

期末專題報告

2025年 台北職能發展學院 Python 機器學習班

專案名稱：台灣ETF與台積電股價之關聯性分析

個人名稱：LIAO CHUN CHING



報告大綱

- ✓ 1 專題名稱與摘要
- ✓ 2 研究動機與問題定義選用理論架構模型
- ✓ 3 選用模型與方法說明數據分析流程說明
- ✓ 4 資料來源與說明結論與建議
- ✓ 5 分析結果摘要
- ✓ 6 結論與應用建議
- ✓ 7 視覺化輔助
- ✓ 8 參考文獻格式

團隊介紹

成員

- 廖純慶

1 專題名稱與摘要

專題名稱:

- 台灣ETF與台積電股價之關聯性分析 — 應用線性回歸模型進行實證探討

摘要:

本研究以台積電（2330.TW）為基準，探討其日報酬率對台灣市場主要ETF（包含0050元大台灣50ETF、0056高股息ETF、00878高股息ETF、00692富邦公司治理ETF）之影響程度。透過線性回歸分析，評估台積電報酬率變動與各ETF報酬率之聯動性，並輔以盒鬚圖、Z分數標準化與相關係數R值進行輔助統計分析，完整呈現不同ETF對台積電依賴度之差異。本研究結果可供投資人進行資產配置與風險管理參考。

[關鍵字]台積電、ETF、線性回歸、機器學習、資料視覺化

2 研究動機與問題定義

台灣股市結構中，台積電佔據極高權重，其股價變動常直接或間接影響各類ETF表現。本研究希望透過實證資料與迴歸模型，量化分析台積電對不同ETF的影響程度，並透過統計視覺化輔助判讀，協助投資人了解ETF背後潛藏的結構風險與波動來源。

3 選用模型與方法說明

1 資料收集

- 下載台積電與ETF歷史股價
- 計算日報酬率 (Daily Return)

2 資料前處理

- 缺失值處理 (刪除NaN值)
- 報酬率標準化 (Z-score, Boxplot)

3 特徵工程

- 特徵變數: 台積電報酬率
- 目標變數: ETF報酬率

4 模型訓練

- 線性回歸 (Linear Regression)

5 模型評估

- 斜率與截距 (β , α)
- 相關係數 R 值
- 視覺化 (散佈圖、趨勢線、盒鬚圖、Z分數圖)

結果呈現

✓ 3 選用模型與方法說明

分析方法：

- 監督式學習 — 線性回歸模型 (Linear Regression)

補充統計方法：

- Pearson 相關係數 R值
- 報酬率盒鬚圖 (Boxplot)
- Z分數標準化盒鬚圖 (Standardized Boxplot)

使用工具：

- Python 3.13
- 核心套件：yfinance、pandas、numpy、scikit-learn、matplotlib、seaborn、scipy

資料區間：

- 2022/01/01 ~ 2025/04/30

✓ 4 資料來源與說明

- 台積電 (2330.TW)

- ETF標的：

- 0050.TW 元大台灣50ETF

- 0056.TW 高股息ETF

- 00878.TW 高股息ETF

- 00692.TW 富邦公司治理ETF

- 資料來源：Yahoo Finance

Price Ticker	Close 2330.TW	High 2330.TW	Low 2330.TW	Open 2330.TW	Volume 2330.TW	Return
Date						
2022-01-03	591.801514	592.739392	579.609089	580.546968	69089158	NaN
2022-01-04	615.248413	615.248413	603.993869	604.931748	79434666	0.039620
2022-01-05	609.621338	627.441039	605.869822	627.441039	69593809	-0.009146
2022-01-06	603.993896	605.869654	596.490867	598.366624	53210211	-0.009231
2022-01-07	594.615051	605.869595	592.739294	603.055959	38249908	-0.015528
...
2025-04-23	873.000000	873.000000	848.000000	853.000000	37863656	0.069853
2025-04-24	864.000000	886.000000	859.000000	886.000000	32238198	-0.010309
2025-04-25	888.000000	901.000000	886.000000	900.000000	34640282	0.027778
2025-04-28	897.000000	898.000000	893.000000	898.000000	21830091	0.010135
2025-04-29	902.000000	902.000000	893.000000	897.000000	24362874	0.005574

(透過yfinance套件下載)

[801 rows x 6 columns]

