

109-1 證券市場微結構

SECURITIES MARKETS MICROSTRUCTURE PRACTICE

HW2

報酬率與量

組員

資管系	B064020005	林岳賢
資管系	B064020026	李昀澍
資管系	B064020053	陳聖勳

<< 作業二：報酬率與量 >>

Select 50 firms and one year intraday 15min data, answer the following questions.

1. Daily Trading volume is a useful info for traders. Try to forecast the daily trading volume from the first half hour's transaction data as well as previous day's transaction data, using whatever info you think relevant. As always, please show descriptive Statistics first, followed by regression analysis or machine learning if you know how. Explain your findings.

我們利用程式隨機抽樣股票 50 家，公司細節舉例如下：

公司規模大小定義

- 大型公司(Big):股本 50 億以上
- 中型公司(Medium):10-50 億
- 小型公司(Small):10 億以下

```
# A tibble: 50 x 4
  code name      股本 firm_size
  <int> <chr>      <dbl> <chr>
1  2348 海悅      967835680 small
2  1323 永裕      912745540 small
3  1439 中和      920000000 small
4  1731 美吾華 1329152440 medium
5  1717 長興 12402794550 big
6  2231 為升 1219166650 medium
7  1218 泰山 4999990380 medium
8  1713 國化 1509517000 medium
9  1712 興農 4204925850 medium
10 1909 榮成 12176857270 big
# ... with 40 more rows
```

◎Step1：我們先只看單一間公司來做線性迴歸

該家股票的描述統計值如下：

```
Call:
lm(formula = x$sumVol ~ lag(x$sumVol, 1) + x$nBuy + x$nI + x$meanPr +
    x$sumVol12 + x$nBuy12 + x$nI12 + x$meanPr12)

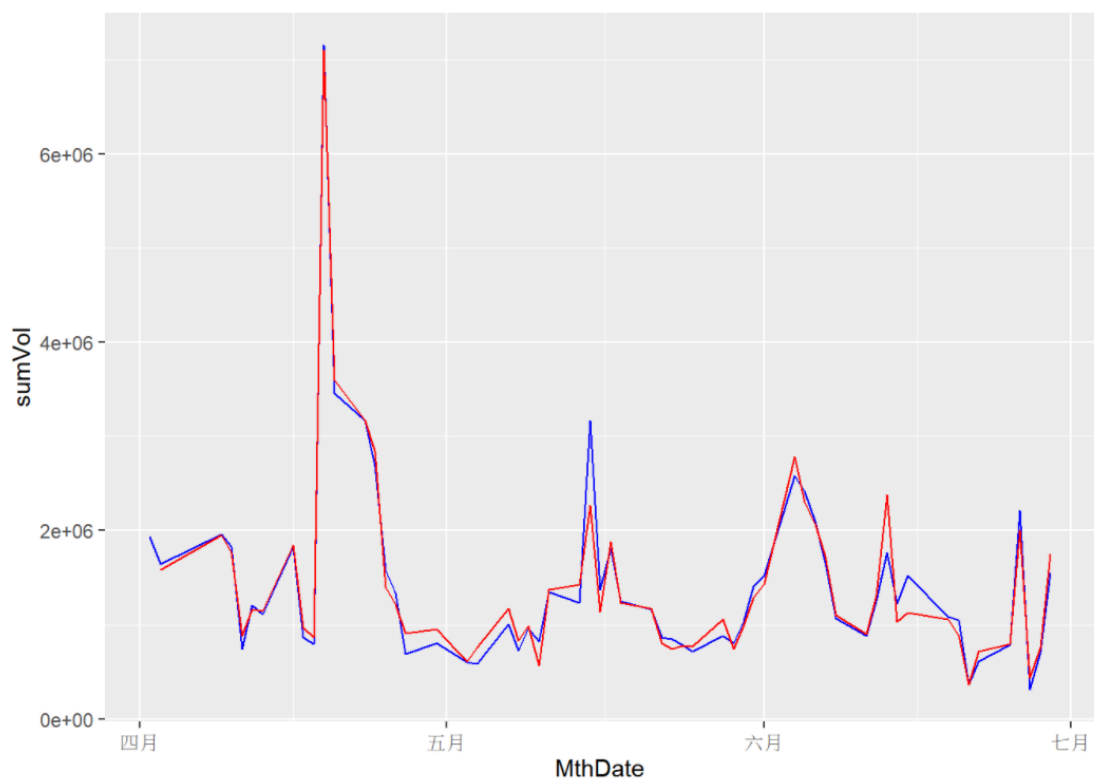
Residuals:
    Min       1Q   Median       3Q      Max
-610104 -104354   -9776    64433   903237

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.917e+06  2.527e+06   1.155  0.253707
lag(x$sumVol, 1) 2.170e-02  2.981e-02   0.728  0.469947
x$nBuy       -1.239e+02  6.723e+02  -0.184  0.854558
x$nI         3.203e+03  4.703e+02   6.811  1.19e-08 ***
x$meanPr     7.349e+05  3.663e+05   2.006  0.050261 .
x$sumVol12    1.484e+00  4.018e-01   3.694  0.000547 ***
x$nBuy12     -4.029e+02  6.184e+03  -0.065  0.948313
x$nI12       -4.755e+03  3.339e+03  -1.424  0.160628
x$meanPr12   -8.462e+05  3.729e+05  -2.269  0.027621 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 205600 on 50 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.9658,    Adjusted R-squared:  0.9603
F-statistic: 176.4 on 8 and 50 DF,  p-value: < 2.2e-16
```

