear regression model of OCH.*

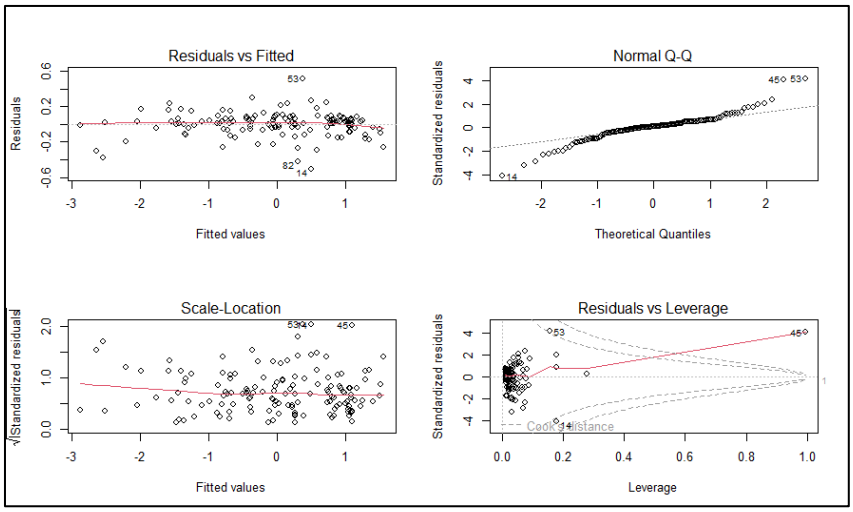

Figure 25. *4 related graphs.*

3.1.5. Cross-model Usability

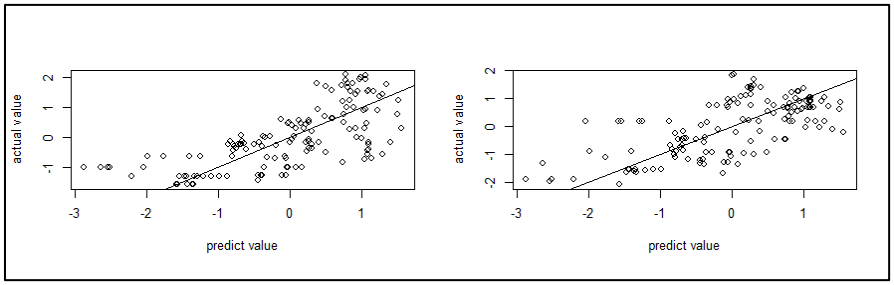

Figure 26. *Apply model of SLS to the others.*

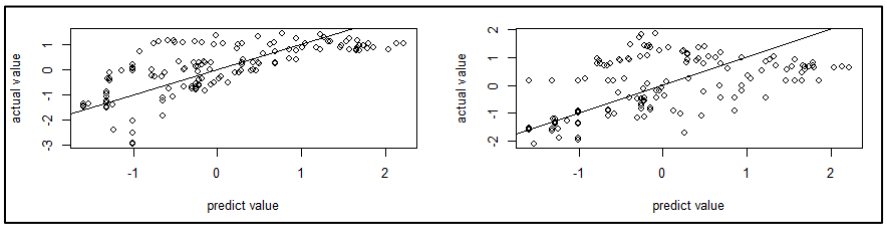

Figure 27. *Apply model of OCH to the others.*

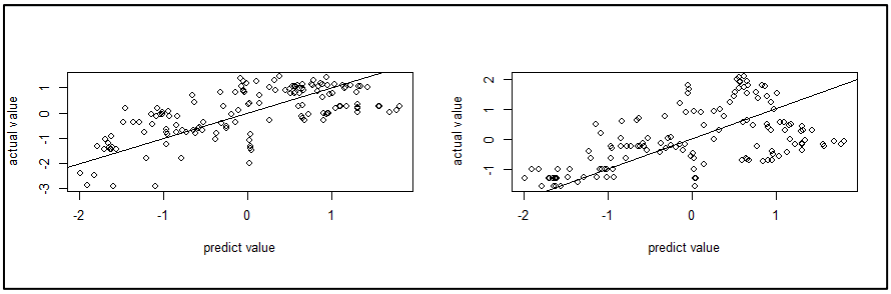

Figure 28. *Apply model of EVS to the others.*

It is obvious that we can apply model of OCH to SLS and EVS although it has some errors.

3.1.6. Conclusion

In conclusion, the influential factors include Low, High and Open.

3.2. Three stocks with the strongest drop rate on HNINDEX in the period from January to July 2022

3.2.1. Source

Stock	Rate of change (ROC)
THD	-0,78343465
SHS	-0,746124031
ART	-0,715151515
BII	-0,703703704
IDJ	-0,701010101
DVG	-0,697674419
SCI	-0,690860215
IPA	-0,688821752

Figure 28. *The Rate of change with ascendsing order.*

- THD: Thaiholdings Joint Stock Company

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=13728

- SHS: Saigon – Hanoi Securities JSC

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=13728

- ART: Bos Securities Corporation

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=12979

3.2.2. Data Visualization

- THD:

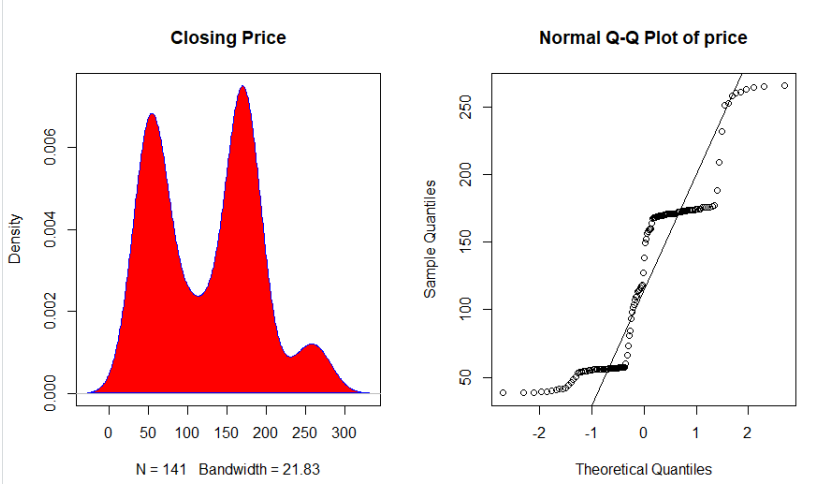

Figure 29. *The distribution of THD.*

- SHS:

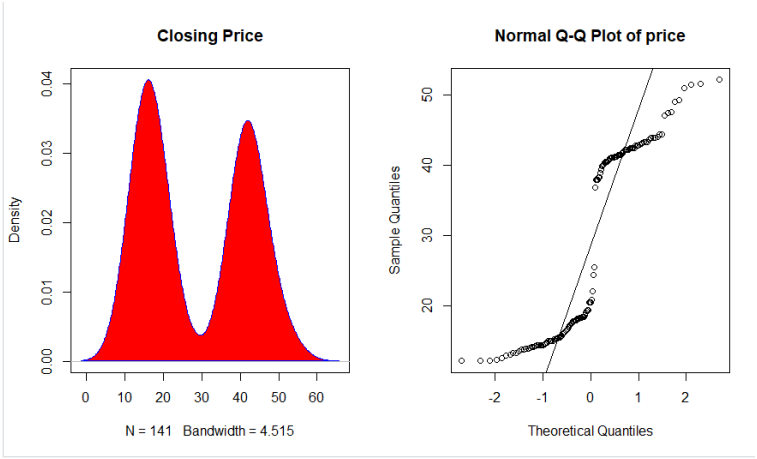

Figure 30. *The distribution of SHS.*

- ART:

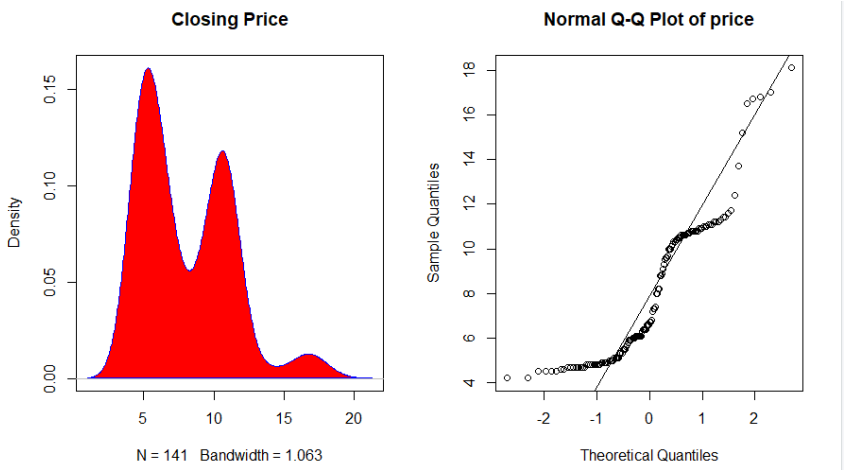

Figure 31. *The distribution of ART.*

3.2.3. Correlation Plot

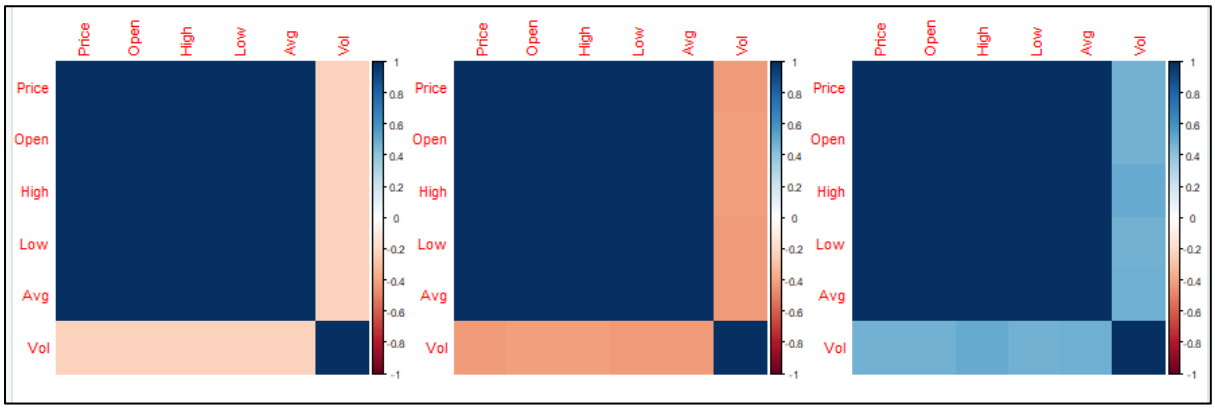

Figure 32. *Correlation plots of THD, SHS ,ART ,repectively.*

3.2.4. Regressive Model

Overall formula:

closing_price = $\alpha + \beta_1$* opening_price + β_2* highest_price + β_3* lowest_price + β_4* volume + β_5* avg + ε , where $\varepsilon \sim N(0, \sigma^2)$

- THD:

```
Call:
lm(formula = Price ~ Low + High + Vol + Open + Avg, data = scaling_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.85552 -0.07587  0.05571  0.10792  0.22588

Coefficients:
                        Estimate             Std. Error t value            Pr(>|t|)
(Intercept) -0.0000000000000004171  0.0150479364364516036   0.000               1.0000
Low          0.1869661995234728480  0.0746073783668983193   2.506               0.0134 *
High         1.0674470664586244162  0.0760405896747291810  14.038 <0.0000000000000002 ***
Vol          0.0021472895742555252  0.0187048074867217015   0.115               0.9088
Open        -0.2358586486650242620  0.0916450682048002147  -2.574               0.0111 *
Avg         -0.0367623894490373485  0.0493888262923014854  -0.744               0.4580
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1787 on 135 degrees of freedom
Multiple R-squared:  0.9692,    Adjusted R-squared:  0.9681
F-statistic:   850 on 5 and 135 DF,  p-value: < 0.00000000000000022
```

Figure 33. *Linear regression model of THD.*

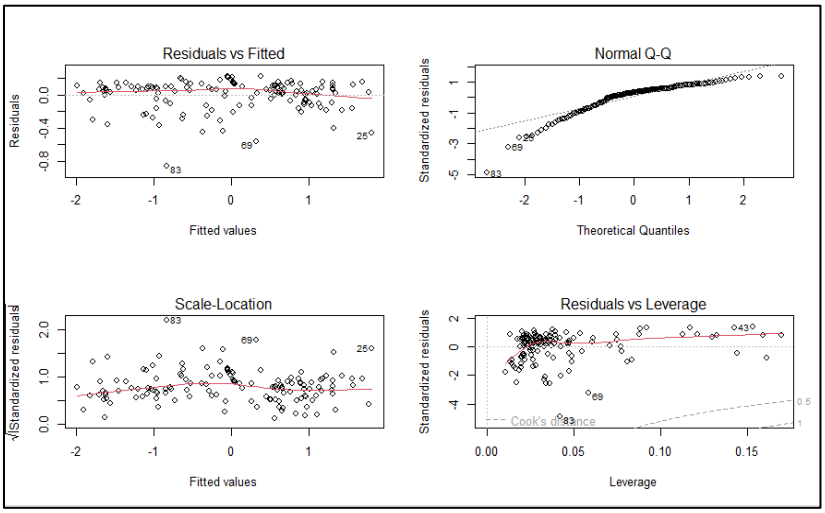

Figure 34. *4 related graphs.*

- SHS:

```
Call:
lm(formula = Price ~ Low + High + Vol + Open + Avg, data = scaling_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.092677 -0.020795 -0.002625  0.019599  0.111267

Coefficients:
                       Estimate             Std. Error t value        Pr(>|t|)
(Intercept)  0.0000000000000001852  0.002798424299642947 9   0.000         1.00000
Low          0.2403405433836611760  0.079764565459385233 8   3.013         0.00309 **
High         0.1153667099131378132  0.107962763882972281 9   1.069         0.28717
Vol          0.0037561369696292937  0.003309425878178281 2   1.137         0.25775
Open        -0.3422279283272789407  0.057476103990488233 7  -5.954 0.0000000213685 ***
Avg          0.9869388931807455689  0.133762513673972016 6   7.378 0.0000000000147 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03323 on 135 degrees of freedom
Multiple R-squared:  0.9989,    Adjusted R-squared:  0.9989
F-statistic: 2.533e+04 on 5 and 135 DF,  p-value: < 0.00000000000000022
```

Figure 35. *Linear regression model of SHS.*

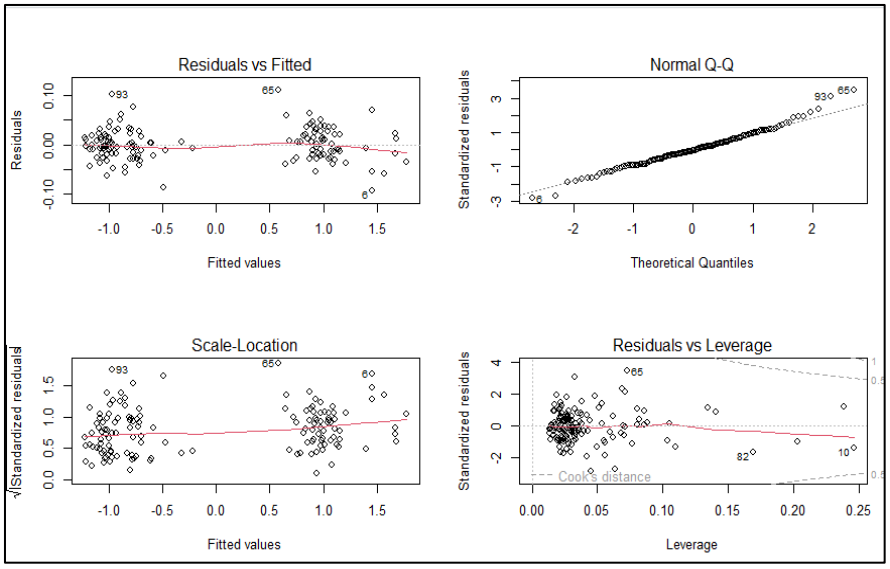

Figure 36. *4 related graphs.*

- ART:

```
Call:
lm(formula = Price ~ Low + High + Vol + Open + Avg, data = scaling_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41853 -0.02991 -0.00913  0.02738  0.21702

Coefficients:
                       Estimate             Std. Error t value        Pr(>|t|)
(Intercept)  0.0000000000000004718  0.005009943392255755 00   0.000         1.00000
Low          0.3086966138302963957 9  0.092199432992207161 49   3.348         0.00105 **
High         0.3077472892053930153 5  0.127925033623286998 90   2.406         0.01750 *
Vol         -0.0164572303796263162 7  0.007203498878843512 74  -2.285         0.02390 *
Open        -0.4435089714865716459 3  0.091659109401474342 38  -4.839 0.000003516688 ***
Avg          0.8322358099431196309 9  0.120622898910643663 87   6.899 0.000000000184 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05949 on 135 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9965
F-statistic:  7885 on 5 and 135 DF,  p-value: < 0.00000000000000022
```