- 計算方式：每日收盤價計算日報酬率

「本專題資料筆數充足，符合監督式學習分析需求」



5

分析結果摘要

其中，0050 ETF與台積電關聯性最高，00878高股息ETF對台積電依賴最低。

ETF代碼	ETF名稱	斜率 (Beta)	截距	R值 (相關係數)
0050.TW	元大台灣50ETF	0.68	-0.0001	0.941042
00692.TW	富邦公司治理ETF	0.60	-0.0001	0.660624
0056.TW	高股息ETF	0.36	0.0001	0.664182
00878.TW	高股息ETF	0.33	0.0002	0.924344

R值 (相關係數)

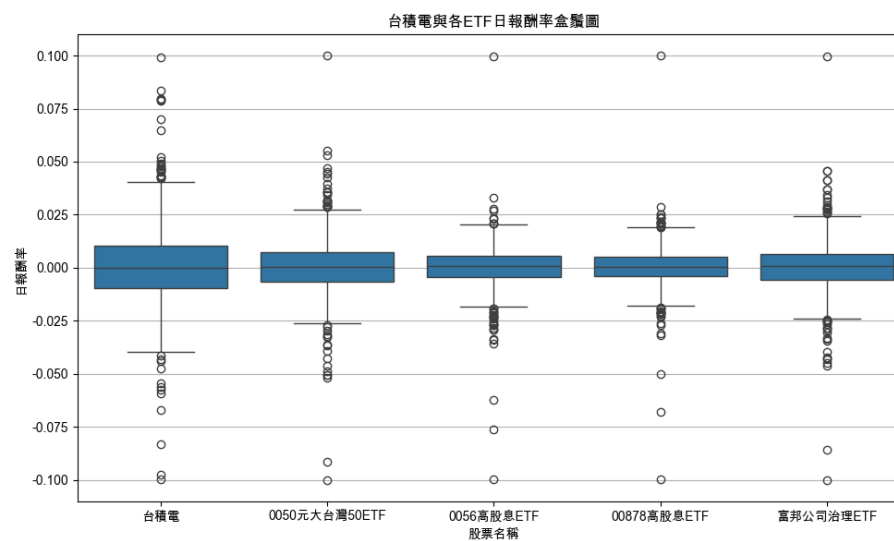
- 越接近1，代表回歸模型的解釋力越好。
- 0050.TW (0.94) 與00692.TW (0.92)：回歸效果極佳，與市場有高度關聯。
- 0056.TW與00878.TW (約0.66)：說明高股息類型的報酬與市場變動的解釋力偏低，可能受股息政策、成分股變動等影響。

