

東吳大學巨量資料管理學院
資料科學系碩士在職專班碩士論文

指導教授：邱嘉豪 博士

運用機器學習進行營收與股利預測以輔助
股票買進決策之研究

Applying Machine Learning to Revenue and
Dividend Forecasting for Supporting Stock
Purchase Decisions

研究生：汪承翰 撰

中華民國 115 年 02 月

謝誌



論文摘要

隨著個人理財意識提升，高股息 ETF (Exchange Traded Fund) 已成為台灣投資市場的重要資產類別，總規模突破新臺幣二兆元，受益人數超過八百萬人。然而，被動式投資熱潮亦衍生出選股邏輯僵化、忽視企業價值及資金過度集中所導致的價格扭曲等結構性問題。為補足此一局限，本研究旨在建構結合財務基本面分析與機器學習技術之主動選股架構，以兼顧穩定配息與超額報酬的投資目標。研究採用系統化量化分析流程，首先利用歷年財報資料及滾動四季 (TTM) 計算方式動態預估全年獲利，並以「長線斜率」指標篩選 EPS 或淨利位於市場前 20% 的績優企業。進一步運用隨機森林 (Random Forest) 與 XGBoost 集成學習模型進行特徵工程與報酬預測，量化財務因子對股價變動的影響，最終選出預期收益最佳的 20 檔個股組成投資組合。實證期間涵蓋 2015 年 1 月 1 日至 2025 年 8 月 30 日。結果顯示，本研究構建之 AI 主動選股模型在累積報酬率與夏普值 (Sharpe Ratio) 上均顯著優於傳統市值型 (0050) 及高股息型 (0056、00878) ETF。雖受成長因子高波動與持股集中度影響，模型於 2020 年及 2025 年期間之最大回撤略高於基準指數，但於 2022 年空頭市場中展現出卓越的下行風險控管能力。進一步透過定期定額 (DCA) 模擬驗證，證實策略具備優異的淨值修復性，能有效將市場短期震盪轉化為低位累積單位數的契機。綜上所述，本研究證明結合基本面預測與機器學習的主動選股策略，於台灣股市具有高度實務可行性，不僅能為散戶投資者提供客觀且紀律化的投資框架，亦可作為現行高股息 ETF 的優化替代方案，以達成長期資產增值與穩定收益的雙重目標。

關鍵字： 高股息交易所買賣基金、機器學習、梯度提升、主動選股、基本面分析、台灣股市

Abstract

Driven by rising financial literacy, high-dividend exchange-traded funds (ETFs) have become a cornerstone of Taiwan's investment market, with assets under management (AUM) exceeding NT\$2 trillion and over 8 million investors. However, the expansion of passive investment has introduced structural challenges, including rigid stock selection logic, insufficient consideration of corporate fundamentals, and price distortions caused by capital concentration. To address these limitations, this study develops an active stock selection framework that integrates fundamental analysis with machine learning to achieve stable dividend income and alpha (excess returns). A systematic quantitative process is employed: historical financial statements and trailing twelve months (TTM) data are used to estimate annual earnings, while a long-term slope indicator identifies firms whose EPS or net profits rank within the top 20% of the market. Ensemble learning models—Random Forest and XGBoost—are then applied for feature engineering and return prediction, and the top 20 stocks with the highest expected returns are selected to construct the portfolio. The empirical analysis covers January 1, 2015, to August 30, 2025. Results show that the proposed AI-driven model significantly outperforms market-cap-weighted (0050) and high-dividend (0056, 00878) ETFs in cumulative returns and Sharpe ratios. Although portfolio concentration led to slightly higher drawdowns during 2020 and 2025, the model demonstrated superior downside risk control in the 2022 bear market. Dollar-cost averaging (DCA) simulations further confirm the strategy's strong recovery capability. Overall, the findings demonstrate that combining fundamental forecasting with machine learning provides a practical and effective alternative to conventional high-dividend ETFs in the Taiwan stock market, offering retail investors a disciplined framework for long-term capital appreciation and stable income.

Keywords: High-Dividend ETFs; Machine Learning; XGBoost; Active Stock Picking; Fundamental Analysis; Taiwan Stock Market

目錄

謝誌.....	i
論文摘要.....	ii
Abstract.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	vii
第一章 緒論.....	1
第一節 背景.....	1
第二節 研究動機.....	3
第三節 研究目的.....	5
第四節 研究問題.....	7
第二章 文獻探討.....	9
第一節 營收成長與股票獲利關係.....	9
第二節 量化投資策略研究.....	10
第三節 機器學習與投資策略.....	11
第四節 高股息 ETF 與主動式選股組合比較.....	13
第三章 研究方法.....	14
第一節 研究架構.....	14
第二節 資料處理及特徵工程.....	16
第三節 資料集劃分.....	24
第四節 模型優化與策略設計.....	25
第五節 綜合評分系統.....	37
第六節 投資組合建構.....	40
第七節 實證設計與評估指標.....	42
第四章 實證結果分析.....	46
第一節 預測模型之歷年績效驗證.....	46
第二節 成長因子的微調效應與最佳參數確立.....	49
第三節 參數優化目標之實證分析.....	52
第四節 模型結構與訓練視窗敏感度檢定.....	57
第五節 增加濾網效益分析.....	61
第六節 投資組合品質實證與選股邏輯一致性檢定.....	64
第七節 策略績效總覽與標竿比較.....	67
第五章 研究結論與未來研究方向.....	76
第一節 結論.....	76

第二節 未來研究方向	78
附錄	80
參考文獻	81



表目錄

表 1 資料蒐集表單	16
表 2 基本面因子變數定義表	21
表 3 籌碼面因子變數定義表	21
表 4 技術面因子變數定義表	22
表 5 評價面因子變數定義表	22
表 6 模型特徵值	23
表 7 資料集劃分	24
表 8 預估 2026 年 Q2 成長選股名單	40
表 9 預估 2026 年 Q3 價值選股名單	41
表 10 動態欄位定義表	44
表 11 測說明與結果	46
表 12 Q2 預測模型驗證	46
表 13 Q3 預測模型驗證	47
表 14 基礎與加權預測比較	49
表 15 歷年優化名單重疊率與替換率統計	50
表 16 參數優化目標之績效總表	54
表 17 特徵與決策邏輯驗證	59
表 18 濾網機制對年度與累積報酬之影響對比	61
表 19 股利發放與殖利率	64
表 20 投資組合年度品質統計表	65
表 21 投資組合與市場標竿之年度績效對照表	70
表 22 投資組合定期定額績效與累積損益表	72
表 23 逐年定期定額累積報酬率	73

圖目錄

圖 1 研究架構圖.....	15
圖 2 研究架構圖.....	27
圖 3 模型流程圖.....	37
圖 4 交易時間軸.....	40
圖 5 Q2 至 Q3 選股名單置換到 Top 20% 流向圖	51
圖 6 獲利最大化比較	54
圖 7 RMSE 比較.....	55
圖 8 Q2、Q3 獲利狀況	56
圖 9 有無濾網年度報酬率比較	61
圖 10 年度報酬率比較	67
圖 11 夏普值 (Sharpe Ratio) 比較	68
圖 12 最大回撤 (MDD) 比較	69
圖 13 投資組合定期定額績效與累積損益表.....	72
圖 14 定期定額累積損益比較	73

第一章 緒論

第一節 背景

在現代個人理財與資產管理中，投資被視為累積財富與創造被動收入的重要途徑。透過股票、ETF 等金融工具，投資人可在長期持有過程中分享企業成長與股息收益，進而達成財務目標。然而，實務上多數散戶投資人傾向採用被動式投資策略，例如投資元大台灣 50 (0050) 或各類高股息 ETF，以追求市場平均報酬或穩定配息收益。這類策略操作簡便，能分散風險且無需頻繁調整投資組合，因此受到廣大投資人的青睞。

然而，在這個背景下，過度依賴被動投資的投資人可能忽略企業基本面與現金流品質，僅以指數或殖利率為依據進行配置。相較於直接投資個股，投資高股息 ETF 雖然操作簡便且能分散風險，但投資人無法挑選特定公司，容易面臨報酬偏離預期、長期成長潛力不足，或被動資金流入造成的價格扭曲問題。反之，直接投資個股雖需投入更多時間與分析能力，但若能依據企業基本面與股息政策進行系統化選股，投資人不僅可能獲得超額報酬，亦可長期參與公司成長，累積資產價值與穩定被動收入。