Figure 37. *Linear regression model of ART.*

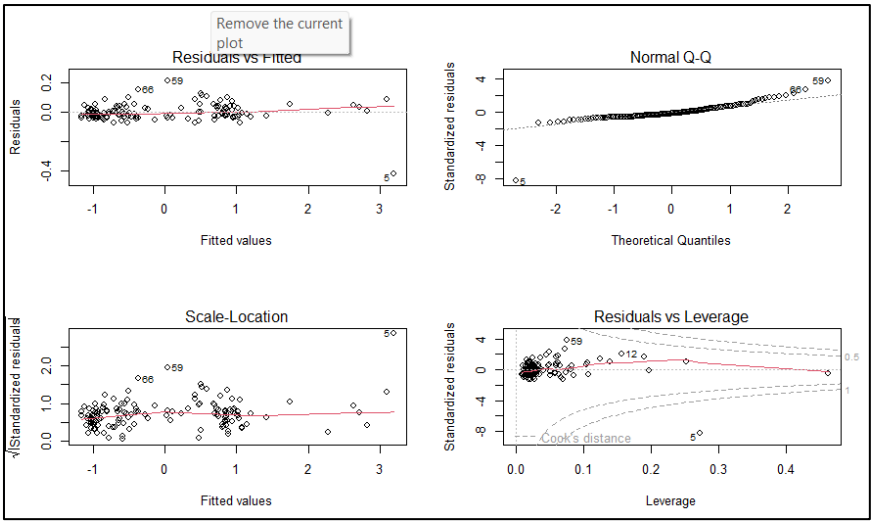

Figure 38. *4 related graphs.*

3.2.5. Cross-model Usability

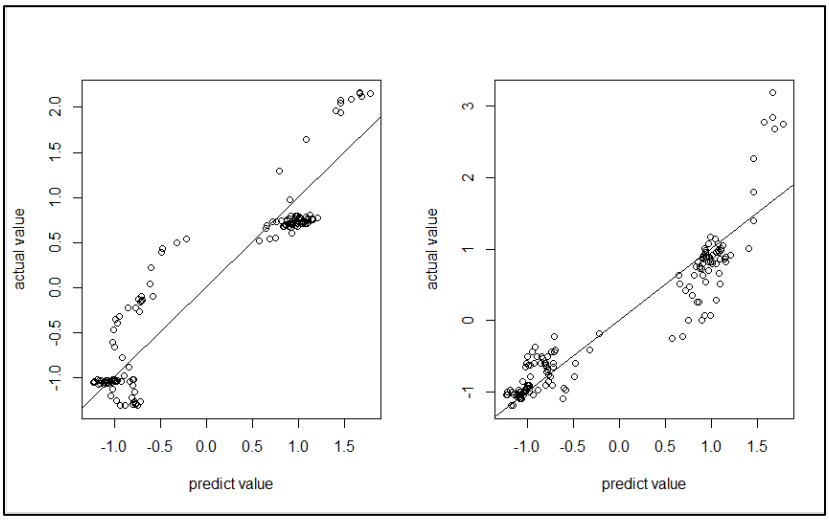

Figure 39. *Apply SHS model to the others.*

We can apply SHS model to THD and ART although it has some errors.

3.2.6. Conclusion

In conclusion, the influential factors include Low and Open.

3.3. Three stocks with the strongest growth rate on HNINDEX in the period from January to July 2022

3.3.1. Source

Stock	Rate of change (ROC)
CLM	1,986622074
VLA	1,366666667
SGD	1,142857143
SDU	1,08125
CJC	0,72972973
SGC	0,627218935
HEV	0,612676056

Figure 28. *The Rate of change with descendsing order.*

- CLM: Vinacomin – Coal Import Export Joint Stock Company

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=12400

- VLA: Van Lang Technology Development & Investment JSC

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=320

- SGD: Educational Book JSC In Ho Chi Minh City

https://finance.vietstock.vn/ket-qua-giao-dich?tab=thong-ke-gia&exchange=2&code=62

3.3.2. Data Visualization

- CLM:

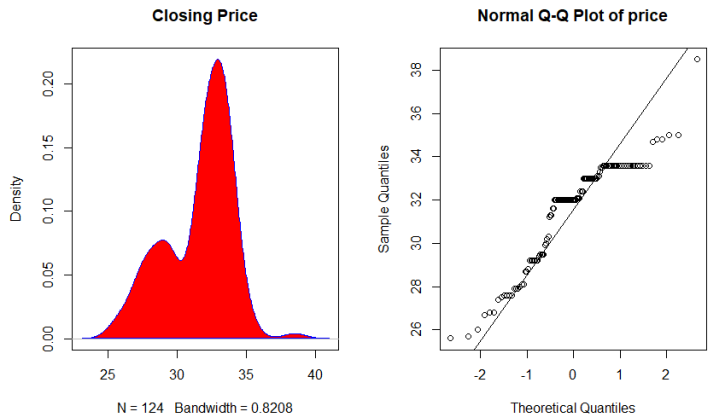

Figure 40. *The distribution of CLM.*

- VLA:

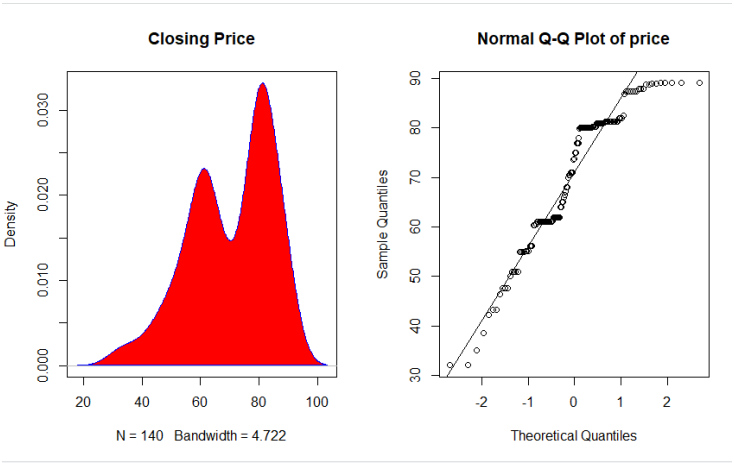

Figure 41. *The distribution of VLA.*

- SGD:

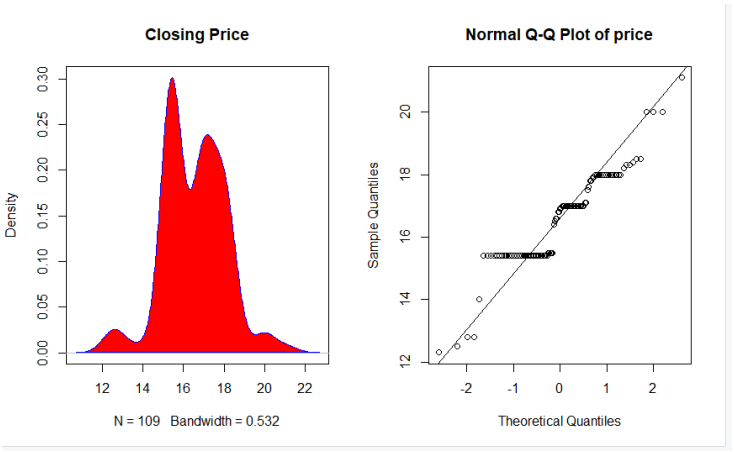

Figure 42. *The distribution of SGD.*

3.3.3. Correlation Plot

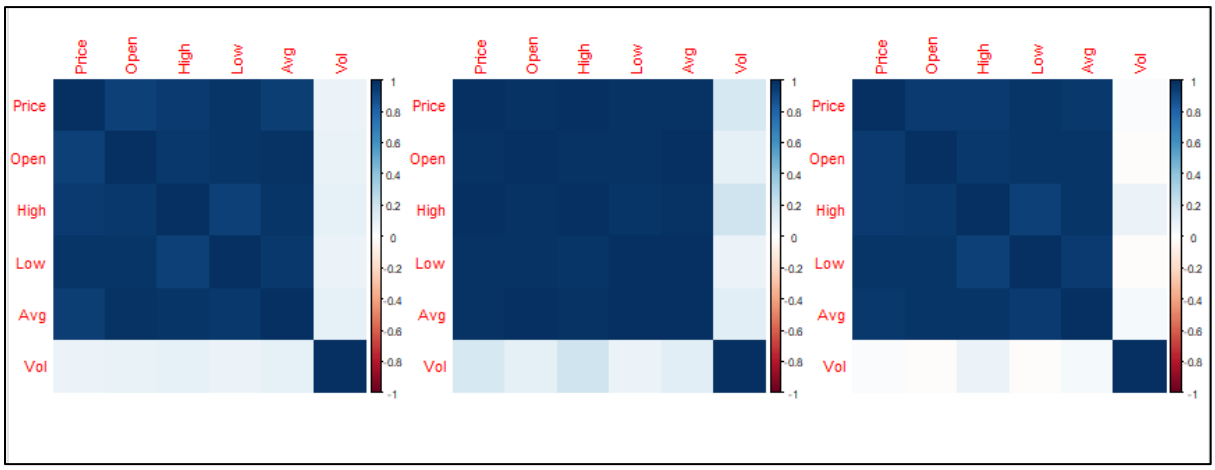

Figure 43. *Correlation plots of CLM, VLA ,SGD repectively.*

3.3.4. Regressive Model

Overall formula:

closing_price = $\alpha + \beta_1$* opening_price + β_2* highest_price + β_3* lowest_price + β_4* volume + β_5* avg + ε , where $\varepsilon \sim N(0, \sigma^2)$