**6**

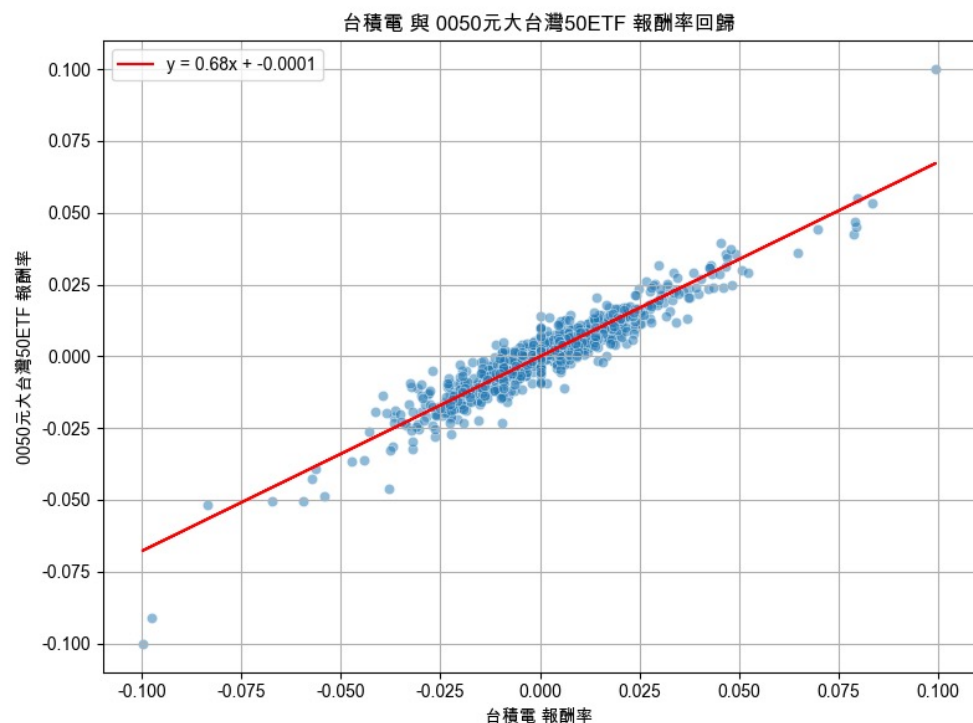
結論與應用建議

1. 台積電報酬率對各ETF表現具有明顯影響，但敏感度 (Beta值) 存在顯著差異。
 2. 以高股息為訴求之ETF（如0056與00878），對台積電依賴度顯著低於市值型ETF（如0050）。
 3. 投資人可依據自身風險承受度，參考ETF對台積電之敏感度進行資產配置。
-

- 各檔ETF與台積電回歸圖（散佈圖 + 趨勢線）
- 報酬率分布盒鬚圖
- Z分數標準化盒鬚圖



1 台積電 與 0050元大台灣50ETF



圖表說明：

- X軸：台積電報酬率
- Y軸：0050 ETF報酬率
- 趨勢線方程式： $y = 0.68x - 0.0001$

解讀重點：

•斜率：0.68

→ 代表每當台積電報酬率變動 1%，0050 ETF 會隨之變動約 0.68%。

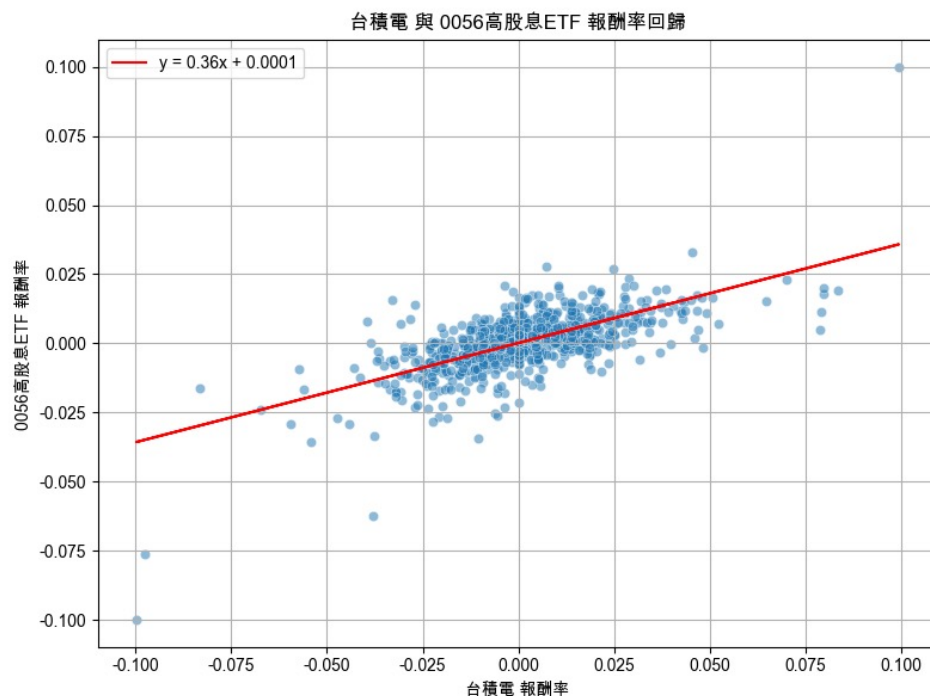
→ **相關性非常強**，因為0050成分中台積電佔比高（目前超過 40%），因此兩者高度聯動。

