台積電 ADR 與台灣台積電

股價報酬率之分析

迴歸分析 B 組期中報告 第一組

06155341 江祐宏

07155209 陳亮妘

07155234 蔡其浩

目錄

摘要 3
第一章 緒論 4
第一節 研究背景與動機 4
第二節 研究流程 5
第三節 資料來源與介紹變數 5
第四節 既有文獻..... 5
第二章 資料分析 6
第一節 時間序列分析 6
第二節 探索性資料分析 7
第三節 摘要五數 7
第三章 原始模型檢定 8
第一節 建立迴歸模型 8
3.1.1 參數估計 8
3.1.2 模型適合度檢定 8
3.1.3 模型解釋能力 9
第二節 原始迴歸模型殘差檢定 9
3.2.1 檢驗常態性 9
3.2.2 檢驗獨立性 10
3.2.3 檢驗變異數同質性 10
第四章 再建模型 11
第一節 離群值分析 11
第二節 探索性資料分析 12

第三節 再建迴歸模型 12
4.3.1 參數估計 13
4.3.2 模型適合度檢定 13
4.3.3 模型解釋能力 13
第四節 迴歸模型殘差檢定 14
4.4.1 檢驗常態性 14
4.4.2 檢驗獨立性 15
4.4.3 檢驗變異數同質性 15
第五章 結論 16
最終模型 16
第六章 附錄 20
參考資料 20

摘要

全球金融市場牽一髮而動全身，資訊與資訊之間，往往互相影響。若是要了解股價間波動的原因，似乎能夠從一些資訊中看出端倪。了解這些訊息，從中套利成了投資人的參考依據。本次研究以台灣台積電股價報酬率作為解釋變數，台積電 ADR 作為應變數進行簡單迴歸分析，研究結果顯示台積電 ADR 對台積電股票報酬率的影響有顯著性。

第一章 緒論

第一節 研究背景與動機

研究動機

在購買股票的時候，我們時常會考量多個因素來做決策。例如：買入估值較低的高成長性股票，或是做技術分析、籌碼分析、市場分析、資訊分析等等。而非靠直覺、感性地做決策，我們不當盲從的跟風人，走在資訊的尾端。身為財金背景的學生，我們要成為投資界的量化開拓者，靠數據說話！

台積電不只是台灣市值最大的公司（佔台灣大盤 29.49%，2021/4/28），也是全世界市值前十名的大企業，並列於蘋果、微軟、特斯拉等公司。同時，它代表著台灣人的護國神山。存股族、小資族、早餐店阿姨、自營商，任何一個身在台灣的散戶或法人無人不曉台積電帶給台灣的重要價值，因此我們選擇了台積電作為研究對象。

作為一個投資人，資訊的可信度相當重要。然而，身處在資訊爆炸的網路世代，每天充斥在新聞媒體、廣告的渲染下卻能夠輕易的影響了我們的思考與行動。舉例來說，我們時常聽到的新聞報導：「昨晚台積電 ADR 大漲 7%，想必今早的開盤會非常亮眼，OO 老師也說，現階段是個非常好的買入時機。」然而…真的是這樣子嗎？我想，股價要是這麼容易被預測出來，那我們就默默成為台灣首富了，不會在這邊打出你看見的這份迴歸報告。因此，我們決定驗證假說，透過本學期所學的簡單迴歸，用真實數據與統計方法去解釋新聞常說的「台積電 ADR 帶動台積電股價」，究竟是不是真的，又會造成多大的影響。

研究目的

我們希望透過此分析結果，了解台積電 ADR 佔台積電報酬率實際的影響權重，並透過統計數據進行深度解釋。我們期望未來聽到一件資訊時，可以更敏銳地觀察並嘗試用數據的觀點解讀它，尤其是「台積電 ADR 與台積電報酬率」相關議題。期待同學在經過這場報告之後，能夠不再受資訊左右，建立成熟健康的投資心態，並利用科學的新觀點去看待投資，相信這一場報告，能夠為班上同學們鋪墊「投資」這個一輩子課題的指引燈。

第二節 研究流程

蒐集資料 → 整理資料 → 資料分析 → 建立簡單迴歸模型 → 殘差檢定 → 結論

第三節 資料來源與介紹變數

本篇研究使用 Yahoo Finance 所提供的 python API，下載 2011 年初至 2020 年底，完整 10 年的台積電與台積電 ADR 股價資料進行分析。由於股價的漲跌是絕對報酬，因此我們選取「調整後收盤價」計算其日報酬率，並抓取出相同交易日做簡單迴歸分析，探討兩者的線性關係與對未來報酬率做預測。