- CLM:

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.82691 -0.00844  0.00585  0.01126  0.30169

Coefficients:
                        Estimate             Std. Error t value             Pr(>|t|)
(Intercept) -0.0000000000000006192  0.0090856102555719152   0.000                1.000
Low          1.1257311509124612314  0.0394603398126408475  28.528 < 0.0000000000000002 ***
High         0.9041135761223276779  0.0425358861886124887  21.255 < 0.0000000000000002 ***
Vol         -0.0019495793899807828  0.0091863374104602762  -0.212                0.832
Open        -0.6275695167162487209  0.0664547133338244805  -9.444 0.000000000000000417 ***
Avg         -0.4040460171603837969  0.0699103671435155616  -5.779 0.000000062298835517 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1012 on 118 degrees of freedom
Multiple R-squared:  0.9902,    Adjusted R-squared:  0.9898
F-statistic:  2380 on 5 and 118 DF,  p-value: < 0.00000000000000022
```

Figure 44. *Linear regression model of CLM.*

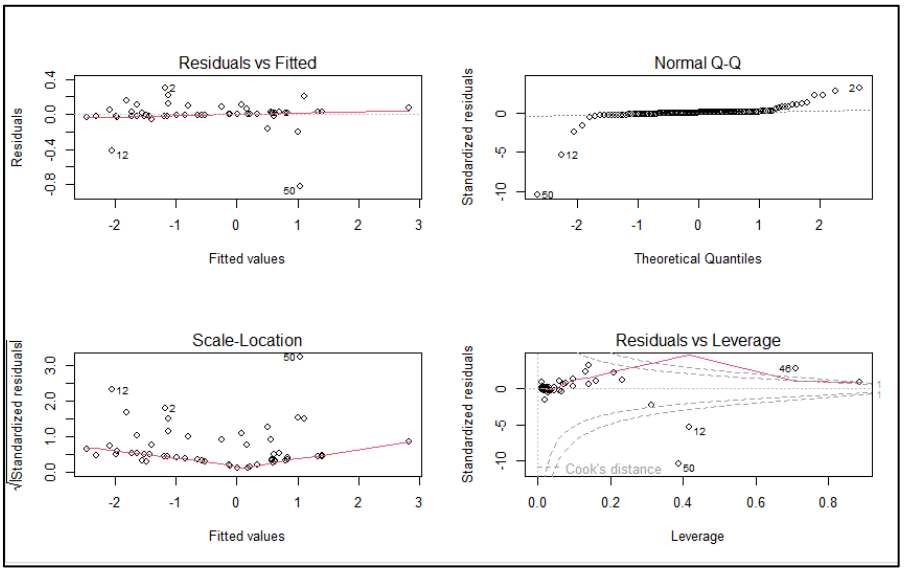

Figure 45. *4 related graphs.*

- VLA:

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.40221 -0.00303  0.00362  0.00744  0.33144

Coefficients:
                        Estimate             Std. Error t value             Pr(>|t|)
(Intercept) -0.0000000000000007435  0.0058491293427694524   0.000                1.000
Low          0.6374482658861491124  0.0625828756121937085  10.186 <0.0000000000000002 ***
High         1.0117374348526455563  0.0500309303867475033  20.222 <0.0000000000000002 ***
Vol         -0.0041888072888284882  0.0073788935396122985  -0.568                0.571
Open        -0.6378899079515207848  0.0605679072648233896 -10.532 <0.0000000000000002 ***
Avg         -0.0093608670174135170  0.0721450141393307487  -0.130                0.897
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06921 on 134 degrees of freedom
Multiple R-squared:  0.9954,    Adjusted R-squared:  0.9952
F-statistic:  5777 on 5 and 134 DF,  p-value: < 0.00000000000000022
```

Figure 46. *Linear regression model of VLA.*

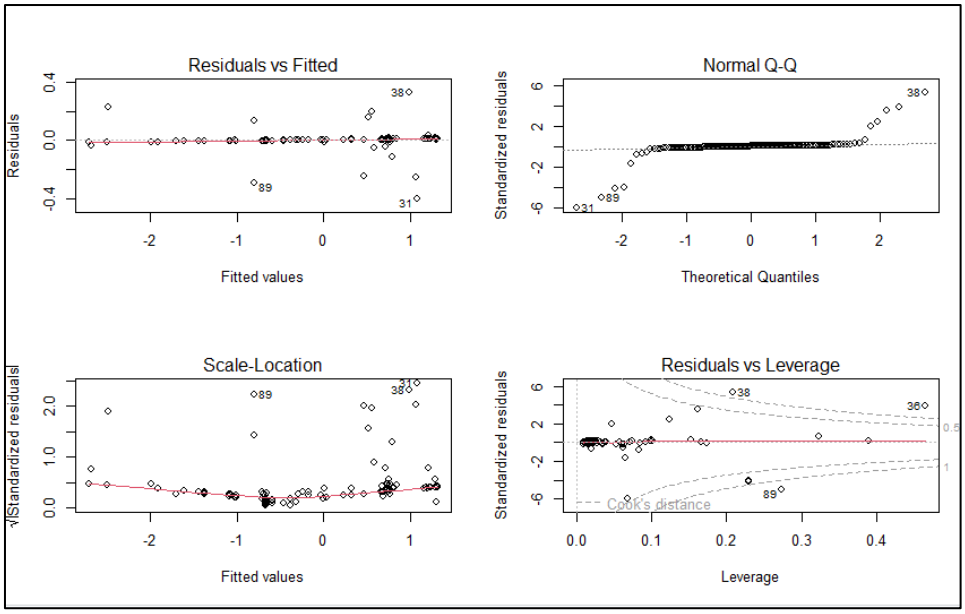

Figure 47. *4 related graphs.*

- SGD:

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.59010 -0.04296 -0.00574  0.01753  0.56551