•截距：約 -0.0001

→ 幾乎接近零，表示在台積電報酬率為 0 時，0050 報酬率平均值也接近 0。

****散布點：** **集中且趨勢線貼近，顯示高度線性關係。

2 台積電 與 0056高股息ETF



圖表說明：

•趨勢線方程式： $y = 0.36x + 0.0001$

解讀重點：

•斜率：0.36

→ 台積電漲跌對0056影響較低，斜率明顯下降。

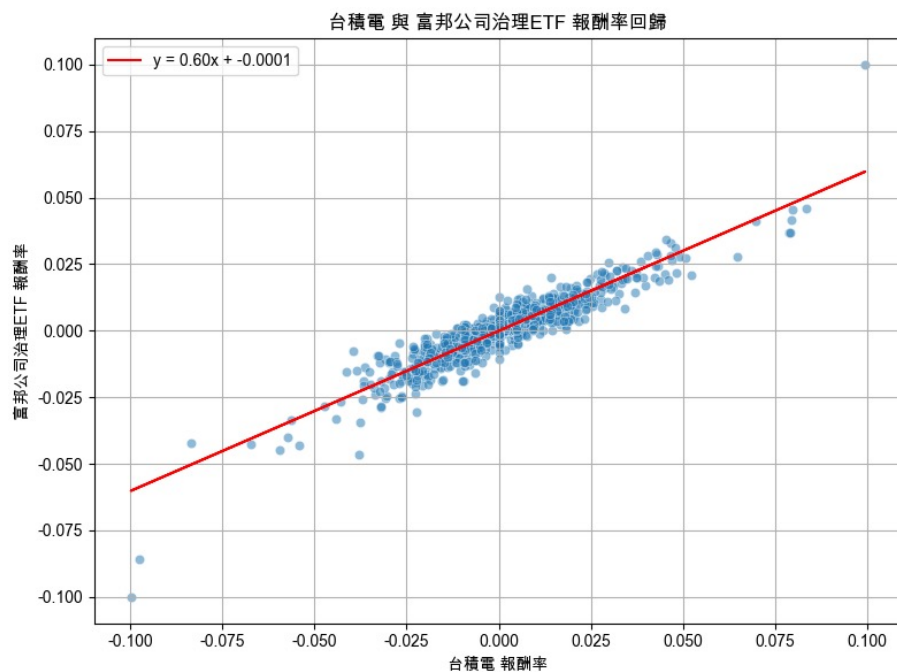
→ 因0056成分股以高股息為主，科技股比例較低，對台積電依賴程度不高。

•截距：+0.0001

→ 接近零，整體無明顯系統性偏移。

•**散布點：**較為分散，顯示與台積電相關性較弱。

3 台積電 與 00692富邦公司治理ETF



圖表說明：

•趨勢線方程式： $y = 0.60x - 0.0001$

解讀重點：

•斜率：0.60

→ 仍有不錯的相關性，但低於0050，高於0056、00878。

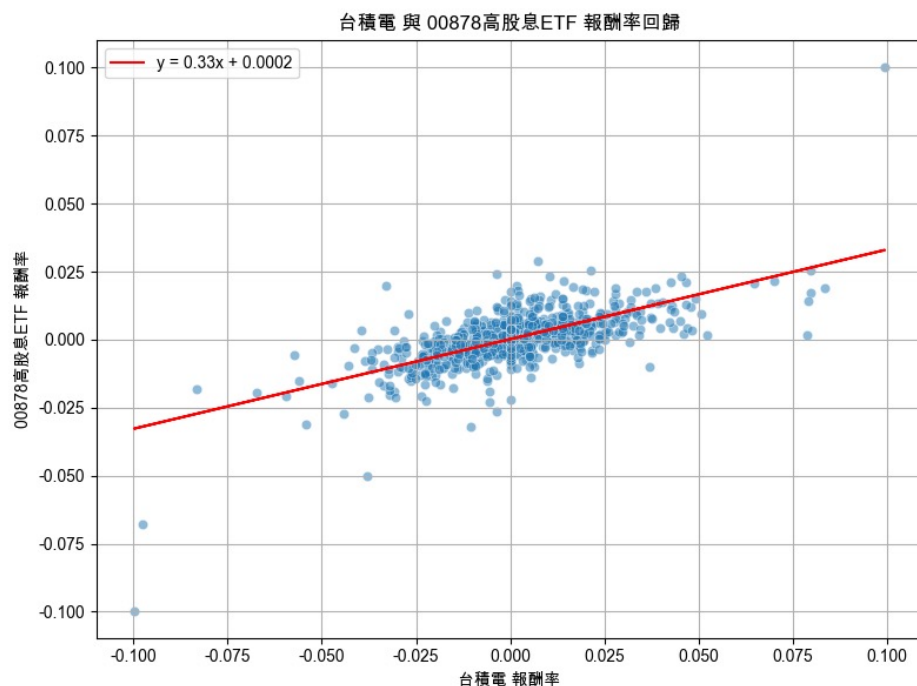
→ 00692也持有不少大型電子股，因此台積電變動對其影響仍算明顯。

•截距：-0.0001

→ 沒有系統性偏移。

•**散布點： **線性關係良好，但離散程度稍高於0050。

4 台積電 與 00878高股息ETF



圖表說明：

•趨勢線方程式： $y = 0.33x + 0.0002$

解讀重點：

•斜率：0.33

→ 與台積電的關聯度最低。

→ 00878成分股主要以穩定配息的金融、傳產、電信等為主，電子股佔比偏低，因此台積電對其**影響力較弱**。

•截距： $+0.0002$