以下為報酬率計算之公式：

$$R_i = \frac{Y_i - Y_{i-1}}{Y_{i-1}}$$ ，R: 報酬率，Y: 股價

反應變數 (Y)：台灣台積電每日股價報酬率

解釋變數 (X)：台積電 ADR 每日股價報酬率

第四節 既有文獻

黃營杉、李銘章(2004)「台灣母公司股價報酬與其 ADR 報酬間資訊傳遞之研究」

沈中華、邱志豪(1999)「交易成本，GDR 與股價的套利--門檻共整合應用」

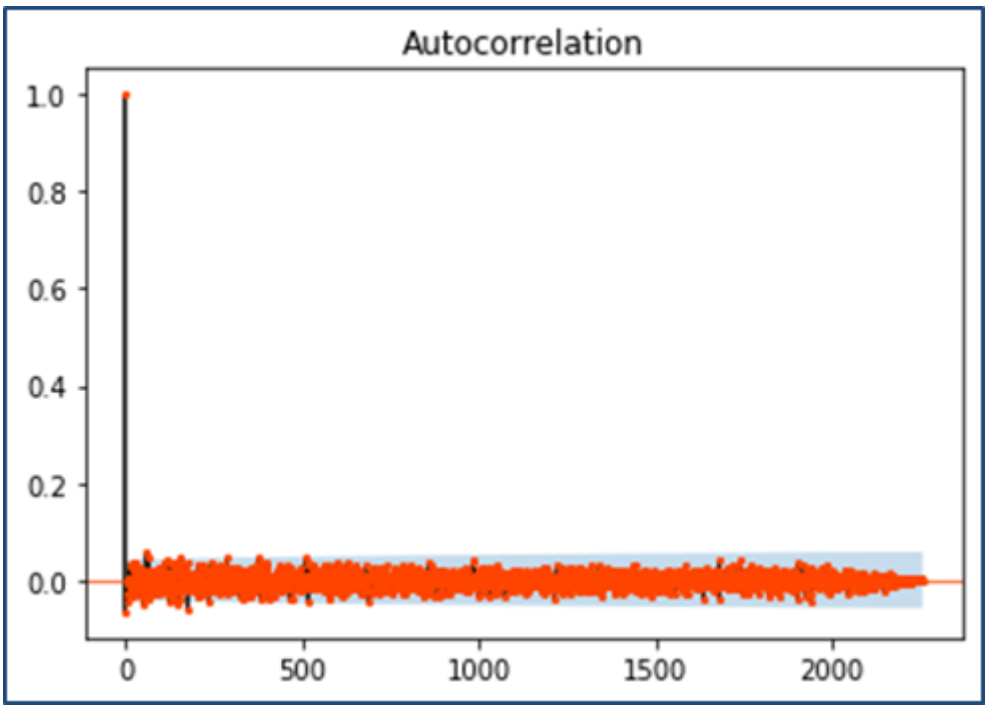
吳禮祥(2000)「美國存託憑證的套利與價差交易」

第二章 資料分析

第一節 時間序列分析

因為股價變動屬於時間序列，而研究資料是採用每日股價報酬率。在簡單迴歸模型的假設下，必須檢測解釋變數是否為時間序列。因此我們採用自我相關函數(Autocorrelation Function, ACF)來做檢測。從圖表得出解釋變數並非時間序列，符合簡單迴歸模型的假設。

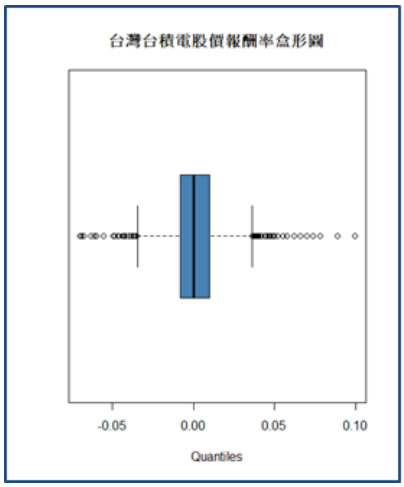
$$ACF(h) = \frac{cov(y_t, y_{t+h})}{cov(y_t, y_t)} = \frac{cov(h)}{var(y_t)} \in [-1,1]$$

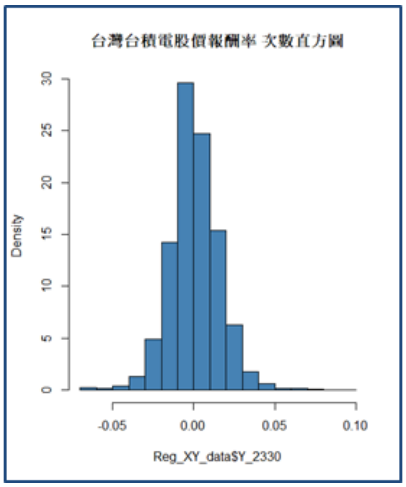


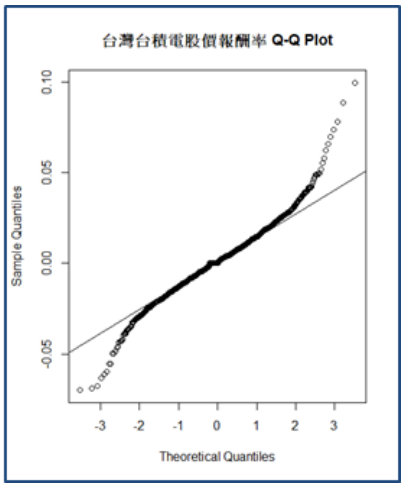
自我相關函數圖

第二節 探索性資料分析(Exploratory Data Analysis)