Coefficients:
                      Estimate            Std. Error t value             Pr(>|t|)
(Intercept)  0.000000000000002629  0.014139742383634732   0.000                1.000
Low          0.8843995701921236 74  0.065007633154456387  13.605 < 0.0000000000000002 ***
High         0.4807937275938039 55  0.072875285442231957   6.597        0.00000000184 ***
Vol         -0.0155154405688664 68  0.015426027368646440  -1.006                0.317
Open        -0.4674847386201795 85  0.088115971276967783  -5.305        0.00000064751 ***
Avg          0.1057745691237794 02  0.080098202258525000   1.321                0.190
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1476 on 103 degrees of freedom
Multiple R-squared:  0.9792,    Adjusted R-squared:  0.9782
F-statistic: 970.6 on 5 and 103 DF,  p-value: < 0.00000000000000022
```

Figure 48. *Linear regression model of SGD.*

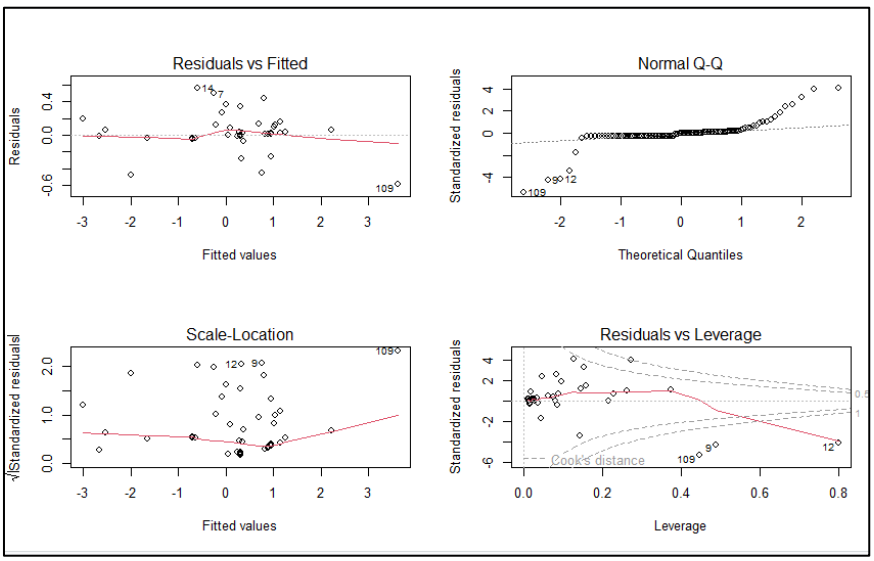

Figure 49. *4 related graphs.*

3.3.5. Cross-model Usability

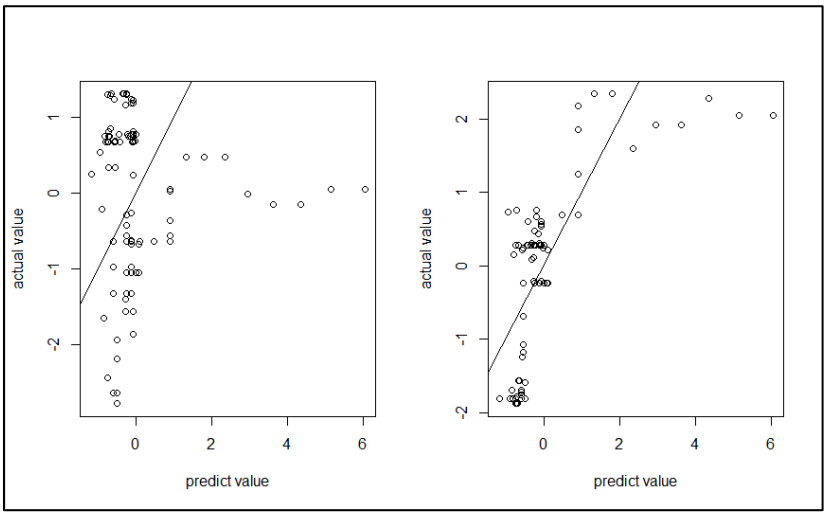

Figure 50. *Apply CLM model to the others.*

If we ignore some outliers, we can apply model of CLM to VLA and SGD. However, if we want to have an accuarate result, this method does not work well.

3.3.6. Conclusion

In conclusion, the influential factors include Low, High and Open.

4. The effect of the Economic Crisis on stock index sector

4.1. Stock Market Sectors

In this part, we will analyze 11 GICS (stands for Global Industry Classification Standard) stock market sectors, including:

Figure 51. *11 GICS stock market sectors.*

For more information, you can access this website:

https://www.fool.com/investing/stock-market/market-sectors/

The table below indicates more details about each stock market sector:

Energy	covers companies that do business in the **oil, natural gas and consumable fuels** industry.
Materials	includes companies that provide various **goods** for use **in manufacturing** and other applications.
Industrials	encompasses a wide range of different businesses like **airlines, railroads, and logistics** companies.
Utilities	encompasses every different type of utility activities such as **making electrical power** available **to residential** and commercial customers, as well as things like **natural gas transmission and distribution** or **water delivery**.
Healthcare	Includes all companies **related to health**; for example, pharmacies, hospitals, etc.
Financials	includes businesses that are primarily related to **handling money**.
Consumer Discretionary	covers **goods** and **services** for which consumer demand **depends upon** consumer **financial status**.
Consumer Staples	provides **necessities** for consumers, **regardless of** their **financial situation**. Things like food and clothing are the examples.
Information Technology	includes businesses engaged in several fields of **technological innovation**.
Communication Services	contains companies that **provide communications services** primarily **through** a **fixed-line, cellular, wireless**, high bandwidth and/or fiber optic cable network.
Real Estate	includes different types of investments related to **real estate**.

4.1.1. Energy

- BKC: CTCP Khoáng Sản Bắc Kạn

https://vn.investing.com/equities/bamcorp.-historical-data

- HGM: CTCP Than Núi Béo - Vinacomin

https://vn.investing.com/equities/vnbc-historical-data

- HLC: CTCP Than Hà Lầm – Vinacomin

https://vn.investing.com/equities/vhlc-historical-data

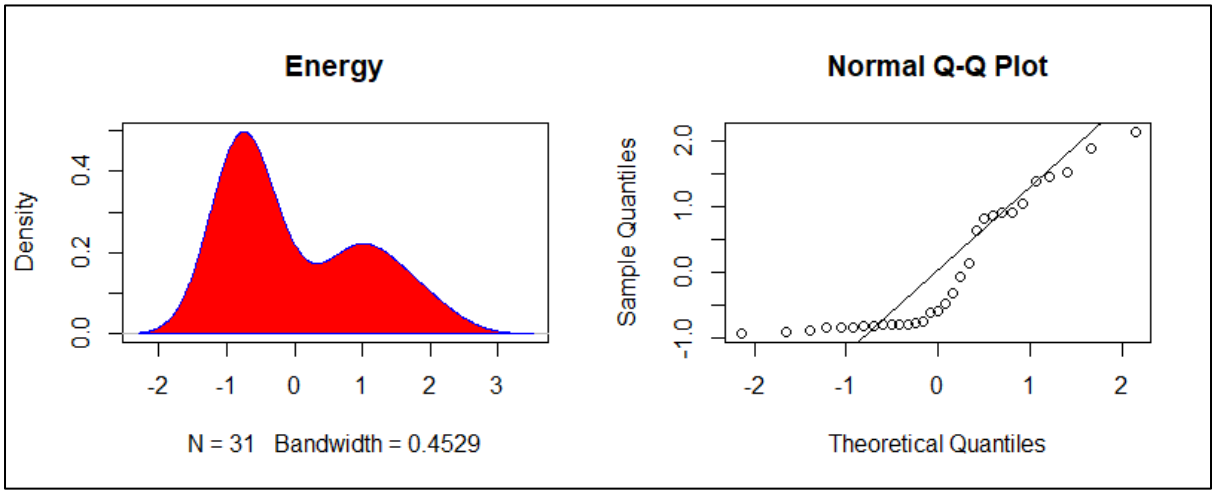

Figure 52. *The distribution of Energy.*

4.1.2. Materials

- BCC: CTCP Xi măng Bỉm Sơn

https://vn.investing.com/equities/bim-son-cement-jsc-historical-data