→可以觀察出當日前半小時交易量的 t value 遠離 0，則表示有差異存在。

然後該股票我們畫出來的預測圖如下：



其中，紅色線為預測，藍色線為實際結果。

◎Step2：再來做全部 50 家公司的線性迴歸

全部 50 家股票的描述統計值如下：

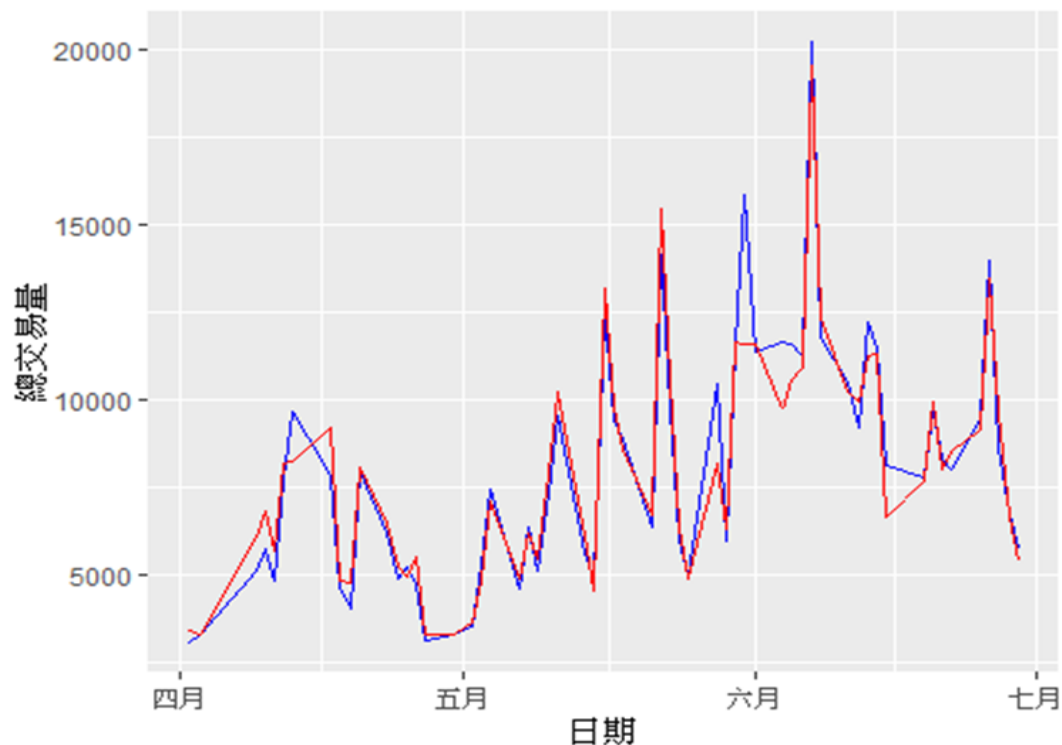
```
Call:
lm(formula = rtdf$sumVol ~ lag(rtdf$sumVol, 1) + rtdf$nBuy +
    rtdf$nI + rtdf$meanPr + rtdf$sumVol12 + rtdf$nBuy12 + rtdf$nI12 +
    rtdf$meanPr12)

Residuals:
    Min       1Q   Median       3Q      Max
-75291838  -375713   254255   589594 125329237

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.984e+05  1.228e+05  -2.430  0.01518 *
lag(rtdf$sumVol, 1)  6.125e-02  5.208e-03  11.760 < 2e-16 ***
rtdf$nBuy      4.109e+03  2.846e+02  14.441 < 2e-16 ***
rtdf$nI        3.416e+03  1.801e+02  18.966 < 2e-16 ***
rtdf$meanPr    -2.435e+04  2.220e+03 -10.966 < 2e-16 ***
rtdf$sumVol12   2.238e+00  4.175e-02  53.617 < 2e-16 ***
rtdf$nBuy12     1.836e+03  1.325e+03   1.386  0.16598
rtdf$nI12      -1.474e+04  9.305e+02 -15.840 < 2e-16 ***
rtdf$meanPr12   2.692e+04  8.386e+03   3.211  0.00134 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5905000 on 2868 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.9781,    Adjusted R-squared:  0.978
F-statistic: 1.598e+04 on 8 and 2868 DF,  p-value: < 2.2e-16
```