近年來，高股息 ETF (Exchange Traded Fund) 在台灣投資市場快速崛起，其資產規模已突破新台幣 2 兆元，受益人數更上探 800 萬人，成為國內投資人主要的被動式投資標的。以元大高股息(0056)、國泰永續高股息(00878)、群益台灣精選高息(00919)等為代表的高股息 ETF，其受益人數與資產規模屢創新高，顯示高配息商品已成為散戶資金的集中焦點。表面上，這有助於提升市場參與率與理財普及，但實際上卻潛藏多項結構性風險與市場失衡問題，也與以主動選股方式獲取企業成長利益的初衷有所背離。

主動選股策略具備靈活運用資訊優勢、基本面分析及風險控管等特性。以台灣市場為例，蔡麗雀、蔡佩恂與徐川皓 (2017) 透過實證研究發現，主動篩選高殖利率且基本

面良好的個股投資組合，在累積報酬與風險調整後績效均優於市場平均水準。綜合相關研究，目前高股息 ETF 熱潮所引發的主要問題可歸納如下：

1. ETF 初衷扭曲與市場價格失真

ETF 原本設計的目的在於提供投資人分散風險、低成本與透明化的投資工具。然而，當投資人過度集中於高股息 ETF 作為主要投資標的時，其選股邏輯往往以配息率為核心，而非企業基本面，導致部分獲利能力不足或成長性有限的公司仍被納入成分股。被動資金集中流入可能推升 ETF 價格偏離基本面，增加價格波動性，扭曲市場價格機制(Ben-David, I., Franzoni, F., & Moussawi, R. 2014)。此外，被動資金流入成分股會推升個股價格，使股價短期高於企業基本面價值，進而降低殖利率，偏離 ETF 設計的高股息目標(Wang, S. M., 2008)，削弱投資者透過高股息獲取穩定收益的效果。

2. 台灣高股息 ETF 選股邏輯與制度限制

以台灣為例，高股息 ETF 的選股主要依據「未來一年預測現金股利殖利率」，例如元大台灣高股息(0056)追蹤「台灣高股息指數」，挑選殖利率最高的 50 家公司作為成分股。選股過程雖考慮市值、流動性、低本益比(P/E)、股東權益報酬率(ROE)以及營收與盈利成長潛力，但仍受市值與流動性等限制影響。

在組合權重與調整上，高股息 ETF 多採用殖利率加權，並依固定周期(半年或每季度)進行成分股調整，以確保交易穩定。部分 ETF 對成分股的資本額、成交量及換手率設有限制，並需公開成分股名單，但公開資訊也可能導致「被吃豆腐」現象。加上配息來源多元化(如資本利得或收益平準金)等操作細節的不透明，使投資人對報酬與風險的理解受限，增加策略執行的不確定性。

3. 投資人對 ETF 機制理解不足

投資人往往誤以為高股息 ETF 代表高報酬，而忽略其運作機制中潛在的複雜性與風險，包括配息來源、成分股調整、交易成本及流動性風險等。若投資人僅依殖利率或過去配息歷史判斷投資價值，容易忽略成分股的獲利穩定性、現金流品質及未來成長性，增加投資風險。此外，ETF 份額價格可能出現折溢價，與實際淨資產價值偏離，使投資人對報酬率與本金風險的理解不足。

綜上所述，台灣高股息 ETF 的被動投資熱潮雖提升投資參與度，但在選股邏輯、價格取得、流動性、報酬穩定性及市場價格發現效率上仍存在結構性缺陷。本研究期望透過主動選股結合基本面分析與現金流品質，建構兼具穩定配息與風險控制的高股息投資組合，以驗證主動策略在台灣市場的可行性，並作為改善現有被動高股息 ETF 投資結構的參考依據。

第二節 研究動機

透過長期持有投資標的，投資人可累積企業成長所帶來的資產增值與股息收益，逐步建立穩定的被動收入來源與財富基礎。尤其在個股投資中，若能依據企業基本面與股息政策進行系統化篩選，不僅可參與公司長期成長，亦有機會獲取超額報酬。

在投資理論上，效率市場假說 (Efficient Market Hypothesis, Fama, 1970) 主張，股票價格已充分反映所有可得資訊，使投資者難以持續取得超額報酬。然而，實務觀察顯示，市場仍存在各種異常現象 (market anomalies)，反映部分個股價格並未完全反映其內在價值。因此，透過基本面分析評估企業價值，投資人仍可能在資訊不完全反映的市場中找到低估或成長潛力的個股，實現長期資產累積與超額報酬的目標。

傳統價值投資策略主要依賴財務報表等基本面資料，以篩選具潛力之投資標的。然而，隨著金融市場資訊量急遽增加與資料結構日益複雜，傳統以人工分析為主的方法已難以有效處理龐大且多維的資料集。近年來，隨著運算能力提升與資料科學技術

的發展，機器學習（Machine Learning）已成為金融研究與量化投資的重要工具。該方法能捕捉資料間的非線性與高維關係，提升預測精準度並強化決策效率。

近年來，機器學習方法在股票報酬率預測與資產定價研究中受到廣泛關注。陳鼎玟（Ding-Wen Chen, 2018）透過應用機器學習與深度學習模型預測台股指數，發現這類方法在捕捉市場非線性變化與提升預測精度方面具有明顯優勢。Junsheng Wang（2022）比較了隨機森林（Random Forest）、普通最小二乘法（OLS）與 XGBoost 在股票回報預測上的表現，結果顯示隨機森林與 XGBoost 模型整體優於傳統線性模型，能有效降低預測誤差並提升投資績效。進一步地，蕭弘宗（2024）以台灣股市為實證場域，運用九種機器學習方法對上市櫃股票進行資產定價分析，並與傳統線性模型比較，透過 133 個特徵變數識別出 14 個主題性集群，揭示影響股票行為的重要特徵因素。在變數分析中，獲利/品質、低槓桿與債務發行對正回報股票表現具有顯著預測作用，而價值變數則在負回報股票中展現重要性。機器學習效能比較結果顯示，Ridge 在解釋能力和誤差指標上表現優越，神經網絡在召回率上突出，XGBoost 則在精確度上最為優異，並據此設計出穩定超越 0050 ETF 報酬的交易策略。綜合上述研究顯示，運用機器學習方法不僅可提高報酬率，也可提升資產定價模型的精準度，為投資者提供更有效的決策依據。

綜上，隨著被動式投資（如高股息 ETF）快速成長，投資市場出現資金集中與選股邏輯僵化等問題，反映出主動策略在結合現代技術後的研究與應用價值。基於此，本研究提出以下研究動機：

1. **探討在個股投資中獲取超額報酬的可能性：**現行高股息 ETF 或相關標的普遍存在資訊揭露不足、成分股篩選機制不透明、以及預估股息易受市場情緒影響等問題。基於此，本研究以個股為主要投資標的，旨在分析是否能透過企業基本面與市場特徵的量化資訊，辨識具獲利潛力的公司，以追求超額報酬。

2. **降低個股投資風險並建立分散化投資組合的機制：**由於單一個股波動度較高，本研究將建立一套以個股為單位的投資組合選取流程。透過系統化的選股與權重配置方法，在保留個股超額收益可能性的同時達成風險分散，以建構可重複驗證且穩定的投資組合。

3. **運用機器學習方法提升獲利能力與預測精準度：**機器學習具備處理高維度與非線性資料的能力，可補足傳統價值投資與線性模型的限制。本研究將評估多種機器學習模型在預測股票報酬或重要特徵上的成效，並透過模型比較與特徵重要性分析，提升投資策略的預測能力與整體績效。

第三節 研究目的

對一般散戶投資人而言，若不依賴高股息 ETF 等被動投資方式，而選擇直接投資個股，雖有機會透過長期持有累積資產與獲取超額報酬，但同時也面臨資訊蒐集困難、財務分析能力不足、篩選標的耗時費力等挑戰。散戶投資人若僅憑歷史股息、短期殖利率或市場風潮進行選股，容易忽略企業基本面、現金流品質及長期成長潛力，進而影響投資報酬與風險控制。