- BTS: CTCP Xi măng VICEM Bút Sơn

https://vn.investing.com/equities/but-son-cement-jsc-historical-data

- GKM: CTCP Khang Minh Group

https://vn.investing.com/equities/khang-minh-brick-historical-data

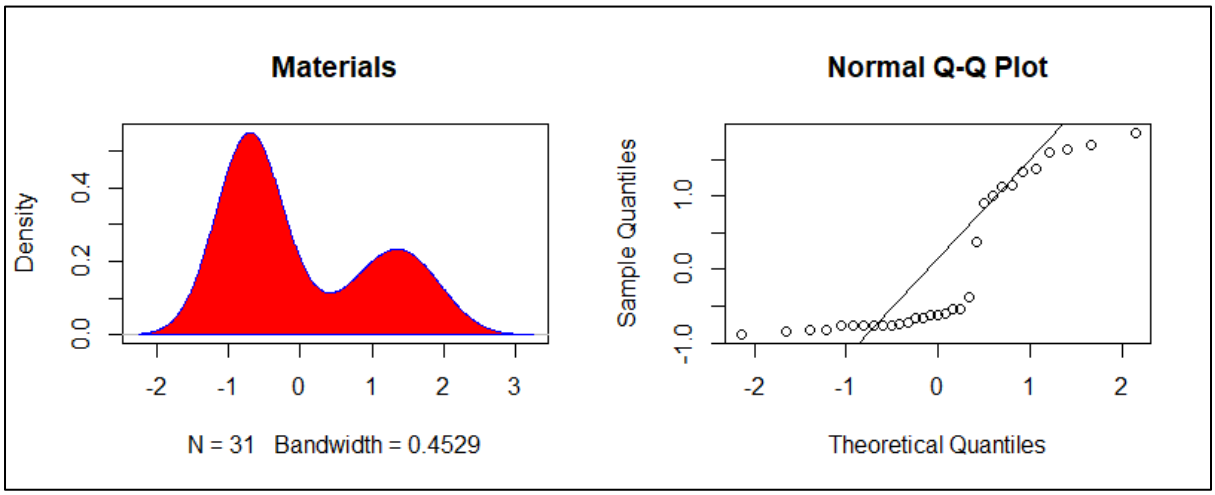

Figure 53. *The distribution of Materials.*

4.1.3. Industrials

- CCR: CTCP Cảng Cam Ranh

https://vn.investing.com/equities/camranh-port-jsc-historical-data

- GMD: CTCP Gemadept

https://vn.investing.com/equities/gemadept-corp-historical-data

- CDN: CTCP Cảng Đà Nẵng

https://vn.investing.com/equities/danang-port-jsc-historical-data

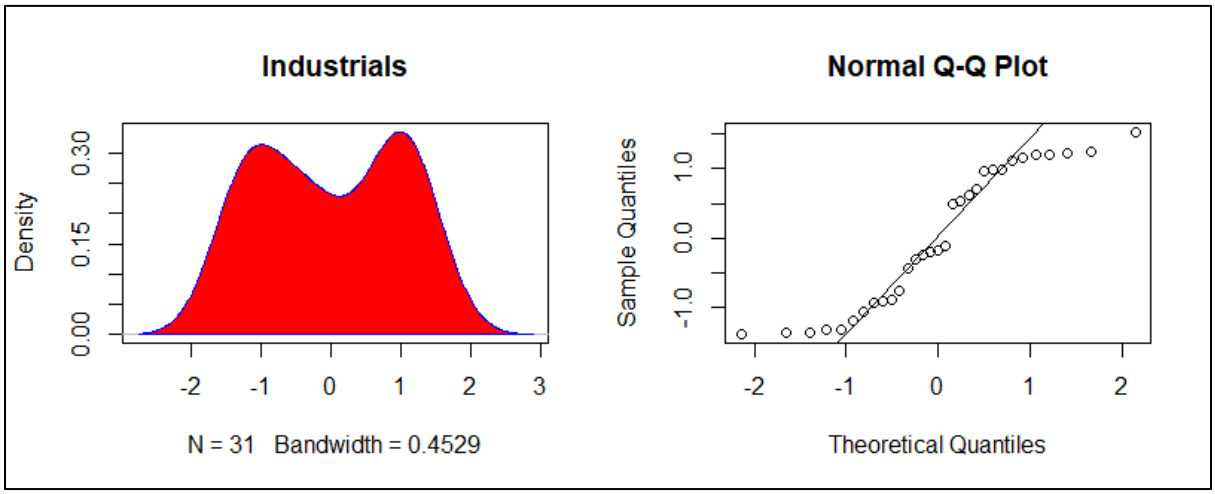

Figure 54. *The distribution of Industrials.*

4.1.4. Utlities

- GAS: Tổng Công ty Khí Việt Nam

https://vn.investing.com/equities/petrovietnam-gas-jscrp-historical-data

- DDG: CTCP Đầu tư Công nghiệp Xuất nhập khẩu Đông Dương

https://vn.investing.com/equities/indochine-import-export-historical-data

- BWE: CTCP Nước - Môi trường Bình Dương

https://vn.investing.com/equities/binh-duong-water-historical-data

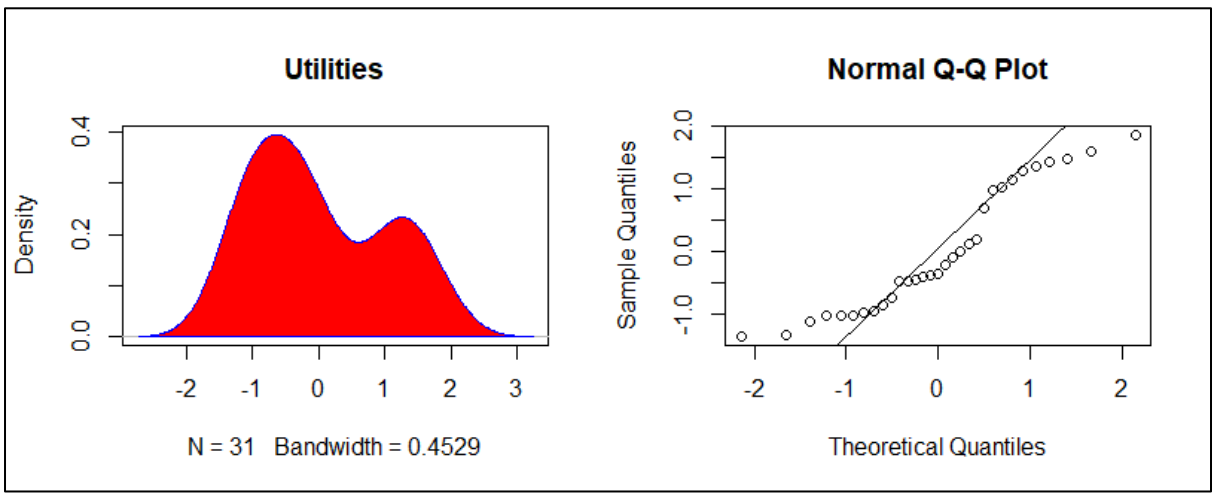

Figure 55. *The distribution of Utilities.*

4.1.5. Healthcare

- IMP: CTCP Dược phẩm IMEXPHARM

https://vn.investing.com/equities/imexpharm-corp-historical-data

- DHT: CTCP Dược phẩm Hà Tây

https://vn.investing.com/equities/hataphar-historical-data

- DHG: CTCP Dược Hậu Giang

https://vn.investing.com/equities/dhg-pharmaceutical-jsc-historical-data

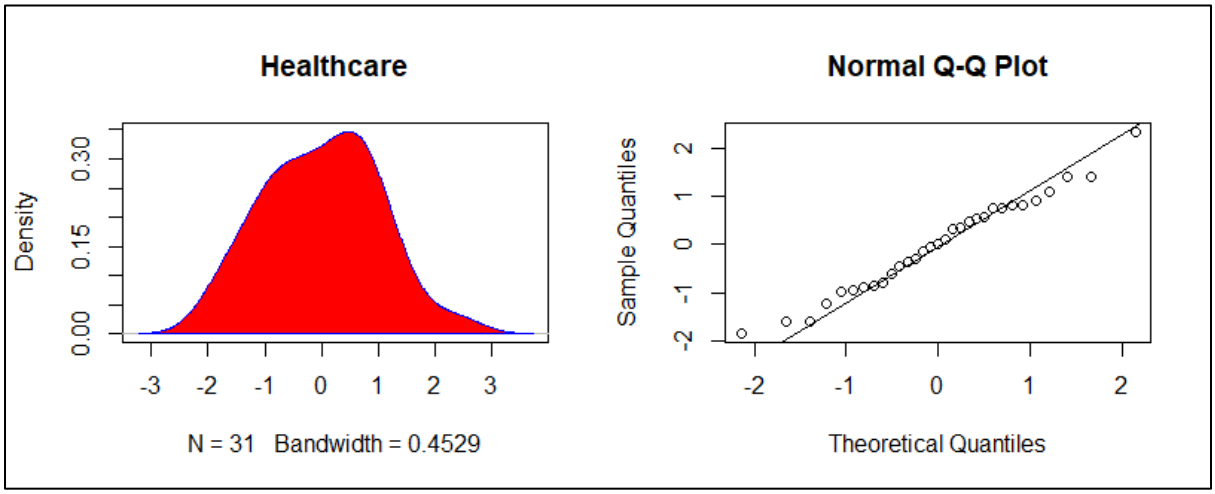

Figure 56. *The distribution of Healthcare.*

4.1.6. Financials

- ACB: Ngân hàng TMCP Á Châu

https://vn.investing.com/equities/asia-commercial-bank-historical-data

- SHB: Ngân hàng TMCP Sài Gòn - Hà Nội

https://vn.investing.com/equities/shb-historical-data

- TVC: CTCP Tập đoàn Quản lý Tài sản Trí Việt

https://vn.investing.com/equities/tri-viet-management-investment-corp-historical-data

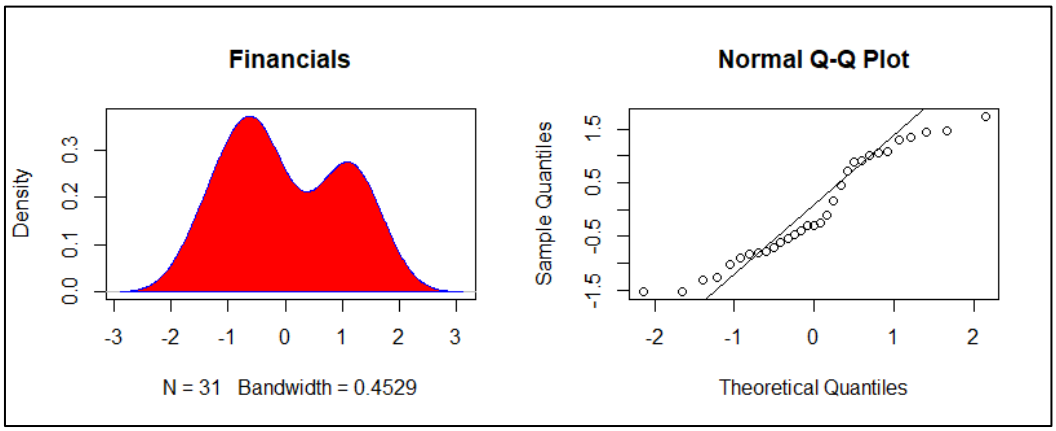

Figure 57. *The distribution of Financials.*

4.1.7. Consumer Discretionary

- DAH: CTCP Tập đoàn Khách sạn Đông Á

https://vn.investing.com/equities/dong-a-hotel-group-jsc-historical-data

- FRT: CTCP Bán lẻ Kỹ thuật số FPT

https://vn.investing.com/equities/fpt-digital-historical-data

- CVN: CTCP Vinam

https://vn.investing.com/equities/vinam.,jsc-historical-data

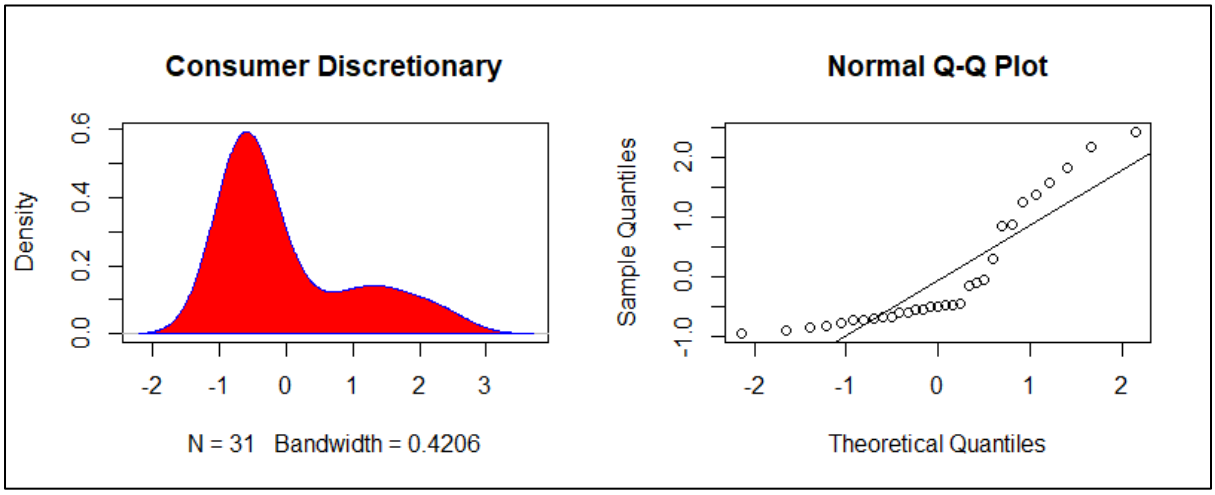

Figure 58. *The distribution of Consumer Discretionary.*

4.1.8. Consumer Staples

- DBC: CTCP Tập đoàn Dabaco Việt Nam

https://vn.investing.com/equities/dabaco-historical-data

- TFC: CTCP Trang

https://vn.investing.com/equities/trangcorp-corporation-jsc-historical-data

- TDT: CTCP Đầu tư và Phát triển TDT

https://vn.investing.com/equities/tdt-invest-historical-data

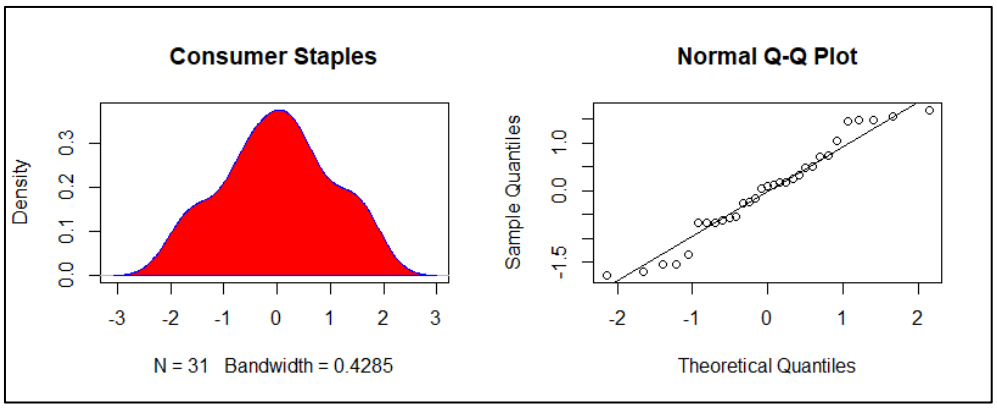

Figure 59. *The distribution of Consumer Staples.*

4.1.9. Information Technology

- KST: CTCP KASATI

https://vn.investing.com/equities/kasati-historical-data

- FPT: CTCP FPT

https://vn.investing.com/equities/fpt-corp-historical-data

- CMG: CTCP Đầu Tư CMC

https://vn.investing.com/equities/cmc-corp-historical-data

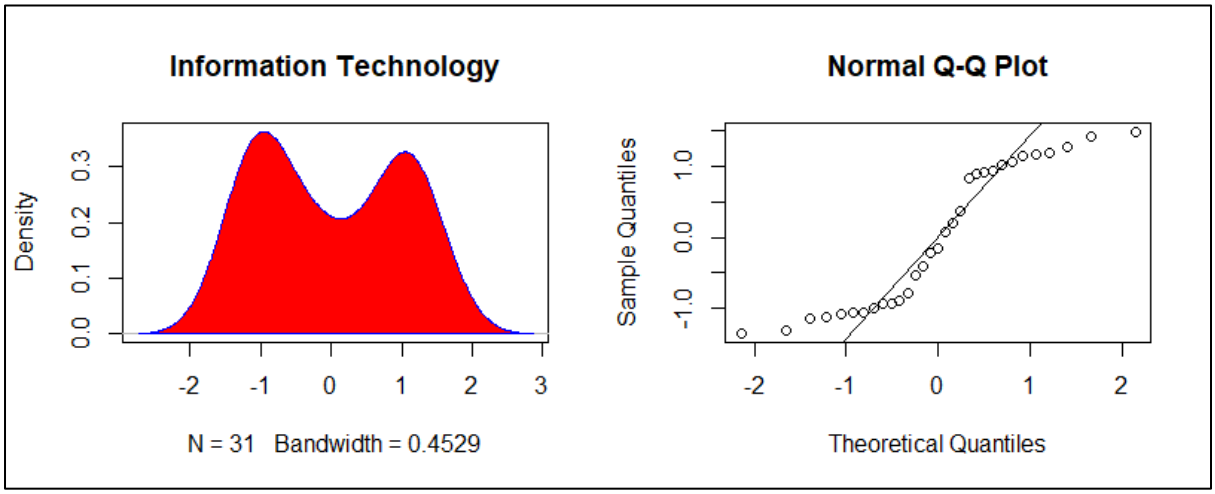

Figure 60. *The distribution of Information Technology.*

4.1.10. Communication Services

- SDA: CTCP Simco Sông Đà

https://vn.investing.com/equities/song-da-manpower-supply-trading-jsc-historical-data

- TV4: CTCP Tư vấn Xây dựng Điện 4

https://vn.investing.com/equities/pecc4-historical-data

- IBC: CTCP Đầu tư Apax Holdings

https://vn.investing.com/equities/apax-historical-data

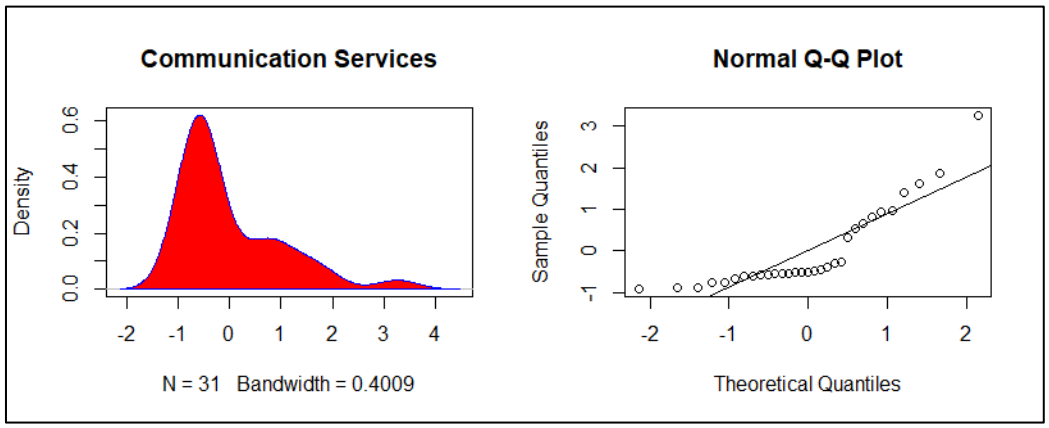

Figure 61. *The distribution of Communication Services.*

4.1.11. Real Estate

- API: CTCP Đầu tư Châu Á - Thái Bình Dương