下列統計圖表由左至右分別為台灣台積電每日股價報酬率的盒形圖、直方圖、QQ 圖。從圖表進行探索性資料分析，盒形圖為左右對稱，有一些離群值，判定近似常態分配。直方圖呈現單峰、鐘形、左右對稱、沒有離島現象，判定近似常態分配。QQ 圖資料點分布接近 45 度線上，有些許離群值，判定近似常態分配。結論反應變數符合常態分配。







第三節 摘要五數(Five Summary)

變數	n	min	Q1	Q2(Med)	Q3	max
X (covariates) 台積電 ADR 股價報酬率	2369	-0.14034064	-0.008555066	0.0009825306	0.010684432	0.12652230
Y (response) 台灣台積電 股價報酬率	2369	-0.06992086	-0.008009115	0.0000000000	0.009732341	0.09974099

此為針對解釋變數和應變數所列出的摘要五數，樣本數共有 2369 筆。其中解釋變數為台積電 ADR 股價報酬率(每日)；反應變數是台灣台積電股價報酬率(每日)。

第三章 原始模型檢定

第一節 建立迴歸模型

```
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Y_2330   R-squared:                       0.165
Model:                            OLS   Adj. R-squared:                  0.164
Method:                 Least Squares   F-statistic:                     467.1
Date:                Fri, 23 Apr 2021   Prob (F-statistic):           1.07e-94
Time:                        12:15:55   Log-Likelihood:                 6730.8
No. Observations:                2369   AIC:                        -1.346e+04
Df Residuals:                    2367   BIC:                        -1.345e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0007      0.000      2.353      0.019       0.000       0.001
X_TSM          0.3527      0.016     21.613      0.000       0.321       0.385
==============================================================================
Omnibus:                      189.689   Durbin-Watson:                   2.540
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              872.091
Skew:                           0.235   Prob(JB):                    4.25e-190
Kurtosis:                       5.935   Cond. No.                         56.2
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
"""
```

3.1.1 參數估計

	coefficient	standard error	t	P>|t|
constant	0.0007	0.000	2.353	0.019
X_TSM	0.3527	0.016	21.613	0.000