→一樣由上面可觀察出這 50 家股票當日前半小時交易量的 t value 非常遠離 0，則同樣表示有差異存在。



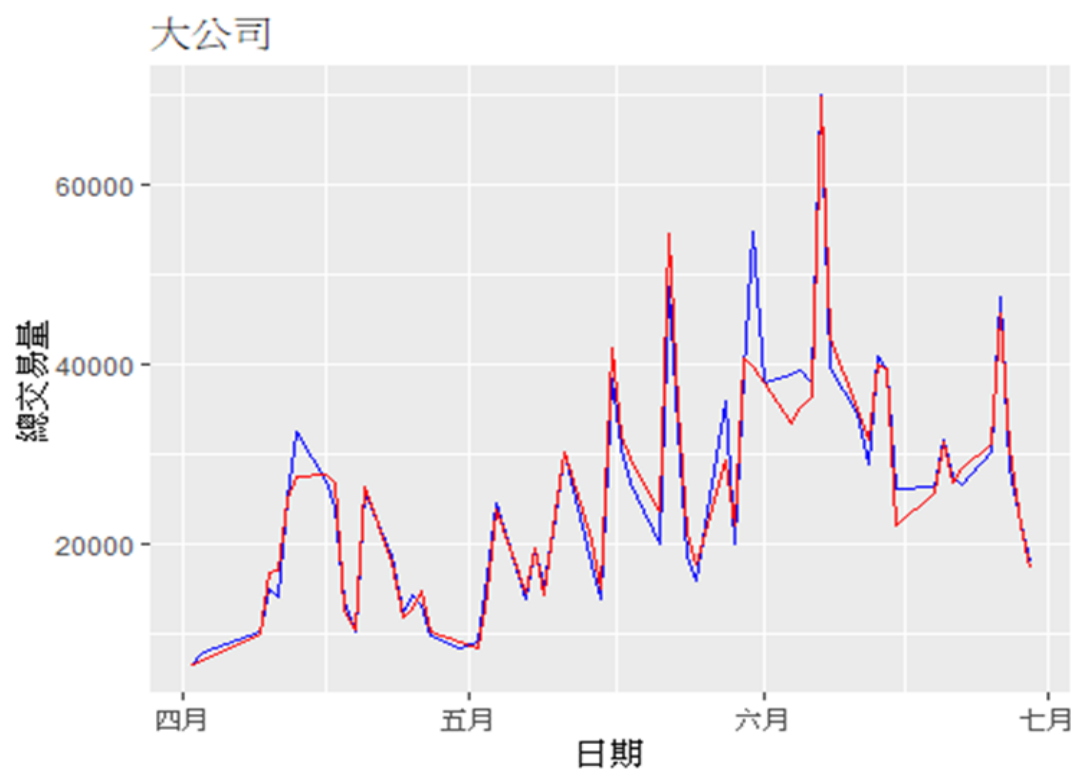
其中，紅色線為預測，藍色線為實際結果。

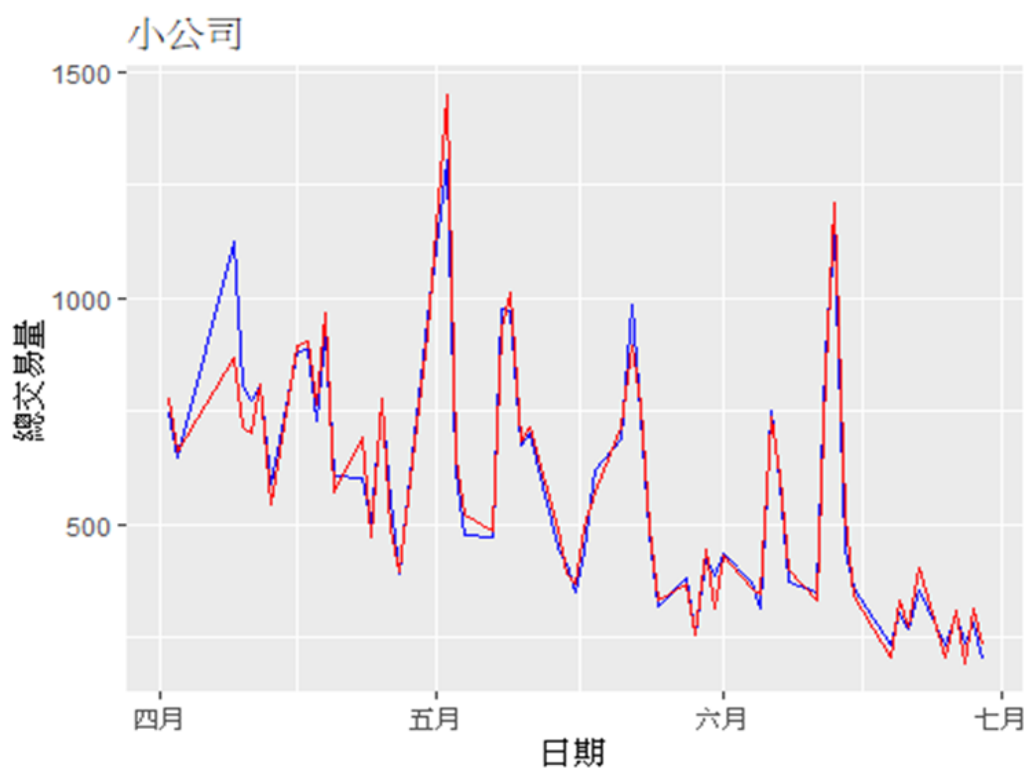
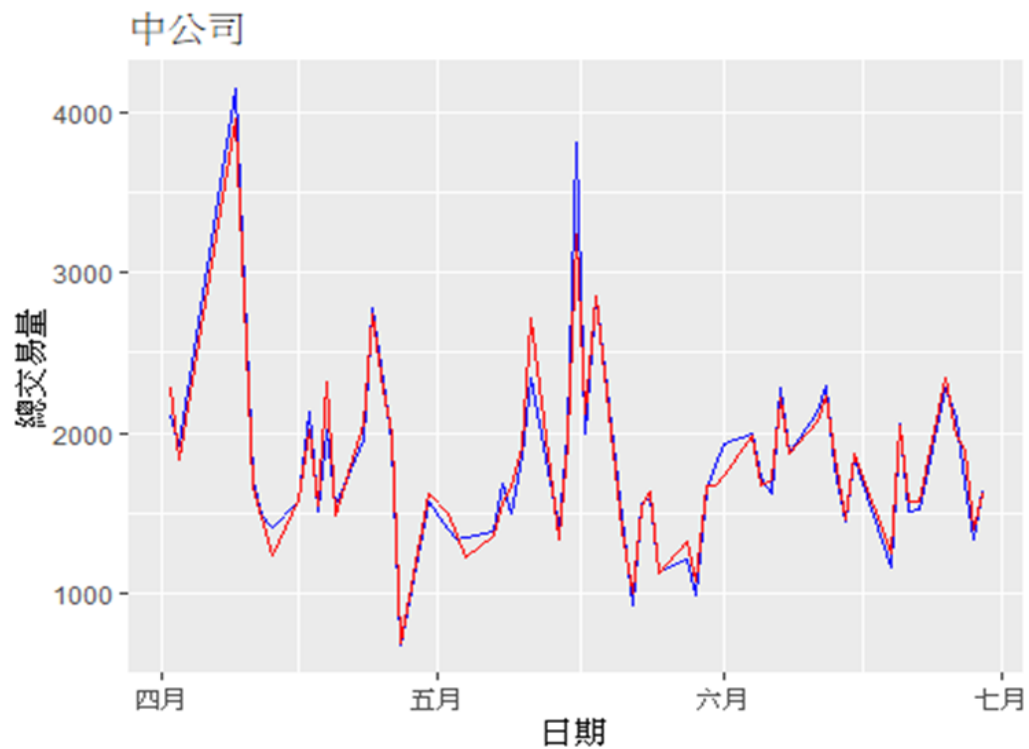
➔我們可以從中發現「前日交易量」對於當日交易量影響較小，然而真的在日內交易中「前半小時的交易量」對於「當日交易量」有相當的影響力。

2. Now, redo the above by subgroups(hint: by firm size, liquidity, volatility, market conditions, etc.).What can you observe?