因此，對散戶而言，若要有效透過個股投資實現長期資產累積，需建立系統化的選股方法，結合企業財務指標（如獲利能力、資產報酬率、負債比率等）及股息政策指標（如股息分配率、現金股利穩定性、歷史殖利率等），以篩選既具穩定配息能力、又具成長潛力的個股，並配合長期持有策略，以提升報酬品質、降低投資風險。

基於此資料架構，本研究將運用多種機器學習模型（如 XGBoost、隨機森林等）建構主動式選股機制，以探討企業基本面、股息特徵與後續股價表現之關聯性，並測試模

型在不同市場環境下之預測能力。藉由系統化、紀律化且可持續更新的模型架構，本研究旨在形成一套具彈性的主動式高股息選股策略，以提升個股篩選的穩定性與客觀性。

此外，本研究將比較機器學習主動選股策略與傳統被動式高股息 ETF（如 0056、00878 等）在累積報酬率、最大回撤與夏普比率等指標上的表現差異，以評估主動策略在報酬品質與風險控制方面的優劣。透過此比較，本研究將進一步檢驗機器學習方法是否具備成為既有高股息 ETF 補充或替代方案的可行性與實務價值。

綜合而言，本研究之主要目的如下：

1. 建立系統化的個股篩選模型

本研究將系統性蒐集與整理企業基本面資料（如獲利能力、財務結構）以及與股息政策相關的量化指標，包括歷史殖利率、股息分配率、現金股利穩定性等。透過整合上述財務與股息特徵，本研究旨在全面衡量企業的配息品質與持續性，進而建立一套能夠快速且有效地篩選高潛力投資標的的系統化個股篩選模型。

2. 設計能降低風險並提升報酬的投資組合策略

有鑑於單押少數個股無法有效分散風險，本研究將建立一套穩健且可重複驗證的投資組合構建流程。透過系統性的風險分散原則、多元化的權重配置方法，以及定期再平衡機制，以有效降低個股波動對組合績效的衝擊，並提升投資組合的整體報酬與穩定性。

3. 運用機器學習模型提升投資組合獲利能力

本研究將比較多種機器學習方法（如隨機森林、XGBoost 等）或其組合，評估模型在股票報酬預測與投資組合報酬提升上的效果。同時透過特徵重要性分析，找出影響個股報酬的關鍵量化變數，以強化投資策略的解釋性與應用價值。

4. 評估機器學習主動策略是否可作為高股息 ETF 的替代方案

本研究將透過實證回測，比較機器學習主動選股策略與現行高股息 ETF 在報酬、風險、最大回撤及夏普比率等關鍵績效指標上的差異。藉由此比較，本研究旨在檢驗主動策略是否擁有長期穩定優勢，並評估其是否具有作為既有高股息投資標的之補充或替代方案的潛在可行性。

第四節 研究問題

基於前述研究動機與研究目的，本研究聚焦於探討機器學習方法結合高股息投資策略於價值投資中的應用成效，並分析其在績效表現與風險管理上的差異與優勢。為此，提出以下三項核心研究問題作為理論推導與實證分析之主軸：

1. 長期持有個股策略之超額報酬與資產累積能力

本研究將檢驗，透過財務指標篩選基本面優異且具穩定配息能力的個股，並結合營收成長預測所建構之主動投資組合，在長期持有下是否能實現超額報酬，累積資產價值，並在報酬率與風險控制層面上顯著優於被動追蹤之高股息 ETF。研究將透過歷史資料回測，評估該策略在不同市場週期下，驗證主動選股策略的可行性與穩健性。

2. 機器學習模型於投資決策之應用效益

本研究將進一步探討機器學習技術在企業基本面與股息政策資料分析中的應用價值，藉由比較 XGBoost 與隨機森林（Random Forest）等模型於短期與長期投資策略下的預測表現，評估其在預測精準度、操作適用性與模型限制等層面的差異，並分析其對散戶投資人進行主動選股的可行性與潛在優勢。

3. 投資組合精簡與績效優化之可行性

本研究亦將探討透過優化機器學習模型參數與選股機制，是否能在**精簡投資組合標的數量**的同時，仍提升整體報酬率與風險調整後績效，並有效降低交易與管理成本，以達成長期超越市場基準（如加權股價指數）與高股息 ETF 的投資目標。

上述三項研究問題將構成本研究理論探討與實證驗證的核心架構，旨在建立結合理論嚴謹性與實務可行性之高股息主動投資策略，為投資人提供兼具**穩定性、報酬性與可操作性的決策參考**。



第二章 文獻探討

第一節 營收成長與股票獲利關係

傳統價值投資理論源自 Graham(1934)所著《證券分析》，以及其後提出的「安全邊際」概念，強調投資人應以公司內在價值作為判斷標準，當股價低於內在價值時買入，以獲取長期超額報酬。巴菲特(Buffett)進一步提出「護城河」理論，主張選擇具長期競爭優勢的公司並長期持有，直到股價充分反映其內在價值。上述理論不僅奠定了現代投資分析的基礎，也強化了營收成長與股票長期表現之間的密切關係。在企業成長與績效研究中，穩健且持續的營收成長常被視為衡量企業未來獲利潛力與市場競爭力的關鍵指標。穩定的營收表現意味著企業產品或服務具備持續需求，亦反映其策略執行能力與市場地位的穩固。實證研究指出，營收成長與企業整體績效及股東報酬之間具有顯著正向關係(Graham & Dodd, 1934; McKinsey & Company, 2023; Njoroge & Nasieku, 2021)。

在台灣的企业成長與績效研究中，營收成長亦被廣泛視為預測企業盈餘與股價表現的重要變數。白翔文(2014)指出，月營收變化率較高的公司，其股票報酬明顯優於變化率較低的公司，顯示營收及盈餘相關的財務指標對股價具有預測能力；陳佑鎮(2013)進一步發現，營收與盈餘的變動對股價報酬具有顯著影響，支持營收與盈餘成長可提升企業市場評價的觀點；陳佳鳳(2001)則指出，月營收成長率對股價報酬率具有因果預測效果，進一步驗證營收成長對投資績效的正向作用。在價值投資與選股策略中，多數研究結合本益比(P/E)、股價淨值比(P/B)、股東權益報酬率(ROE)、負債比率等多項財務指標與營收成長率，設計出具實證效果的投資組合。例如，黃鈺棋(2023)發現，低P/B、ROE高且營收持續成長的公司，其

投資組合報酬率顯著優於市場平均，顯示穩定的營收成長可提升長期投資績效與報酬穩定性。

綜合上述文獻，營收成長不僅是反映企業基本面健康的重要指標，也能有效提升投資決策的精準度。尤其在成長型及新興產業中，若僅依賴歷史盈餘資料，容易低估企業潛在價值與未來成長動能。因此，將營收成長納入基本面評估架構，不僅強化價值投資策略的系統性與科學性，也為後續結合量化分析與機器學習技術提供堅實的理論與實證基礎。

第二節 量化投資策略研究

量化投資策略結合財務指標、統計方法與程式化運算，透過系統化機制制定客觀的投資決策。相較於仰賴專家主觀經驗與市場判斷的傳統投資方式，量化策略強調以數據為依據與規則化執行，能有效降低投資人因情緒波動或短期市場噪音所導致的非理性行為（Chan, Jegadeesh, & Lakonishok, 1993）

其中，營收成長被視為衡量企業經營表現與未來獲利潛力的關鍵指標。Fama 與 French（1993）指出，營收的持續成長通常代表企業具有穩固的市場競爭優勢，其財務穩健性與獲利能力亦可能反映在股價與股利的穩定提升上。此外，結合其他財務指標，如股東權益報酬率（ROE）、每股盈餘（EPS）、負債比率、本益比（P/E）及股價淨值比（P/B）等，可更全面地篩選出財務健全且具長期成長潛力的企業（Piotroski, 2000）。

在實務應用層面，量化策略透過歷史資料回測不同指標組合的績效，不僅能減少主觀判斷所帶來的風險，也可協助投資人避免因短期市場波動而做出錯誤的交易決策。此方法有助於持續且穩健地評估策略於不同市場階段的表現。隨著市場資料