→ 略為偏正，但幾乎可以忽略。

•**散布點： **呈現低相關、雜訊較多的特性。

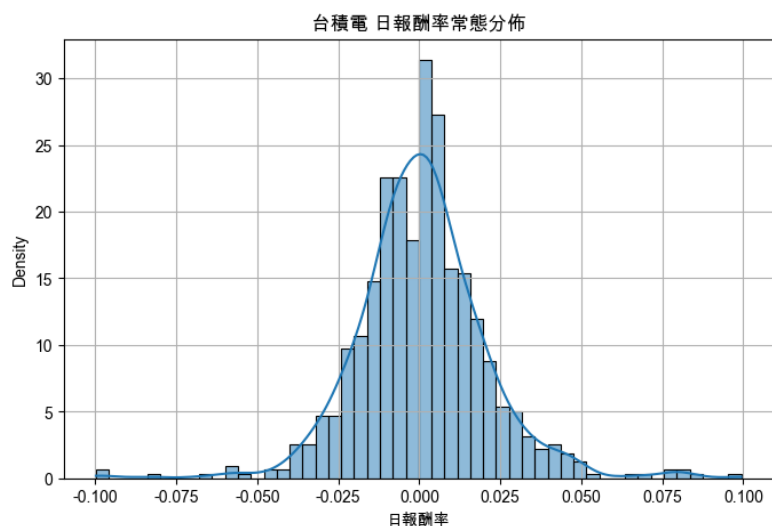
參考文獻格式

- Yahoo Finance (yfinance 套件)
 - Scikit-Learn Documentation
 - Seaborn Documentation
 - 《Python機器學習實作入門》
-

被問時演練

• NaN值說明

日期	Close	Return	
6/10	100	NaN	← 第一天沒有前一天
6/11	105	0.05	
6/12	103	-0.0190	



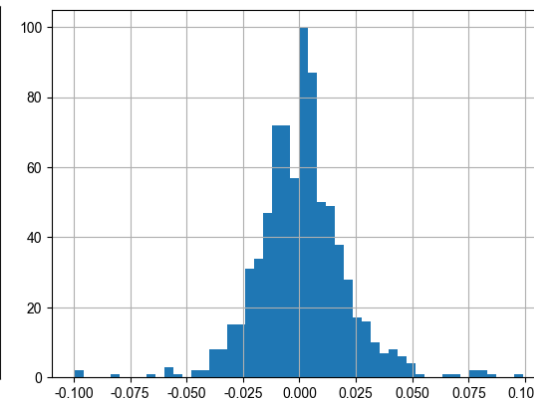
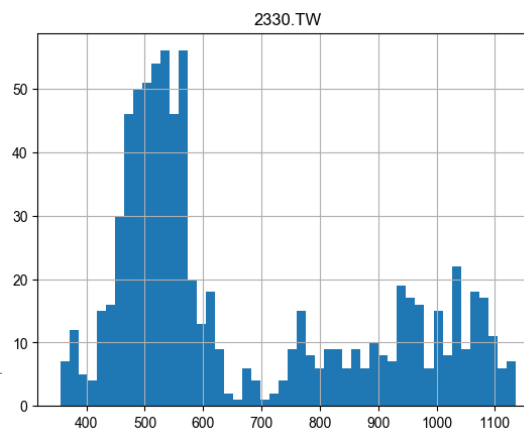
選用「日報酬率」而不是「收盤價」來產生常態分布圖？

✓ 原因一：收盤價不是「穩定分布的變數」(圖中)

- 收盤價 (Close) 是累積的結果，不斷上升或下降，沒有明確的上下邊界或平均值。
- 比如台積電十年前 100 元，今天 800 元，這樣的價格分布一定是右偏、非對稱。
- 所以畫出來的 histogram 會非常偏態 (skewed)，不適合用常態分布來分析。

✓ 原因二：報酬率才是「變動」的衡量單位(圖右)

- 報酬率才是「今天相對昨天漲跌多少」的度量，是金融分析中最重要的變異性單位。
- 報酬率常常接近常態分布（雖然有肥尾），這使得：
- 我們可以使用統計工具（Z-score、VaR、標準差等）
- 做風險分析、投資組合分析、資產配置等等





Thanks.
Happy Hacking

2024 TVDI Python AI ML