由上表的參數估計可得迴歸模型 $\hat{y} = 0.0007 + 0.3527x$

3.1.2 模型合適度檢定

$\beta 1$ 之假設檢定

H0: $\beta 1 = 0$, H1: $\beta 1 \neq 0$ $\alpha = 0.05$

p-value < 1.07e^-94 <0.05

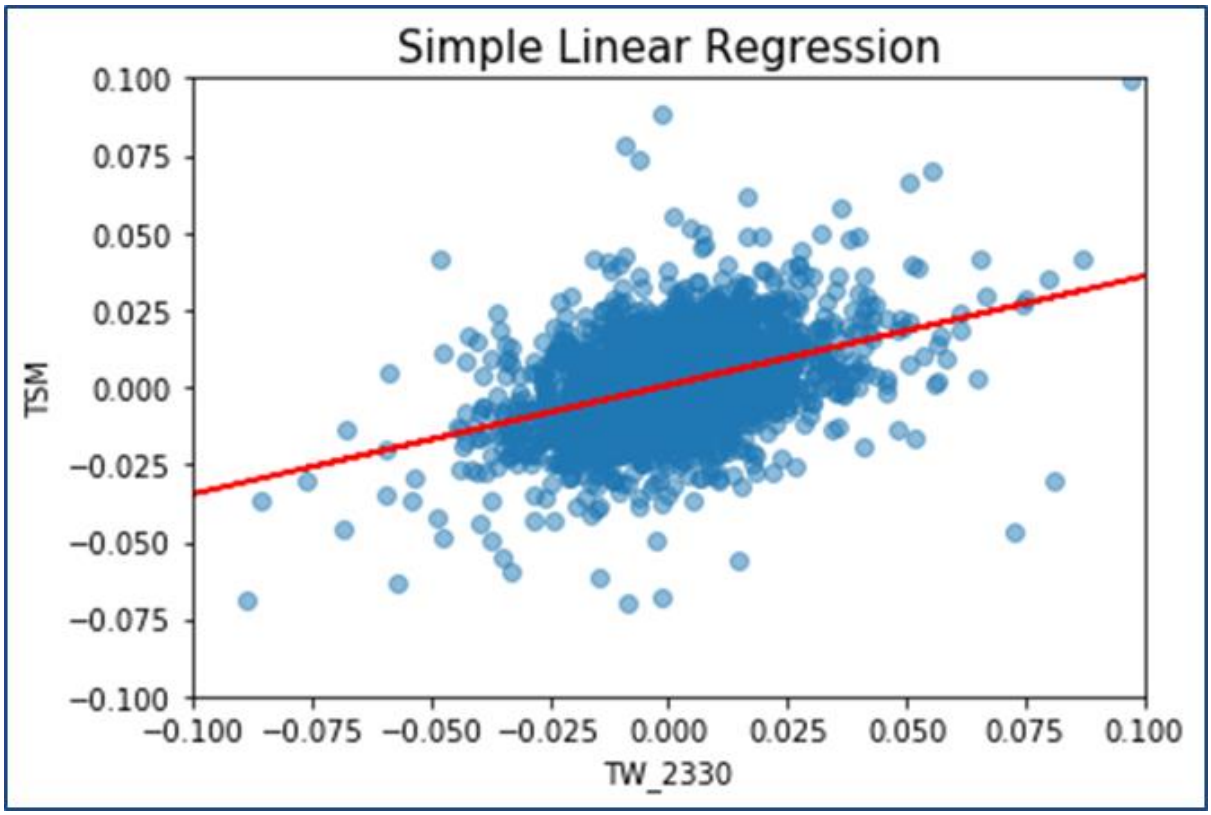
拒絕 H0，接受 H1（$\beta 1 \neq 0$）

我們驗證 x 和 y 是否存在線性關係。假設檢定中所設定的虛無假設為 1 = 0，對立假設為 1 ≠ 0，在顯著水準 α = 0.05 下，虛無假設下的拒絕域為 p-value < 0.05。從分析結果可知 p-value < 0.05，因此拒絕 H0。所以 x 和 y 存在線性關係，意思是 x 對 y 的影響是顯著的。

3.1.3 模型解釋能力

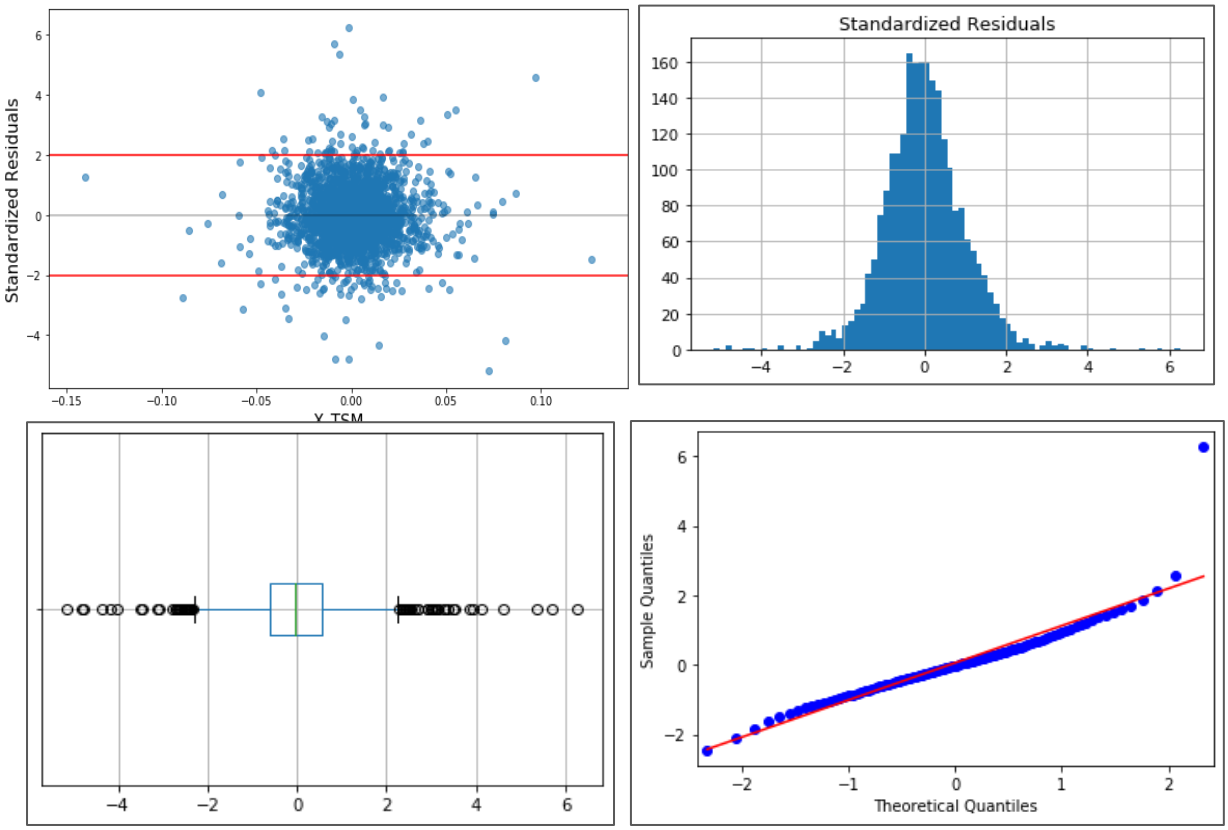
模型解釋分析表			
R-squared	0.165	Adj. R-squared	0.164