從第 1 題以公司規模重新做一次，我們分成大、中、小公司 3 個 subgroups 來做測試。

3 個 subgroups 其預測與實際結果的日總交易量分別於下圖
(紅色線為預測，藍色線為實際結果)





➔然而再從描述統計值中(如以下)，我們可以發現在「中公司」的 t value 非常地遠離 0，有差異存在，表示具有相當影響力，也可對照上圖的「中公司」來看預測與實際線的貼的比較齊。

大公司

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.641e+06	4.757e+05	-7.654	5.83e-14	***
lag(brtdf\$sumVol, 1)	1.744e-02	9.211e-03	1.893	0.05867	.
brtdf\$nBuy	1.499e+03	7.377e+02	2.032	0.04247	*
brtdf\$nI	5.245e+03	4.341e+02	12.083	< 2e-16	***
brtdf\$meanPr	6.075e+04	2.067e+04	2.939	0.00339	**
brtdf\$sumVol12	1.727e+00	8.500e-02	20.314	< 2e-16	***
brtdf\$nBuy12	3.253e+03	2.738e+03	1.188	0.23517	
brtdf\$nI12	-1.219e+04	1.838e+03	-6.630	6.33e-11	***
brtdf\$meanPr12	1.475e+04	9.071e+04	0.163	0.87086	

中公司

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.526e+05	2.258e+04	6.759	1.99e-11	***
lag(mrtdf\$sumVol, 1)	2.086e-02	7.520e-03	2.774	0.00561	**
mrtdf\$nBuy	-2.775e+03	3.566e+02	-7.784	1.31e-14	***
mrtdf\$nI	4.108e+03	2.044e+02	20.103	< 2e-16	***
mrtdf\$meanPr	-4.444e+03	4.445e+02	-9.998	< 2e-16	***
mrtdf\$sumVol12	2.903e+00	5.443e-02	53.330	< 2e-16	***
mrtdf\$nBuy12	5.331e+03	1.090e+03	4.893	1.10e-06	***
mrtdf\$nI12	-1.063e+04	6.317e+02	-16.821	< 2e-16	***
mrtdf\$meanPr12	1.292e+04	1.661e+03	7.779	1.36e-14	***

小公司

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.326e+04	1.200e+04	1.105	0.269646	
lag(srtdf\$sumVol, 1)	2.842e-02	1.058e-02	2.686	0.007443	**
srtdf\$nBuy	3.087e+03	4.591e+02	6.723	4.22e-11	***
srtdf\$nI	4.471e+02	2.519e+02	1.775	0.076441	.
srtdf\$meanPr	-9.085e+02	2.486e+02	-3.654	0.000281	***
srtdf\$sumVol12	3.069e+00	1.613e-01	19.034	< 2e-16	***
srtdf\$nBuy12	-9.243e+03	1.442e+03	-6.410	2.98e-10	***
srtdf\$nI12	-1.728e+03	7.954e+02	-2.173	0.030204	*
srtdf\$meanPr12	1.710e+03	1.329e+03	1.286	0.198918	

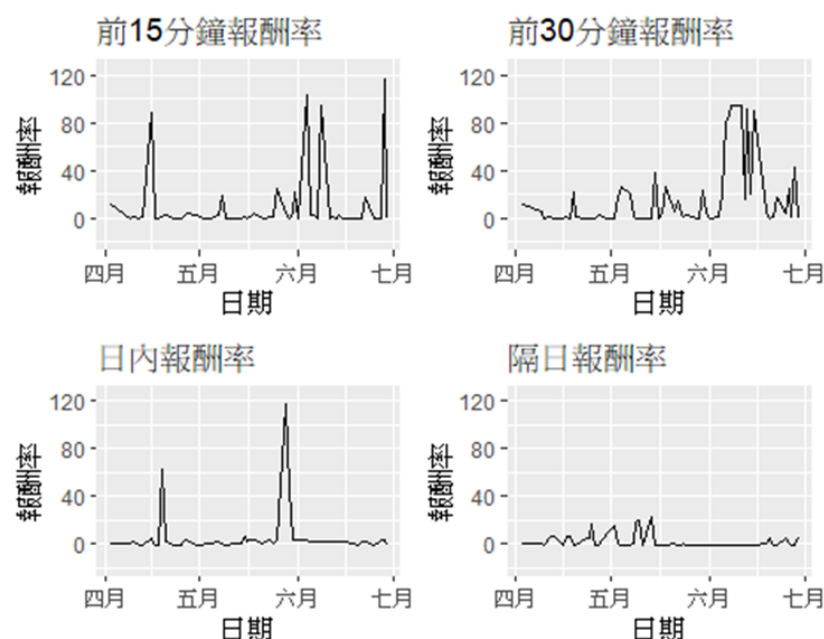
3. Calculate the overnight returns(that is, the difference between yesterday's closing price and today's opening price, in percentage terms), the first 15 minutes returns, the first 30 minutes returns, and intraday returns(close minus open price in percentage) Show the descriptive statistics of each.