規模快速增長與運算能力提升，營收成長逐漸成為基本面量化模型中備受重視的核心因子。

針對台灣市場的研究，吳幸姬與李顯儀（2006）探討「產業月營收變化與股價報酬」的關聯性，結果顯示營收變化率與股價表現之間存在顯著正向關係，驗證營收成長指標的預測價值。

綜合而言，量化投資不僅能有效降低主觀判斷與情緒干擾的風險，也能透過回測與系統化運算提升投資決策的穩定性與可重複性。將營收成長納入量化策略架構，有助於增強決策的科學性，並為後續應用機器學習建立自動化選股系統奠定堅實基礎。

第三節 機器學習與投資策略

隨著資料科學與計算能力的快速進展，機器學習（Machine Learning）逐漸成為股市預測與投資策略設計的重要工具。不同於傳統統計方法依賴線性假設，機器學習能捕捉非線性關係與變數間複雜互動，特別在高維度財務特徵資料中展現優異表現（Gu, Kelly, & Xiu, 2020）。國際研究顯示，隨機森林（Random Forest）、梯度提升樹（XGBoost）、神經網路（Neural Networks）與長短期記憶模型（LSTM）等演算法，在預測股價、營收與股利方面均展現穩定且精確的結果（Mintarya et al., 2023）。

以台灣股市為例，近年多項實證研究已驗證機器學習在投資決策中的可行性與優勢。謝曜宏、許資沄、朱涵宇與林祝興（2019）運用隨機森林模型，結合技術指標與財務指標，建立智慧型選股策略。研究結果顯示，透過特徵重要性篩選出的關鍵指標構建的策略，其投資報酬率高於採用全部指標的策略；此外，策略參數經優化後，報酬進一步提升，且整體獲利表現優於同期台灣 50（0050 ETF）。

田鉉元（2022）則以隨機森林預測市值前 300 大台股公司下一季報酬，並依特徵重要度篩選股票，再結合價值投資原則建構策略。研究結果顯示，這種方法可有效提高投資組合績效，兼顧報酬與風險控制。

此外，許雅筑（2024）運用機器學習技術，結合公司基本面與財務指標，建立台灣股票市場的選股模型。研究結果顯示，此模型在中小型股篩選上具有良好的預測準確性與穩定性，證實人工智慧技術能有效輔助量化投資策略，並為台灣股市的投資組合建構提供新的方法論依據。

本研究的設計架構中，運用最新財報資料與歷史時間序列數據，預測企業翌年之營收成長率，並根據預測結果進行初步篩選。此預測能即時反映企業最新經營狀況，為後續機器學習選股策略提供量化依據。

隨後，結合歷史基本面指標，採用 XGBoost 與隨機森林模型直接預測公司未來一年的投資報酬率。第二階段運用 Pareto 前緣進行多目標優化，平衡投資組合的預期收益與風險，確保模型於準確率與實際投資績效間達到最佳平衡。經 2018 至 2024 年第三季財報資料實證驗證，該雙階段策略能有效提升股票預測精度與投資組合績效。

綜上，機器學習在投資決策中具備三項核心優勢：其一，能有效處理高維度且複雜的財務資料；其二，能捕捉非線性關係與特徵交互作用；其三，能融合傳統基本面指標建構出精準且系統化的投資模型。儘管現有研究多集中於股價走勢預測，本研究將聚焦於營收成長與投資報酬率之間的實證關係，期望拓展機器學習於基本面選股策略的理論深度與實務應用價值。

第四節 高股息 ETF 與主動式選股組合比較

在台灣股市中，高股息 ETF（Exchange Traded Fund）長期以來被視為穩健收益型投資的主要工具。此類 ETF 通常集中於市值較大、獲利穩定且具高現金股利支付能力的企業，如金融、傳產與部分電信產業。由於其具備穩定現金流與低波動性特徵，因此成為追求穩定收益且風險承受度較低投資者的首選。

然而，高股息 ETF 的總報酬率（Total Return）通常受到限制，尤其在成長型或高科技產業比重偏低的情況下，其長期資本增值潛力相對有限。換言之，高股息策略雖能提供穩定現金流，但可能無法充分捕捉市場成長機會，對於追求資本增值的投資者存在一定限制。

相較之下，主動式選股組合結合基本面分析與量化方法，透過對企業財務結構、營收成長率、資產報酬率（ROA、ROE）等指標進行篩選，並可引入機器學習模型預測企業未來營收與股利變化，具備更高的靈活性與潛在報酬（Gu, Kelly, & Xiu, 2020；Mintarya et al., 2023）。主動式策略能針對成長型企業及高潛力標的進行布局，在有效控制風險的前提下，追求超越市場平均報酬的可能性較高。

根據林子培（2014）的研究，雖然高殖利率股票投資組合可以輕鬆打敗大盤，但要穩定超越主動式操作並不容易，顯示主動式選股在績效優化與資本增值上仍具一定優勢。總體而言，高股息 ETF 適合追求現金流穩定與低風險的投資者，而主動式選股組合則適合具備研究能力與風險承受度、希望獲取超額報酬的投資者。

第三章 研究方法

第一節 研究架構

本研究旨在建構一套結合基本面分析與機器學習的系統化選股模型，並探討其在台灣股市中長期持有個股的實證績效。本節將詳細說明研究方法、實驗設計及評估流程。

1. **資料蒐集與前處理：** 蒐集台灣上市櫃公司之歷史財報數據，涵蓋營收、淨利、每股盈餘（EPS）、股利發放及其它關鍵財務指標。
2. **基本面營收預測：** 先將累積財報還原為單季數據，利用滾動四季（TTM）總和動態預估全年獲利。篩選機制結合「長線斜率」指標，並鎖定EPS或淨利排名全市場前 20% 的績優前段班，旨在精準挖掘基本面穩健且具備高成長潛力的優質標的。
3. **機器學習模型建構：** 基於特徵工程處理後之數據集，採用隨機森林（Random Forest）與 XGBoost 等集成學習演算法進行模型訓練。透過特徵重要性分析進行量化各財務因子對股價報酬之解釋力，據此優化預測模型，精準篩選出具備最佳預期收益之投資標的。
4. **交易策略執行：** 依據模型預測結果，設定當前預測清單前 20 名做為買入清單，旨在建立兼具高成長潛力與高股息收益之投資組合。
5. **實驗設計：** 為客觀評估本研究提出之主動式選股策略績效，本研究設定以下實驗組與對照組進行比較分析：

(1) **回測期間：**2015 年 01 月 01 日至 2025 年 08 月 30 日。

(2)投資組合：透過營收預測+ AI 預測模型 + 策略組合，所建構之年度投資組合。

(3)市場 ETF：選取市場上具代表性之高股息指數股票型基金（ETF）作為績效基準（例如：高股息:0056、00878，市值型 ETF 0050、00692 等），以比較主動策略與被動式指數投資之差異。

6. 績效評估指標

最後，透過歷史回測與績效評估，比較主動選股策略與被動 ETF 在報酬率、股利穩定性等指標上的差異，驗證系統化機器學習策略在實務投資中的有效性與可行性。本研究流程旨在建立一套科學、系統化且操作性強的高股息個股投資模型，以提升選股效率、優化投資組合績效，並提供散戶投資者具實務參考價值的策略框架。

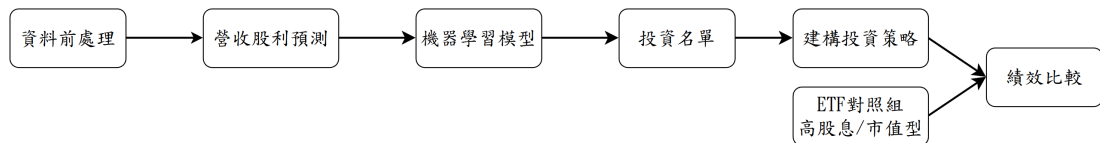


圖 1 研究架構圖

第二節 資料處理及特徵工程

一、資料蒐集

本研究以臺灣上市公司為研究對象，探討其高現金股利分配行為與股息策略之特徵與影響因素。研究選取主要上市企業為樣本，蒐集其年度財務報表資料，包含每股盈餘（EPS）、預估每股盈餘、預估淨利、實際淨利、現金股利及股利分配比率等基本財務指標。同時，輔以法人買賣動向（外資與投信）、除息日前後股價表現等市場行為變數，並以每日股價資料計算移動平均線、本益比及殖利率等技術指標。