這是原始資料的迴歸模型，R-square 是 0.165，代表模型能夠解釋這組資料的比例只有 16.5%。如果單憑分析表所呈現出來的結果，是沒有辦法充分的解釋 x 和 y 的關係。但是前面做了 1 的檢定，可以看的出來 x 對 y 的關係是顯著的，所以接下來我們檢查殘差項是否滿足三大假設。



第二節 原始迴歸模型殘差檢定

3.2.1 檢驗常態性



由圖表判別常態性。從標準化後的殘差項可以發現雖然大多數都集中在 [-2,2] 之間，而且可以觀察到圖形呈現一個圓形，有少數點是在之外，但因為樣本數足夠多，因此判定近似常態分配；從直方圖和盒形圖也可以觀察到圖形幾乎是左右對稱；QQ 圖也可以看到點大多落在 45 度直線上，所以我們判定殘差符合常態性。

在檢驗獨立性和變異數同質性會常用這兩種檢驗方式。

```
> durbinWatsonTest(origin.lm)
 lag Autocorrelation  D-W Statistic
   1       -0.3353416        2.669344
Alternative hypothesis: rho != 0
> bptest(origin.lm)
    studentized Breusch-Pagan test
data:  origin.lm
BP = 0.1982, df = 1, p-value = 0.6562
```

3.2.2 檢驗獨立性

採用 Durbin-Watson 表檢驗

統計量= $\hat{e}_t = Y_t - \hat{Y}_t$ $\frac{\sum_{t=2}^{n}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{n}\hat{e}_t^2}$