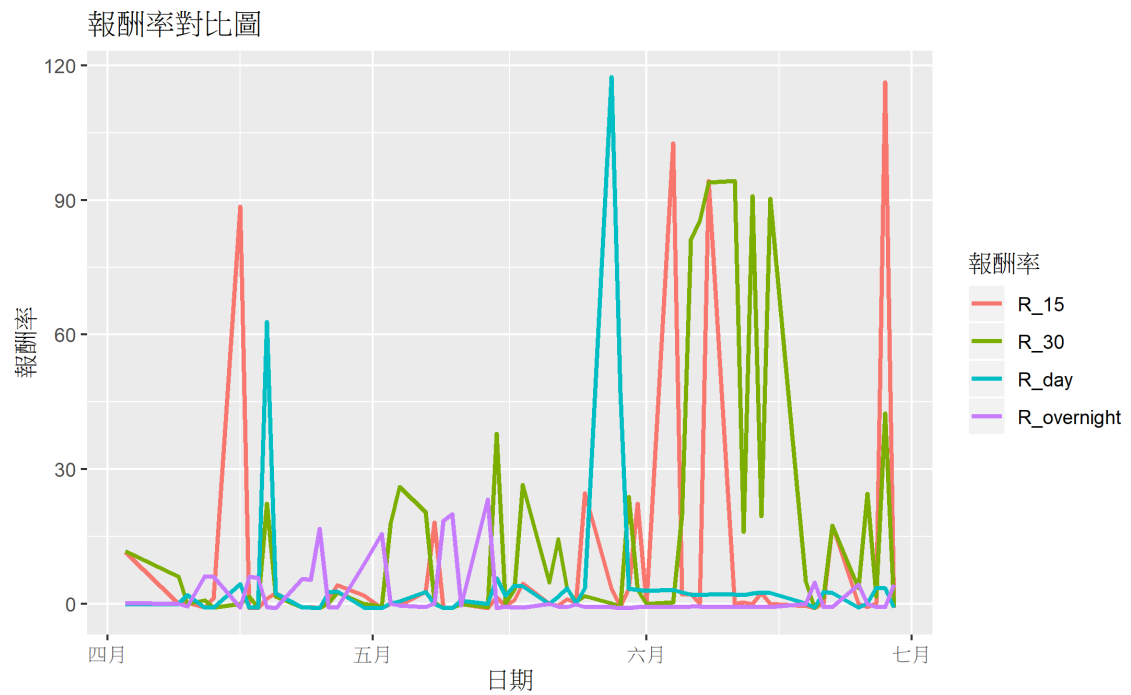
隔夜報酬計算公式
$$\frac{P_t - P_{t-1}}{P_{t-1}}$$

➔ 「隔夜報酬」、「當日前 15 分鐘報酬」、「當日前 30 分鐘報酬」以及「當日報酬」，4 個描述性統計如以下所示：

MthDate	f_first	l_last	last15	last30
Min. :2018-04-03	Min. : 6.00	Min. : 22.84	Min. : 6.45	Min. : 6.31
1st Qu.:2018-04-27	1st Qu.: 10.40	1st Qu.: 37.14	1st Qu.: 14.16	1st Qu.: 22.85
Median :2018-05-19	Median : 13.25	Median : 39.35	Median : 31.57	Median : 222.25
Mean :2018-05-19	Mean :115.75	Mean : 98.47	Mean : 143.04	Mean : 282.59
3rd Qu.:2018-06-07	3rd Qu.:169.00	3rd Qu.: 39.94	3rd Qu.: 44.74	3rd Qu.: 245.38
Max. :2018-06-29	Max. :913.00	Max. :1048.10	Max. :1220.00	Max. :1225.00
R_overnight	R_15	R_30	R_day	
Min. :-0.9934	Min. :-0.9918	Min. :-0.9919	Min. :-0.9522	
1st Qu.: -0.7483	1st Qu.: -0.4537	1st Qu.: -0.0021	1st Qu.: -0.1093	
Median :-0.6670	Median : 0.0500	Median : 2.1157	Median : 1.9955	
Mean : 2.1604	Mean : 8.9339	Mean :15.6530	Mean : 5.1969	
3rd Qu.: 3.2472	3rd Qu.: 2.3896	3rd Qu.:19.7023	3rd Qu.: 2.8697	
Max. :23.2948	Max. :116.3184	Max. :94.3307	Max. :117.4087	