https://vn.investing.com/equities/apec-investmen-historical-data

- BII: CTCP Louis Land

https://vn.investing.com/equities/bao-thu-industrial-develop-invest-historical-data

- HUT: CTCP Tasco

https://vn.investing.com/equities/hud---tasco-historical-data

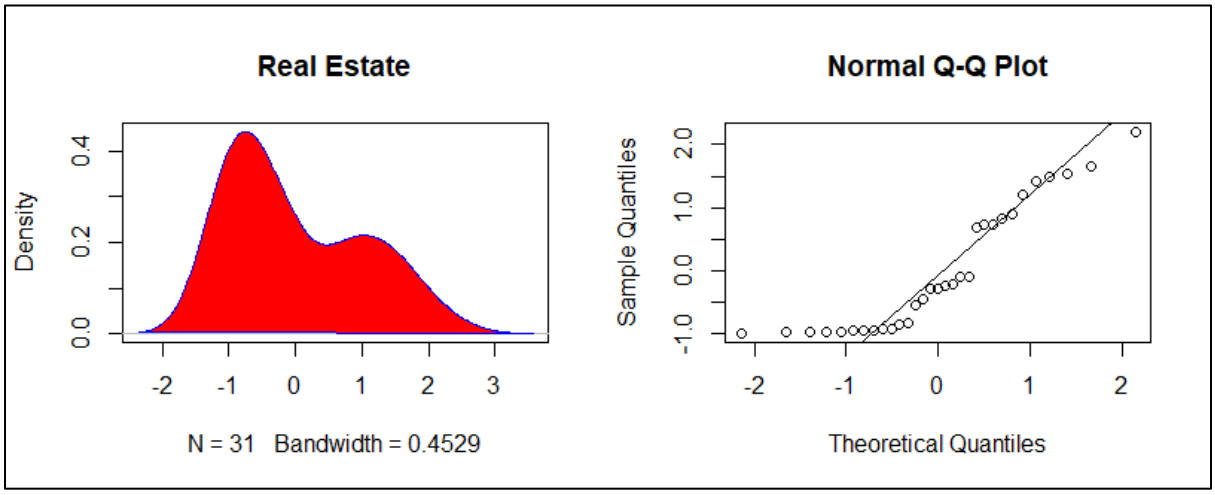

Figure 62. *The distribution of Real Estate.*

4.2. Impact Factors

Then, we choose 5 factors that impose big impact on the economy, including: Stock Market, Inflation Rate, Balance of Trade, Manufacturing PMI, Currency. We will explain why we make this selection with the following rationales:

- **Stock market** refers to several exchanges in which shares of publicly held companies are bought and sold. This means that it allows buyers and sellers of securities to meet, interact, and transact. Therefore, the competitions between market participants serve as a barometer for the overall economy.

- **Inflation** can be simply interpreted as a rise in prices. Therefore, high and variable rates of inflation can impose major costs on an economy, which make this one of important factors to the economy.

- **Balance of Trade** (BOT) is the difference between the value of a country's exports and the value of a country's imports for a given period. This index is also one of the key factors that ensure economic stability.

- **Purchasing Managers' Index** (PMI) is an index of the prevailing direction of economic trends in the manufacturing and service sectors. Therefore, paying attention to the value and movements in the PMI can yield profitable foresight into developing trends in the overall economy.

- **Currency** fluctuations are a natural outcome of floating exchange rates (determined by supply and demand on the open market), which is the norm for most major economies.

Energy	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -1.328e-16 1.041e-01 0.000 1.0000 Stock_Market 1.075e+00 1.803e-01 5.965 3.15e-06 *** Inflation_Rate 2.213e-01 1.152e-01 1.921 0.0662 . Balance_of_Trade 2.919e-01 1.195e-01 2.443 0.0220 * Manufacturing_PMI -1.252e-01 1.159e-01 -1.080 0.2904 Currency 9.261e-02 1.584e-01 0.585 0.5641
Materials	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -3.130e-17 1.180e-01 0.000 1.0000 Stock_Market 9.750e-01 2.044e-01 4.771 6.75e-05 *** Inflation_Rate 1.652e-01 1.306e-01 1.265 0.2175 Balance_of_Trade 2.488e-01 1.355e-01 1.836 0.0783 . Manufacturing_PMI 9.070e-02 1.314e-01 0.690 0.4963 Currency 1.673e-01 1.796e-01 0.932 0.3604
Industrials	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 1.522e-16 6.448e-02 0.000 1.000 Stock_Market 1.124e+00 1.117e-01 10.060 2.84e-10 *** Inflation_Rate 6.175e-02 7.139e-02 0.865 0.395 Balance_of_Trade 9.303e-02 7.405e-02 1.256 0.221 Manufacturing_PMI -1.107e-01 7.180e-02 -1.541 0.136 Currency 1.348e-01 9.817e-02 1.374 0.182
Utilities	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -6.908e-16 9.368e-02 0.000 1.0000 Stock_Market 1.115e+00 1.623e-01 6.870 3.36e-07 *** Inflation_Rate 1.683e-01 1.037e-01 1.622 0.1173 Balance_of_Trade 1.186e-01 1.076e-01 1.102 0.2808 Manufacturing_PMI 1.506e-01 1.043e-01 1.444 0.1612 Currency 3.554e-01 1.426e-01 2.492 0.0197 *
Healthcare	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 8.826e-16 7.513e-02 0.000 1.0000 Stock_Market 6.775e-01 1.302e-01 5.205 2.19e-05 *** Inflation_Rate -9.194e-02 8.318e-02 -1.105 0.2796 Balance_of_Trade 1.209e-01 8.628e-02 1.402 0.1733 Manufacturing_PMI 3.042e-03 8.367e-02 0.036 0.9713 Currency -2.889e-01 1.144e-01 -2.525 0.0183 *