α = 0.05 , K=1

Durbin-Watson 值=2.669>1.927（外插法）

→殘差項相互獨立

3.2.3 檢驗變異數同質性

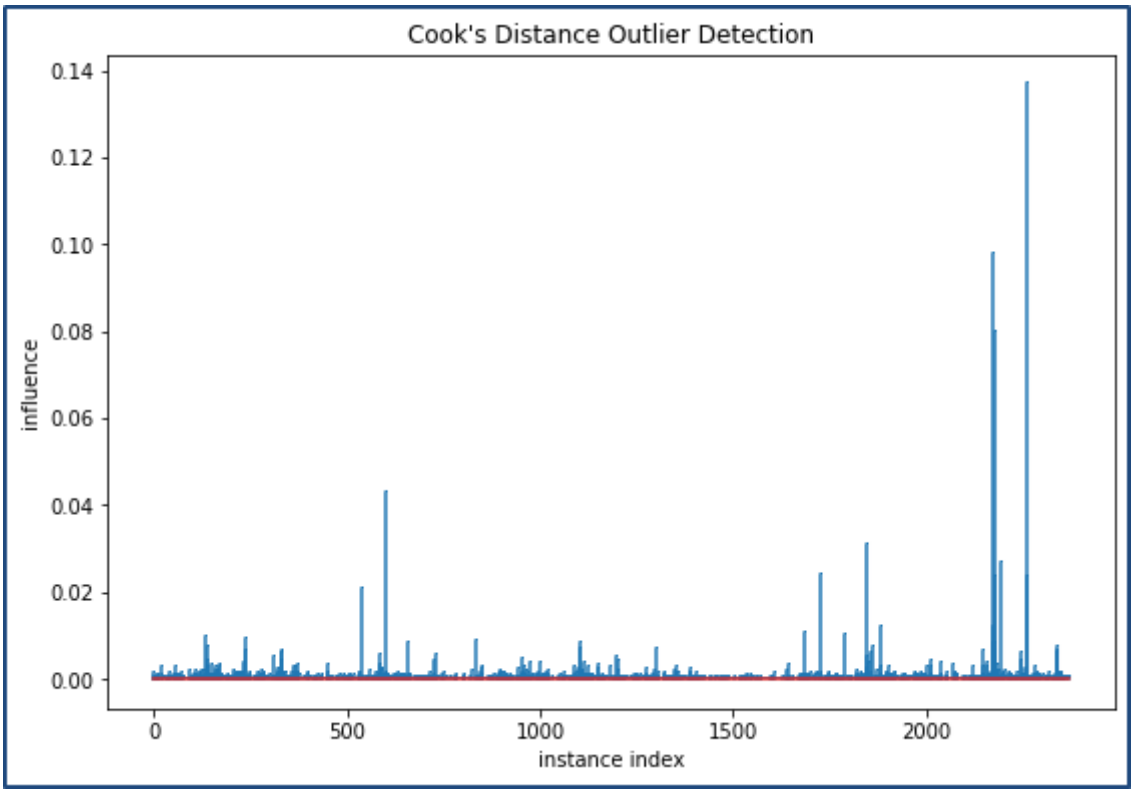
H0: 殘差項的變異數具有同質性, H1: 殘差項的變異數不具有同質性

p-value=0.6562>α=0.05，因此接受 H0，及代表殘差項的變異數具有同質性。

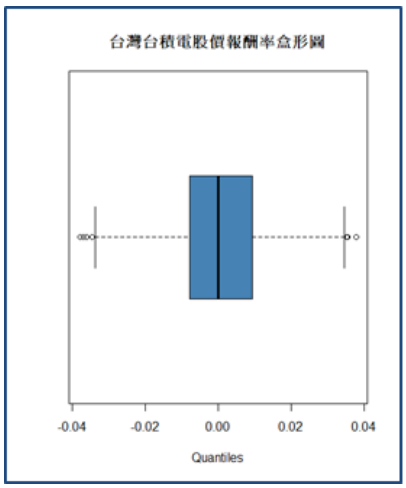
做完殘差檢定，得知常態性和變異數同質性和獨立性都有通過，意即得出的結果符合迴歸的三個假設前提，但是由於我們的解釋能力太低，所以我們認為可能有些離群值影響到 R-square，採用離群值分析 Cook's Distance。

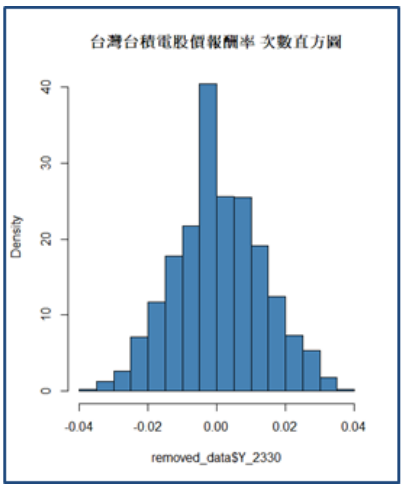
第四章 再建模型

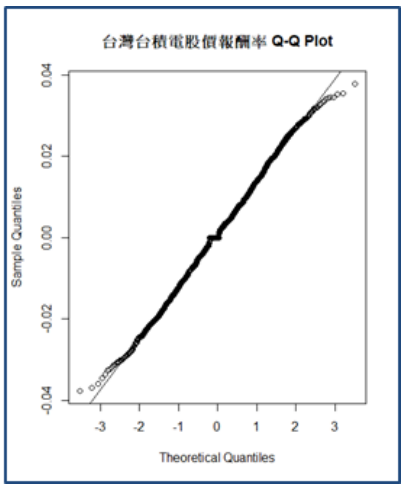
我們採用 Cook's Distance 偵測離群值，其中紅線為參考 Bollen and Jackman (1990) 所提議的 4/n(i.e. 4/100=0.04) 標準線，超過 0.04 者可能為離群值。而我們的 4/n=0.00169，共刪除 114 筆離群值。



第一節 探索性資料分析(刪除離群值後)







由刪除離群值後的統計圖表可以看到，盒形圖左右對稱直方圖單峰且更接近鐘形沒有離島現象，QQ 圖的樣本點更貼近 45 度線上，因此判定資料近似常態分配。

第二節 摘要五數(Five Summary)

變數	n	min	Q1	Q2(Med)	Q3	max
X (covariates) 台積電 ADR 股價報酬率	2255	-0.8590854	-0.007981529	0.001002136	0.010429351	0.07976143
Y (response) 台灣台積電 股價報酬率	2255	-0.03771270	-0.007874050	0.000000000	0.009317841	0.03786197

此為刪除離群值之後所列出的摘要五數，樣本數共有 2355 筆。其中解釋變數為台積電 ADR 股價報酬率(每日)；反應變數是台灣台積電股價報酬率(每日)。

第三節 迴歸模型(刪除離群值後)

```
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Y_2330   R-squared:                       0.181
Model:                            OLS   Adj. R-squared:                  0.181
Method:                 Least Squares   F-statistic:                     497.6
Date:                Fri, 23 Apr 2021   Prob (F-statistic):          9.33e-100
Time:                        12:23:18   Log-Likelihood:                 6834.3
No. Observations:                2255   AIC:                        -1.366e+04
Df Residuals:                    2253   BIC:                        -1.365e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0005      0.000      1.834      0.067   -3.14e-05       0.001
X_TSM          0.3485      0.016     22.307      0.000       0.318       0.379
==============================================================================
Omnibus:                        2.266   Durbin-Watson:                   2.439
Prob(Omnibus):                  0.322   Jarque-Bera (JB):                2.313
Skew:                           0.075   Prob(JB):                        0.315
Kurtosis:                       2.951   Cond. No.                         63.5
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
"""
```