1. 資料範圍與來源（Data Scope and Sources）本研究之選股範圍涵蓋台灣證券交易所（TWSE）所有上市掛牌公司。研究資料期間為 2015 年 1 月 1 日至 2025 年 8 月 30 日，主要數據來源包括：

- (1) 臺灣證券交易所 Open API（TWSE Open API）：獲取每日盤後交易資訊、個股收盤價、成交量及三大法人買賣超數據。
- (2) 公開資訊觀測站（MOPS）：擷取除權息日程、現金股利發放紀錄及財務報表資訊。

表 1 資料蒐集表單

資料表名稱	資料欄位
每日股價	Name：股票名稱
	TradeVolume：成交股數
	TradeValue：成交金額
	OpeningPrice：開盤價
	HighestPrice：最高價
	LowestPrice：最低價
	ClosingPrice：收盤價
	Change：漲跌價差
	Transaction：成交筆數

	TradeDate：交易日期
	UpdateTime：更新時間

資料表名稱	資料欄位
三大法人買賣超	Code：股票代碼
	Name：股票名稱
	Foreign_Investment_Trust_Buy：外資與陸資買進張數
	Foreign_Investment_Trust_Sell：外資與陸資賣出張數
	Foreign_Investment_Trust_Difference：外資與陸資買賣超張數
	Foreign_Dealers_Buy：外資券商買進張數
	Foreign_Dealers_Sell：外資券商賣出張數
	Foreign_Dealers_Difference：外資券商買賣超張數
	Investment_Trust_Buy：投信買進張數
	Investment_Trust_Sell：投信賣出張數
	Investment_Trusts_Difference：投信買賣超張數
	Dealers_Difference：自營商買賣超張數（總）
	Dealers_Buy_Own：自營商自營買進張數
	Dealers_Sell_Own：自營商自營賣出張數
	Dealers_Difference_Own：自營商自營買賣超張數
	Dealers_Hedging_Buy：自營商避險買進張數
	Dealers_Hedging_Sell：自營商避險賣出張數
	Dealers_Hedging_Difference：自營商避險買賣超張數
	Trading_Volume：成交張數

	TradeDate：交易日期
	UpdateTime：更新時間

資料表名稱	資料欄位
本益比、殖利率與股價淨值比	Code：股票代碼
	Name：股票名稱
	PEratio：本益比
	DividendYield：殖利率
	PBratio：股價淨值比
	Statisticsdate：統計日期
	UpdateTime：更新時間
綜合損益表	公司代號
	公司名稱
	營業收入
	營業成本
	營業毛利（毛損）
	營業利益（損失）
	本期淨利（淨損）
	基本每股盈餘（元）
資產負債表	公司代號
	公司名稱
	流動資產
	非流動資產
	資產總計
	流動負債
	非流動負債
	負債總計
	權益總計
	每股參考淨值

現金流量表	公司代號
	公司名稱
	營業活動之淨現金流入（流出）
	投資活動之淨現金流入（流出）
資料表名稱	資料欄位
每月營收	Code：股票代碼
	Name：公司名稱
	產業別：產業分類
	營業收入-當月營收：本月營業收入
	營業收入-上月營收：上月營業收入
	營業收入-去年當月營收：去年同期營業收入
	營業收入-上月比較增減(%)：月增率(%)
	營業收入-去年同月增減(%)：年增率(%)
	累計營業收入-當月累計營收：本年累計營業收入
	累計營業收入-去年累計營收：去年同期累計營業收入
	累計營業收入-前期比較增減(%)：累計年增率(%)
股息股利	股票代號：股票代號
	股票名稱
	除權除息日期
	除權息類型
	無償配股率
	現金股利
	股息除權基準日

二、變數處理

1. 資料結構與特徵工程原始資料經清理後，依「年度」、「公司代號」及「產業類別」進行分組建檔。本研究將歷史交易數據轉化為量化投資策略所需之技術特徵與進場訊號，關鍵變數定義如下：

(1) 評價面指標：計算「現金殖利率目標區間」與「本益比 (P/E Ratio) 相對低點」，以評估股價是否低估。

(2) 技術面指標：標註「長期均線支撐點」與「進場時機訊號」，用於輔助擇時策略。

(3) 籌碼面指標：記錄「三大法人連買天數」與「區間買超合計」，以量化籌碼動能 (Momentum) 對股價之影響，作為機器學習模型之重要特徵輸入。

2. 交易回測與績效追蹤為精確評估策略損益，本研究建立詳盡的交易紀錄資料庫。

(1) **進場紀錄**：針對每一筆模擬交易，記錄其進場基準日、成交價格、當下市值、預期殖利率及目標利差。

三、主要特徵指標

本研究將影響股價表現之因子歸納為四大構面：基本面、籌碼面、技術面與評價面。為使機器學習模型能有效捕捉市場特徵，本研究針對各構面定義了具體的量化變數，並結合「預估值」與「自比價」邏輯進行特徵工程。各構面變數定義詳述如下：

1. **基本面因子 (Fundamental Factors):**為選股模型之核心，旨在評估公司體質、獲利能力與成長潛力。本研究之特色在於引入「預估值 (Forecasted Values)」取代傳統僅依賴歷史數據之做法。

表 2 基本面因子變數定義表

類別	變數名稱	變數	定義與說明
獲利能力	每股盈餘	EPS	歷史已公告之年度與季度每股盈餘。
	預估每股盈餘	EPS_{est}	結合歷史趨勢與季度權重推估之年度 EPS
	股東權益報酬率	ROE	近一季 ROE，衡量公司運用自有資本之效率。
成長性	營收動能	Rev_Mom	包含單月營收年增率、近 3 月營收年增率。
	EPS 成長率	EPS_G	計算公式： $(EPS_{est} - EPS_{t-1})/EPS_{t-1}$ 。
股利政策	預估現金股利	Div_{est}	基於預估 EPS 與歷史平均發放率推算。
	配息穩定度	Div_{cv}	盈餘分配變異係數，數值越低代表配息越穩定。(3 年)
濾網門檻	基本獲利門檻	$Filter$	Growth 模型要求 $EPS \geq 2.0$ ；Value 模型要求 $Div \geq 2.0$ 。

2. **籌碼面因子 (Institutional Factors):**用於追蹤市場主要法人動向，判斷籌碼安定度與主力進駐跡象。

表 3 籌碼面因子變數定義表

類別	變數名稱	定義與說明
成長性	外資連買天數	外資連續買超之交易日數，用於捕捉波段買盤。
	外資買超合計	特定區間內外資買賣超張數或金額總和。
股利政策	投信連買天數	投信連續買超之交易日數，通常代表中小型股行情。
	投信買超合計	特定區間內投信買賣超總和。
濾網門檻	三大法人合計	外資、投信與自營商之買賣超總和。

3. 技術面因子 (Technical Factors): 用於捕捉價格動能 (Momentum) 與市場趨勢，並作為進出場時機之輔助判斷。

表 4 技術面因子變數定義表

類別	變數名稱	變數	定義與說明
價格動能	短中長期報酬	R_{1m}, R_{3m}, R_{6m}	分別為近 1 月、3 月、6 月之累積報酬率。
趨勢指標	長線乖離率	$Bias_{100}$	股價與年線 (MA100) 之乖離率 $\left(\frac{Price}{MA_{100}} - 1\right)$
	中期趨勢	MA_Cross	月線 (MA20) 與季線 (MA60) 之乖離程度。
波動與量能	波動率	$Volatility_{3m}$	近 3 個月股價之標準差。
	流動性濾網	$Liquidity$	預估日均成交額 ≥ 1500 萬，以規避流動性風險。

4. 評價面因子 (Valuation Factors): 本研究引入「自比價 (Self-Relative Valuation)」邏輯，透過比較個股當前估值與其歷史水準，判斷股價之相對高低點。

表 5 評價面因子變數定義表

類別	變數名稱	變數	定義與說明
本益比	歷史本益比	PE_{hist}	$Current \frac{Price}{Trailing EPS}$
	平均本益比	PE_{avg}	過去 3 至 5 年之平均本益比。
	自比價指標	PE_{Rel}	$\frac{PE_{current}}{PE_{avg}}$ ，用於判斷估值偏離程度。
殖利率	預估殖利率	$Yield_{est}$	預估 EPS*過去三年的平均股息派發率。
	目標價推算	$P_{target(7\%)}$	回推當殖利率為 7% 時之對應股價。
濾網門檻	價值門檻	$Filter$	Value 模型限制 $PE \leq 15$ 且 $Yield \in [4\%, 20\%]$