Financials	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 2.825e-16 7.109e-02 0.000 1.000000 Stock_Market 9.868e-01 1.231e-01 8.013 2.28e-08 *** Inflation_Rate 1.002e-01 7.870e-02 1.273 0.214663 Balance_of_Trade 3.550e-02 8.164e-02 0.435 0.667394 Manufacturing_PMI -3.230e-01 7.916e-02 -4.081 0.000403 *** Currency -8.261e-02 1.082e-01 -0.763 0.452460
Consumer Discretionary	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -1.950e-16 1.363e-01 0.000 1.000000 Stock_Market 9.501e-01 2.361e-01 4.025 0.000465 *** Inflation_Rate 1.669e-01 1.509e-01 1.106 0.279093 Balance_of_Trade 1.656e-01 1.565e-01 1.058 0.300076 Manufacturing_PMI 1.617e-01 1.517e-01 1.066 0.296822 Currency 3.676e-01 2.075e-01 1.772 0.088616 .
Consumer Staples	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 1.382e-16 4.960e-02 0.000 1.00000 Stock_Market 8.260e-01 8.593e-02 9.613 7.07e-10 *** Inflation_Rate -1.789e-01 5.492e-02 -3.258 0.00322 ** Balance_of_Trade 2.036e-01 5.696e-02 3.574 0.00147 ** Manufacturing_PMI 3.154e-02 5.524e-02 0.571 0.57307 Currency -1.129e-01 7.552e-02 -1.495 0.14746
Information Technology	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -2.593e-16 7.488e-02 0.000 1.0000 Stock_Market 1.167e+00 1.297e-01 8.995 2.6e-09 *** Inflation_Rate 1.198e-01 8.291e-02 1.445 0.1609 Balance_of_Trade 1.352e-01 8.600e-02 1.573 0.1284 Manufacturing_PMI -6.896e-02 8.339e-02 -0.827 0.4161 Currency 2.092e-01 1.140e-01 1.835 0.0785 .
Communication Services	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 1.804e-16 1.186e-01 0.000 1.000000 Stock_Market 8.358e-01 2.055e-01 4.067 0.000417 *** Inflation_Rate 1.376e-01 1.313e-01 1.048 0.304712 Balance_of_Trade 3.191e-01 1.362e-01 2.342 0.027419 * Manufacturing_PMI 1.302e-01 1.321e-01 0.986 0.333748 Currency 1.600e-02 1.806e-01 0.089 0.930137
Real Estate	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 5.850e-17 9.126e-02 0.000 1.0000 Stock_Market 1.018e+00 1.581e-01 6.440 9.61e-07 *** Inflation_Rate 1.554e-01 1.010e-01 1.538 0.1366 Balance_of_Trade 2.049e-01 1.048e-01 1.955 0.0618 . Manufacturing_PMI 2.206e-02 1.016e-01 0.217 0.8299 Currency 6.197e-02 1.389e-01 0.446 0.6595

After getting the results, we can summarize those in the following table:

Level of statistical importance	Stock Market	Inflation Rate	Balance of Trade	PMI	Currency
$\alpha = 0.01$	All of stock sectors	Consumer Staples	- Consumer Staples - Energy	Financials	- Utilities - Healthcare
$\alpha = 0.05$		Energy	- Materials - Real Estate - Communication Services		Consumer Discretionary
$\alpha = 0.1$		- Utilities - Real Estate	- Information Technology - Healthcare	- Industrials - Utilities	- Consumer Staples - Industrials - Information Technology

From the table, we can have some significant points:

- All of the stock sectors are **strongly** affected by the stock market.
- Consumer Staples, sector that provide basic necessities such as food and clothing, **largely** depends on Inflation Rate and Balance of Trade. This is true because if an economic crisis occurs, which results in the increase of Inflation Rate and the imbalance between Export and Import, daily needs will be more expensive and lack in quantity. Therefore, it will badly damage this sector. Also, Consumer Staples somehow relates to the currency. Foreign capital tends to flow into countries that have strong governments, dynamic economies, and **stable** currencies. A nation needs a relatively stable currency to attract capital from foreign investors. Otherwise, the prospect of exchange-rate losses inflicted by currency depreciation may deter overseas investors. As a result, we can conclude that economic crisis has the most detrimental effect on Consumer Staples.
- Other sectors, that would be negatively impacted by economic recession, are Energy, Financials, Utilities and Healthcare.