4. What is the relationship between the overnight return (day t~ t-1) and the first 15, 30 minutes and intraday returns (day t)? Make innovative graphs to depict the relations and answer the questions. Explain what you observe.





➔ 以上我們可以觀察到「當日前 15 分鐘報酬」和「當日前 30 分鐘報酬」的報酬上下震幅較為激烈，其報酬則為較大，然而在「隔夜報酬」的部分上下震幅不如前二者來的大，但是其報酬相對來說也較穩定。

5. Use regression analysis to test the relation between various intraday return variables in #4(three dependent variables)and overnight returns of t-1(the main independent variable of interest). You may add other control variables such as liquidity, trading volume of t-1 and t, market return, as well as order imbalance (buy orders volume minus sell order volume in percentage term) of either t and t-1, etc. What is the most significant regression result?

```

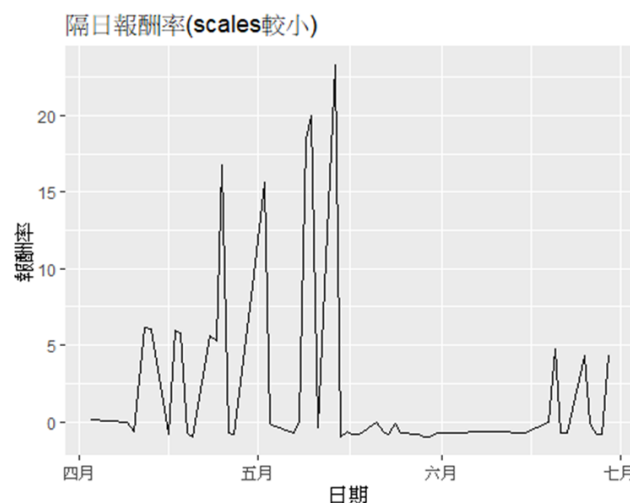
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.67342    0.88404   4.155 0.000117 ***
R_15          -0.03233    0.02856  -1.132 0.262644
R_30          -0.05843    0.02702  -2.163 0.035016 *
R_day         -0.05958    0.04028  -1.479 0.144953
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.496 on 54 degrees of freedom
Multiple R-squared:  0.1403,    Adjusted R-squared:  0.09249
F-statistic: 2.936 on 3 and 54 DF,  p-value: 0.04141

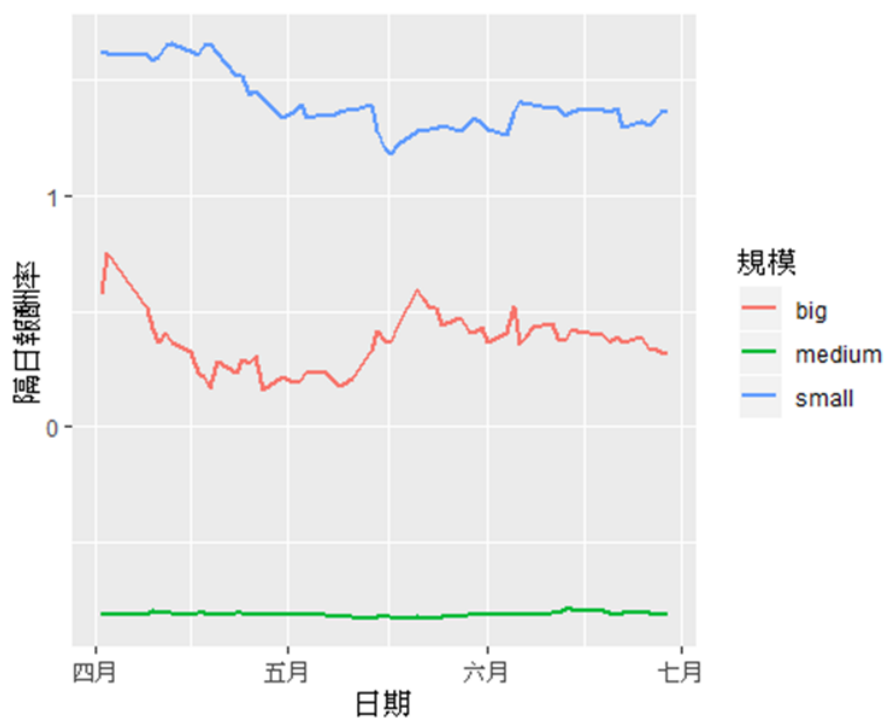
```