表 6 模型特徵值

Year	Code	Name	IndustryCode	EPS	預估 EPS	預估淨利	淨利
2021	1101	台泥	1	4.32	4.18	25719989	27991393
現金盈餘分配率 平均數	預估現金股利	現金股利	盈餘分配變異 係數	篩選類型	進場基準日	進場基準價	外資連買天數
0.8251	3.5811	3.5	0.0252	Q2	2020/9/15	42.7	0
外資買超合計	投信連買天數	投信買超合計	三大法人買超	結算交易日	結算收盤價	結算還原利差	還原殖利率
-3052451	0	-4582854	-6609305	2021/8/31	48.45	9.25	0.2166
7%交易日	7%價格	市值(7%)	資本利差(7%)	資本利差百分比 (7%)	現金殖利率(殖 利率)	還原殖利率(殖 利率)	本益比小於 10 交易日
2020/9/15	42.7		5.75	13.47	8.2	21.66	2020/9/23
本益比小於 10 價 格	資本利差(本益 比<10)	資本利差百分 比(本益比 <10)	現金殖利率(本 益比)	還原殖利率(本 益比)	300 日交易日	300 日價格	資本利差(300)
2021/11/15	112	-23.7	-21.16	16.07	-5.09		
資本利差百分比 (300)	最新日期	最新股價	市值(最新)	當前現金股利合 計	當前還原利差	當前還原殖利率	當前資本利差 (7%)
7.45	18.17	8.54	26.71	2020/9/15	42.7	5.75	13.47
當前還原殖利率 (7%)	當前資本利差 (本益比)	當前還原殖利 率(本益比)	當前資本利差 (300)	當前還原殖利率 (300)	return_1m	return_3m	return_6m
2025/12/4	22.95		7	-12.75	-0.2986	-12.75	-29.86
volatility_3m	Price_vs_MA100	MA20_vs_MA60	市值	Is_Signal_7pct	Is_Signal_6pct	Is_Signal_5pct	Is_Signal_PE10
-26.95	-12.75	-29.86	-0.0149942	-0.0035006	0.0990991	0.0183062	-0.020732
Is_Signal_MA300	單月營收年增率	近 3 月營收年 增率	成交量變異	近一季 ROE			
1	8.33448	0.87	0.83	28.38			

第三節 資料集劃分

研究資料涵蓋 2015 年 1 月 1 日至 2025 年 8 月 30 日。首先建立 2016 年度完整特徵與回測績效資料作為第一個測試基礎，2017 年度以 2016 年作為訓練，其後每年度以前 3 年-5 年資料作為訓練集，將最新年度作為測試集進行隔年預測。此方式模擬實務投資者「以過去經驗預測未來」的過程，有助於提升模型預測準確性。

研究之原始資料蒐集期間自 2010 年 1 月 1 日起，主要考量模型特徵中包含長期技術指標（如 300 日移動平均）以及需回溯三年以上之財務統計指標（如現金股利變異係數）。因此，2010 年至 2014 年期間僅作為特徵計算之暖機資料，未納入模型訓練與績效評估。在確保所有特徵均可於不含未來資訊的前提下完整建構後，本研究之有效樣本期間自 2015 年 1 月 1 日開始，資料涵蓋至 2025 年 8 月 30 日。其後，各年度皆以前 3 至 5 年之歷史資料作為訓練集。由於模型所需之最短訓練視窗為三年，故可執行之前推驗證之最早年度為 2016 年。

表 7 資料集劃分

資料表名稱	涵蓋區間	用途說明	筆數
訓練集	2017 - 2025	模型訓練（逐年累積方式）	2016:326
			2017:329
			2018:337
			2019:340
			2020:361
			2021:369
			2022:371
			2023:368
			2024:396
			2025:402
測試集	2018 - 2026	每年度預測下一年度表現	2026:421
全部資料	2015. 01. 02 - 2025. 04. 30	原始樣本期間	3599

第四節 模型優化與策略設計

本研究針對台灣上市股票市場應用機器學習演算法，並針對回測過程中所揭露之結構性問題，提出從參數調整至系統性策略設計之完整優化架構，以提升模型之實務適配性與投資績效。

一、動態 EPS 預測

由於財報公告存在時間滯後性 (Time Lag)，為提升下半年 (Q3, Q4) 選股之時效性與準確度，本研究提出「動態權重 EPS 預測模型」。此方法旨在解決季節性因素導致的預測失準，確保模型能持續捕捉具備顯著獲利成長潛力之標的。

1. 單季盈餘趨勢斜率

本模型不採取傳統的年增率 (YoY)，而是針對連續四個季度的單季 EPS (SQ_1, SQ_2, SQ_3, SQ_4) 進行線性迴歸分析。透過最小平方法 (OLS) 求得之迴歸係數 β_1 即為動能斜率 (Slope)，用以衡量獲利成長的連續性與強度：

$$EPS_{it} = \beta_0 + \beta_1 \cdot Time_t + \epsilon_{it}$$

當 $Time = \{0, 1, 2, 3\}$ ，當 $\beta_1 > 0$ 時，代表該企業獲利具備向上之動能。

2. 第二季 (Q2) 預測邏輯

於第二季結算點，模型整合當年度上半年 (Q1-Q2) 與前一年度下半年 (Q3-Q4) 之數據。此階段之預測邏輯側重於**獲利穩定性**，並透過「綜合百分比排名」進行標的篩選：

$$Score_{Q2} = W_{Value} \cdot Rank(EPS_{TMM}^{Q2}) + W_{Growth} \cdot Rank(\beta_{SGS}^{Q2})$$

本研究設定 $W_{Value} = 0.90$ ， $W_{Growth} = 0.10$ 。此權重分配旨在確保標的具備實質獲利支撐之基礎下，具備向上之成長慣性。

3. 第三季 (Q3) 預測邏輯：

進入第三季後，模型更新預測窗口至當年度前三季。此時，模型不僅更新 EPS_{TTM} 之基準，更強調**最新季度與歷史季度交替點的動能轉折**。針對最新序列 ($Q4_{last}, Q1, Q2, Q3$) 計算之 β_{SGS}^{Q3} ，能捕捉企業是否進入年度爆發期：

$$Score_{Q3} = 0.9 \cdot Rank(EPS_{TTM}^{Q3}) + 0.1 \cdot Rank(\beta_{SGS}^{Q3})$$

標的篩選採取滾動式分位數過濾，僅保留綜合評分位於全市場前 20% 之個股。

4. 股利分配穩定性因子：變異係數 (CV) 之應用：

除獲利預測外，本研究進一步考量股利政策之連續性。透過計算過去三載之**現金盈餘分配率平均數** (μ_{payout}) 與其變異係數 (Coefficient of Variation, CV)：

$$CV_{payout} = \frac{\sigma_{payout}}{\mu_{payout}}$$

藉此評估企業在獲利變動下維持穩定股利政策之能力。預估現金股利，計算如下：

$$Div_{est} = EPS_{TTM} \cdot \mu_{payout}$$

其中， CV_{payout} 較低者代表其股利預估可信度較高，作為後續風險溢價調整依據。

二、基於獲利最大化之參數優化

傳統機器學習文獻多以統計誤差最小化（如 RMSE, MSE）作為模型收斂目標，然而實證顯示，低預測誤差並不必然保證高投資回報。本研究在 XGBoost 超參數調整階段，採用貝葉斯優化與人工網格搜索並行比對。

1. 參數篩選差異分析：雖然貝葉斯優化能快速收斂至統計意義上的最佳解，但本研究發現，受限於金融市場的高雜訊特性，模型產出之「預測報酬率」與「實際報酬率」之間始終存在顯著誤差，難以達到精確的點預測。
2. 在此前提下，單純追求預測誤差最小化並無法保證選股排序（Ranking）的有效性。因此，本研究捨棄以誤差為基準的參數選擇標準，轉而直接選取能使「回測累積獲利」最大化之參數組合。意即，即便該組參數的統計預測誤差較大，只要其選出的股票組合能創造最高之實際報酬，即視為本研究採納之最佳模型。

```
print(f"\n [Value] 訓練中 (樣本: {len(df_train_value)})...")
model_value = XGBRegressor(n_estimators=1000,
                             max_depth=5, learning_rate=0.03,
                             subsample=0.8, colsample_bytree=0.8,
                             reg_alpha=1, reg_lambda=1, gamma=0.2,
                             random_state=42,
                             n_jobs=-1)
model_value.fit(df_train_value[FEATURE_COLUMNS], df_train_value[TARGET_COL])
```