4.3.1 參數估計

	coefficient	standard error	t	P>|t|
constant	0.0005	0.000	1.834	0.067
X_TSM	0.3485	0.016	22.307	0.000

由上表的參數估計可得迴歸模型 $\hat{y} = 0.0005 + 0.3485x$

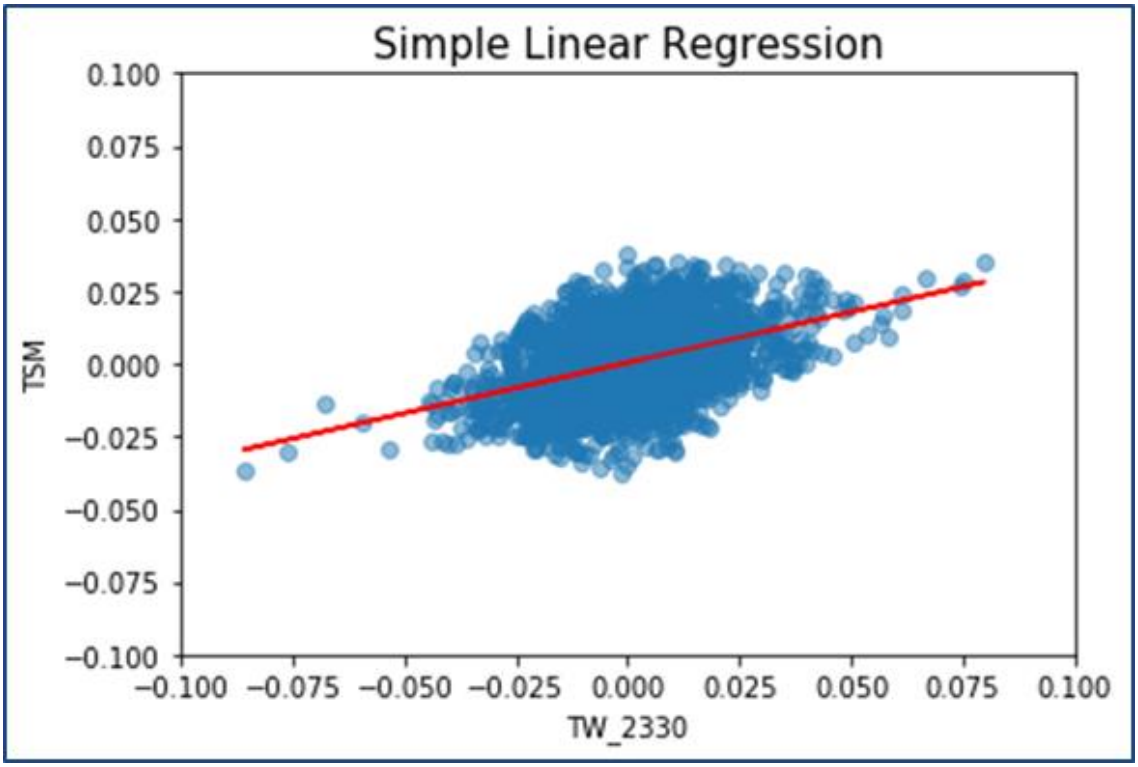
4.3.2 模型合適度檢定

$\beta 1$ 之假設檢定
H0: $\beta 1 = 0$, H1: $\beta 1 \neq 0$ $\alpha = 0.05$
p-value < 9.33e^-100 <0.05
拒絕 H0，接受 H1($\beta 1 \neq 0$)