➔ 我們可以從以上迴歸模型中看出，在「當日前 30 分鐘報酬」這個變數在報酬方面最為顯著有影響力（因為其 Pr 最小以及 t value 值相對其他更遠離 0，表示更有存在差異）。

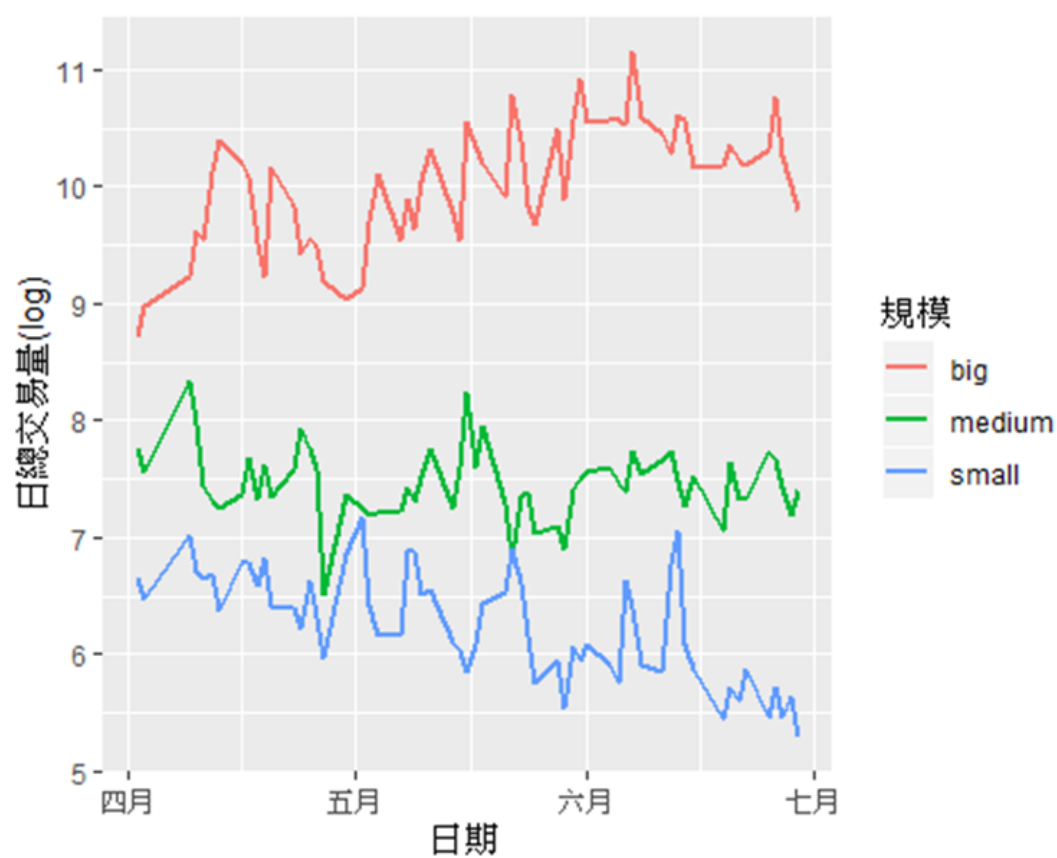
6. Do you observe any price continuation, price reversal patterns between overnight returns and next day intraday returns (15, 30 min opening interval returns and all-day returns)? Grouping your samples by turnover ratios, trading volume, firm size, etc., do you see any difference between the regression results? Explain what you find.



➔ 我們可以觀察隔日報酬率，看出四五月起伏較大，在六月時較有連續性



➔ 我們依公司規模觀察隔日報酬率，看出小公司起伏較大，中公司起伏最小



➔ 我們可以觀察日總交易量，看出大公司交易量遠遠領先於中小公司

7. Some researchers think overnight returns are proxies of stock sentiment, do you agree?

Can you think of a trading strategy from the above findings?

➔ 我們可以針對像是中公司其預測出的成交量相對來說可能較為準確，所以我們可以利用此來測試隔夜報酬或是當日前 30 分鐘報酬，或許在交易上的勝率可以有所提升。