圖 2 研究架構圖

三、訓練視窗與資料長度最佳化

於機器學習應用於時間序列數據時，訓練集長度（Training Horizon）的選擇存在一權衡關係（Trade-off）：資料過短容易導致樣本不足而過度擬合（Overfitting）；資料過長則可能納入已失效的歷史結構（Structural Break），導

致預測滯後。本研究針對成長 (Growth) 與價值 (Value) 雙軌模型，分別進行了 2 年、3 年與 5 年之滑動視窗 (Rolling Window) 敏感度測試。實證結果顯示，兩類模型因因子特性不同，其最佳訓練長度存在顯著差異。

1. 訓練視窗與資料長度最佳敏感度測試設計：本研究設定三種不同長度之回溯期間 (Look-back Period) 進行模型訓練成效比對：

(1) 短期視窗 (2 年)：側重捕捉極短線之市場情緒，但雜訊比 (Noise-to-Signal Ratio) 較高。

(2) 中期視窗 (3 年)：涵蓋單一波段之完整擴張期。

(3) 長期視窗 (5 年)：涵蓋一個完整的小型景氣循環 (Business Cycle)，包含擴張與收縮段。

2. 成長模組最佳化：採用 3 年訓練集：決策依據：成長股之股價驅動力主要來自於「短期營收爆發力」與「近期市場趨勢 (Recent Trends)」。

(1) 測試發現：當採用 5 年資料訓練成長模組時，由於納入過多早期數據，導致模型對近期營收動能 (Momentum) 的反應變得遲鈍 (Lagging)，預測準確度反而下降。

(2) 學理支持：成長型企業之經營環境變化快速，過往 (如 4-5 年前) 之財務特徵對現況之解釋力較低 (即資料存在 Non-stationarity)。

(3) 結論：設定 3 年 (36 個月) 為最佳訓練視窗，以在「樣本充足度」與「趨勢敏感度」間取得平衡。

3. 價值模組最佳化：價值股之投資邏輯建立於「均值回歸 (Mean Reversion)」與「股利穩定性 (Dividend Stability)」。

- (1) 測試發現：若僅使用 2 年或 3 年數據訓練價值模組，容易因處於景氣循環之單一階段（如多頭末升段）而誤判估值位階，導致選入在高點之標的（即落入 Value Trap）。
- (2) 學理支持：價值因子（如 PE, Yield）需要跨越一個完整的景氣循環才能展現其統計顯著性。較長的訓練集能讓模型學習到股價在不同位階下的回歸行為。
- (3) 結論：設定 5 年（60 個月）為最佳訓練視窗，確保模型能捕捉完整的估值循環特徵。

四、成長與價值風格分離機制

針對單一機器學習模型在引入價格動能（Momentum）因子後，易產生特徵主導（Feature Dominance）現象，導致投資組合過度集中於高估值之熱門題材（如特定期間之 AI 概念股），進而產生風險過度集中之偏誤。為解決此一問題，本研究建構「成長（Growth）與價值（Value）雙軌模組架構（Dual-Track Architecture）」，以實現投資風格之有效分散。

本研究發現，若對現行模型特徵進行調整，同時學習成長與價值，模型往往會在權重分配上產生衝突，導致最終預測結果趨於平庸（Mediocrity）。因此，本研究將選股邏輯拆分為獨立運作之雙模組：

1. **成長模組（Growth Module）**：以營收動能與盈餘成長率為核心特徵，並將價格動能視為輔助因子，旨在捕捉具備基本面支撐的強勢股。
2. **價值模組（Value Module）**：以估值分數（Valuation Score）為核心決策依據，但引入動能分數作為「有效性濾網（Validity Filter）」。意即，僅挑選具備初步止跌回穩或反彈跡象的低估值個股。

3. **策略優勢：**此雙軌機制有效規避了單一模型在處理衝突特徵時的預測失準風險，解決了傳統量化策略常面臨的兩難局面：

(1) 避免價值股陷入「低波動陷阱 (Value Trap)」(即股價便宜但缺乏上漲催化劑)。

(2) 防止成長股面臨「估值過熱 (Overvaluation)」(即基本面成長已無法支撐過高股價)。

五、風格因子強化

為強化雙軌模組之策略顯著性並優化選股準確度，本研究針對成長與價值兩類風格，分別設定了因子過濾門檻。此機制旨在進一步剔除雖經模型預測評分較高，但缺乏實質基本面支撐或面臨流動性風險之標的，確保最終投資組合之品質。

1. **成長股強化：盈餘品質與續航力**，成長模組之核心目標在於捕捉具備顯著獲利爆發力之個股。為驗證成長動能之可持續性並排除短期非經常性損益之干擾，本研究設定雙重檢驗門檻：

(1) **獲利成長門檻 (g_{EPS})**：設定年度預估 EPS 成長率需大於 20%。此標準用於確保入選標的之獲利動能顯著優於市場平均水準，具備高成長特徵。

(2) **營收動能門檻 ($g_{Rev,3m}$)**：設定近三個月累計營收年增率需大於 15%。營收為獲利之先行指標，此門檻用於驗證企業之成長趨勢並非短期波動，而是具備營收實質支撐。

2. **價值股強化**，獲利護城河與殖利率支撐，相較於成長股追求爆發力，價值股模組側重於資產的安全性與收益性。本研究導入智慧評分機制，透過「獲利位階」與「股息保護」雙因子，確保入選標的兼具基本面支撐與高防禦屬性：

(1) **評分邏輯**：設定全市場前 5% 為滿分（25 分），隨排名下降遞減至前 50% 歸零，此機制能有效識別營運績效頂尖的「資優生」（如聯電、瑞鼎）。即便其本益比稍高，透過高獲利排名的加分補償，模型仍會將其納入投資組合，確保核心持股具備強大的獲利護城河。

(2) **殖利率加分 (*Yield Bonus, S_{Yield}*)**：以殖利率 7.5% 為計分甜蜜點 (Sweet Spot)，越接近此數值加分越高；同時設定 5% 為基礎門檻，凡高於此門檻者皆給予額外保底加分，透過殖利率加分機制，一方面篩選出具備高現金流回報之標的，另一方面在市場修正時，利用高股息作為股價之下檔保護。

六、均值回歸與非線性懲罰機制

金融市場存在顯著的均值回歸 (Mean Reversion) 特性，即價格或估值在偏離歷史平均過遠後，有極高機率向均值收斂。為解決成長股面臨之估值過高 (Overvaluation) 風險，以及價值股可能落入之價值陷阱 (Value Trap)，本研究設計了一套非線性懲罰機制 (Non-linear Penalty Mechanism)，針對偏離常態之個股進行分數懲罰。

1. **成長股調整：歷史相對估值法 (Historical Relative Valuation)**，針對成長型個股，本研究捨棄單一絕對本益比標準，改採「自比價 (Self-Relative Valuation)」邏輯。透過比較個股當前估值與其自身歷史水準，判斷股價是否過熱。

本研究定義 相對本益比率 (Relative PE Ratio, ρ) 如下：

$$\rho = \frac{PE_{current}}{PE_{avg_5yr}}$$

基於 ρ 值，設定動態懲罰函數 (Penalty Function) P_{growth} ，以修正原始模型預測分數 $Score_{raw}$ ：

$$Score_{adj} = Score_{raw} \times (1 - P_{growth})$$

其中懲罰係數 P_{growth} 定義為分段函數 (Piecewise Function)：

$$P_{growth} = \begin{cases} 1.0 & \text{if } PE_{current} > 50 \\ \frac{\rho - 1.2}{0.6} \times \lambda & \text{if } 1.2 \leq \rho \leq 1.8 \\ 0 & \text{if } \rho < 1.2 \end{cases}$$