4.3.3 模型解釋能力

模型解釋分析表			
R-squared	0.181	Adj. R-squared	0.181

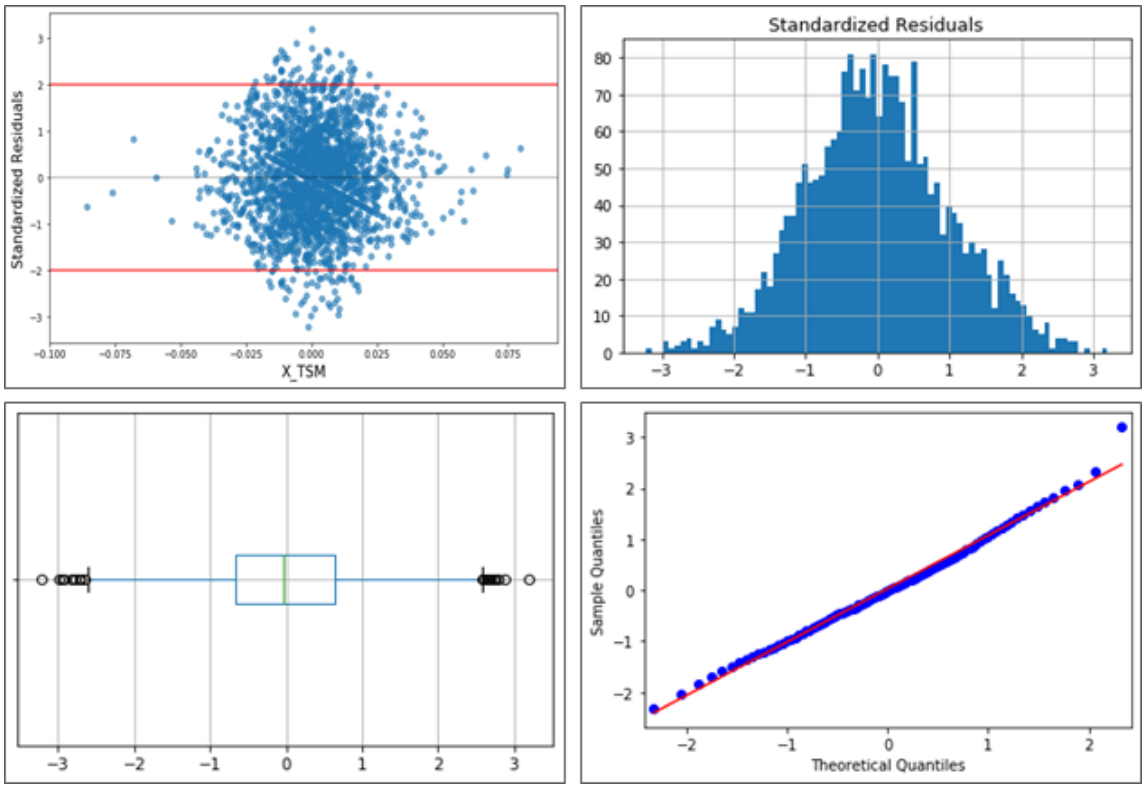
這是刪除離群值的迴歸模型，R-square 是 0.181，代表模型能夠解釋這組資料的比例有 18.1%，有比原始資料提升了一些，但是單憑分析表所呈現出來的結果，還是沒有辦法充分的解釋 x 和 y 的關係。但是前面做了 1 的檢定，可以看的出來 x 對 y 的關係能然是顯著的，所以接下來我們檢查殘差項是否滿足三大假設。



上圖為刪除離群值後的迴歸圖形，可以看得出圖形的集中度比未刪除前還要高，正相關程度也比較高。

第四節 再建迴歸模型殘差檢定

4.4.1 檢驗常態性



由圖表判別常態性。從標準化後的殘差項可以發現雖然大多數都集中在 [-2,2] 之間， 而且可以觀察到圖形呈現一個圓形，有少數點是在之外，但因為樣本數足夠多，因此判定近似常態分配；從直方圖和盒形圖也可以觀察到圖形幾乎是左右對稱；QQ 圖也可以看到點大多落在 45 度直線上，所以我們判定殘差符合常態性。

在檢驗獨立性和變異數同質性也是採取和前面相同的兩種檢驗方式。

```
> durbinWatsonTest(switched.lm)
 lag Autocorrelation D-W Statistic p-value
   1       -0.2457503       2.489683       0
 Alternative hypothesis: rho != 0

> bptest(switched.lm)
        studentized Breusch-Pagan test
data:  switched.lm
BP = 2.8396, df = 1, p-value = 0.09197
```

4.4.2 檢驗獨立性

採用 Durbin-Watson 表檢驗

統計量= $\hat{e}_t = Y_t - \hat{Y}_t$ $\quad \dfrac{\sum_{t=2}^{n}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{n}\hat{e}_t^2}$

α = 0.05 , K=1

Durbin-Watson 值=2.439>1.927（外插法）

→殘差項相互獨立

4.4.3 檢驗變異數同質性

H0: 殘差項的變異數具有同質性, H1: 殘差項的變異數不具有同質性 p-value=0.09197>α=0.05，因此接受 H0，及代表殘差項的變異數具有同質性。

第五章 結論

最終迴歸模型

$$y = 0.0005 + 0.3485x$$

在此次研究中，我們使用簡單迴歸這個統計方法來探討台積電與台積電 ADR 報酬率的關係。最終結果顯示，台積電 ADR 報酬率確實與其母公司股價報酬率有相當程度的影響。然而透過 R-squared 顯示變數的解釋能力差，代表其並非影響母股報酬率最重要的因素。因此，假如未來再聽到相關議題時，可不必過度反應，媒體人所說「有影響」確實為真，但我們不用依據此單一因素而全然改變了原本的決策。

第六章 附錄

參考資料

資料來源：https://finance.yahoo.com/
台灣台積電大幅溢價的省思：https://pse.is/3g2gbl
ADR 是什麼：https://pse.is/3dx8mn

程式碼（來源自 GitHub）：

1. 資料取得：https://pse.is/3fwhrn
2. 探索性資料分析_去除離群值前：https://pse.is/3g8nwj
3. 探索性資料分析_去除離群值後：https://pse.is/3ffplb
4. 迴歸分析：https://pse.is/3gdpcd