(1) 絕對天花板：當 $PE > 50$ 時，懲罰係數為 1（即得分歸零），直接排除泡沫化風險。

(2) 區間懲罰：當 ρ 介於 1.2 至 1.8 倍之間，隨偏離程度線性增加懲罰力道（ λ 為懲罰強度參數）。

(3) 安全區間：當 $\rho < 1.2$ 時，視為合理估值波動，不予懲罰。

2. 價值股調整：雙尾殖利率校正 (Two-Tailed Yield Correction)，針對價值型個股，傳統策略常誤入「高殖利率陷阱」，即股價因基本面惡化崩跌導致殖利率虛高。為此，本研究設定 目標殖利率中位數 ($Y_{target} = 7.5\%$)，並採用雙尾懲罰機制。

若個股預估殖利率 (Y_{est}) 過度偏離此中位數（無論過高或過低），均視為異常訊號。其校正公式如下：

$$Score_{adj} = Score_{raw} - \delta \cdot |Y_{est} - Y_{target}|$$

(1) 右尾風險（過高）：當 $Y_{est} > 7.5\%$ （例如 $> 12\%$ ），通常隱含重大除息缺口或獲利衰退風險，透過減分將其排序後移。

(2) 左尾風險（過低）：當 $Y_{est} < 7.5\%$ ，則不符合高股息策略之收益目標，同樣給予扣分。

七、流動性濾網與產業階層風險控管

為確保策略具備實務可執行性（Executability），並有效分散非系統性風險，本研究針對流動性與產業集中度實施以下限制機制：

1. **流動性風險排除：** 考量實際交易之滑價成本（Slippage Cost）與進出便利性，本研究針對日均成交金額（Daily Turnover）進行敏感度測試。經測試 1,000 萬至 3,000 萬台幣之區間後，最終設定 1,500 萬元為硬性門檻。凡日均成交額低於此標準之個股，將予以剔除，以確保模型選入標的具備充足之市場流動性。
2. **階層式產業權重控管，**針對價值（Value）模型易過度集中於特定高殖利率產業之特性，本研究採用三層級（Three-Tier）產業控管機制，依據產業屬性與市場權重設定持股上限：

(1) **Tier 1 主流產業：** 包含半導體、電子零組件等權值產業。考量其市場代表性與流動性，容許較高之配置佔比，上限設定為 60%。

(2) **Tier 2 景氣循環產業：** 包含航運、生技等波動較大之產業。為避免受單一景氣循環週期過度影響，限制佔比上限為 20%。

(3) **Tier 3 其他產業：** 其餘非主流產業，合計限制佔比上限為 20%。

3. **產業內集中度懲罰機制**，為防止單一細分產業（如營建、海運）因同質性過高而集體入榜，本研究引入「產業龍頭優選機制（Sector Leader Prioritization）」：

(1) **龍頭保留原則**：在同一非主流產業（Non-Tier 1）中，原則上優先保留模型預測分數最高之「領頭羊」個股。

(2) **次順位懲罰**：若同產業第二檔（含）以上之個股進入候選名單，將給予顯著之分數懲罰（Score Penalty）。此機制旨在強迫資金分散，避免投資組合承受過高之單一產業政策或特定事件風險。



八、多模型進場時機與動態切入策略

本研究之核心預測邏輯並非隨機入場，而是基於財報公告週期之「非對稱資訊揭露」進行定時切片。以日頻作為預測與交易決策頻率，季度營收與財報資料僅用於建構低頻的基本面狀態變數，該狀態在實際公告後於其有效期間內保持不變，並透過每日股價與交易資料之切面觀測，動態評估市場對該基本面狀態的反應。

1. **財報驅動之季度資料切片邏輯**: 考量台股財務報告公告之法規時限 (Q1-Q3 為季後 45 天內，年報為年後 3 個月內)，模型之訓練與預測採「資訊落後同步化」處理。

- (1) **Q2 資料切片 (半年報效應)**: 於每年 8 月中旬完成切片。此時點上半年實際獲利已全數揭露，模型側重於利用「預估值工程」推算全年 EPS，捕捉下半年展望動能上漲之標的。

- (2) **Q3 資料切片 (年終衝刺效應)**: 於每年 11 月中旬完成切片。此時點市場進入「高股息預期」期，模型重點轉向預測隔年之現金股利發放能力與評價回升空間。

2. **雙模型動態切入機制**: 本研究不採取單一模型一體適用之做法，而是根據市場評價水位 (Valuation Level) 動態調整模型權重：

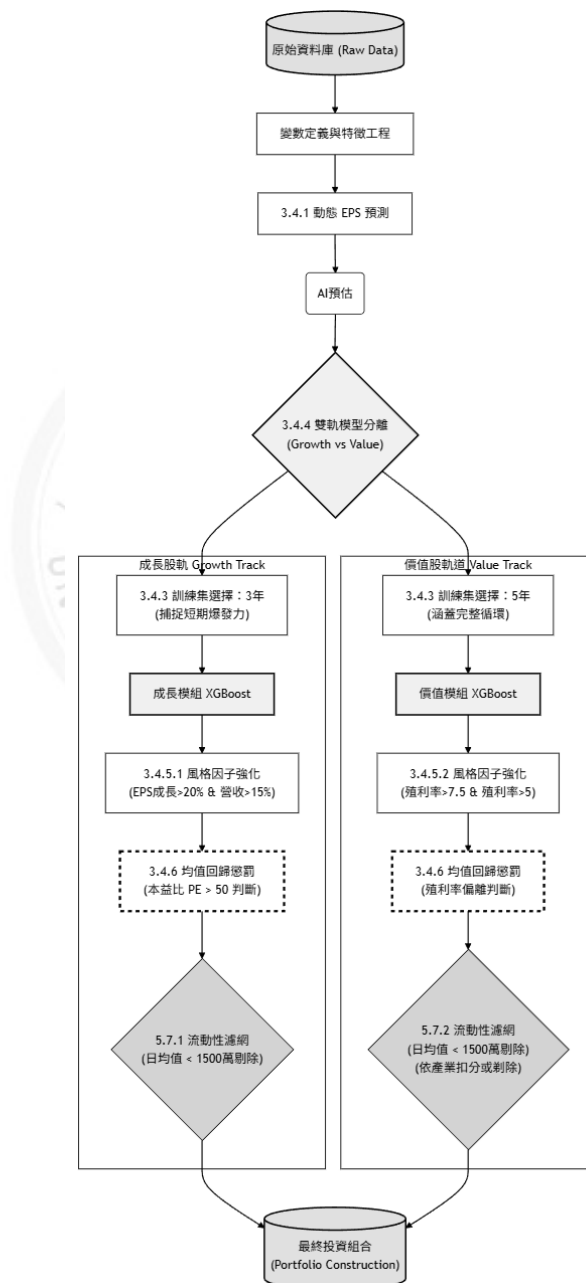
- (1) **Value 模型觸發條件**: 自比價邏輯：當個股「當前 PE」低於「歷史 5 年平均 PE」之一倍標準差，且「預估殖利率」大於 7% 時，模型優先權提升，進場基準日：設定於財報公告後之一個月內完成買入。

- (2) **Growth 模型觸發條件：**動能加速邏輯：側重於「單月營收年增率」連續三個月遞增，且「近一季 ROE」表現優於同產業均值之標的，進場基準日：設定於財報公告後之一個月內完成買入。



第五節 綜合評分系統

本研究構建之選股決策系統採用「複合式評分策略」，整合了 XGBoost 非線性機器學習模型之預測能力，以及基於各參數改善。從動態 EPS 到最終投資組合之產出流程共分為六個階段：



1. **第一階段：初步預估名單**，透過 3.4.1 節 之動態 EPS 預測，分別針對 Q2（結合去年財報與 Q1 季報）與 Q3（結合前三季季報）之時間點，計算全市場個股之「預估 EPS」。此步驟確保輸入模型之本益比（PE）與殖利率（Yield）變數皆具備預測性。
2. **第二階段：基本面相對強弱錨定**，為避免模型過度擬合技術面雜訊，本研究引入基本面錨定機制。依據全市場個股之「預估淨利」與「預估 EPS」進行百分位排名（Percentile Rank, PR）。

計算邏輯：

$$Score_{fund} = w_1 \cdot PR(NetIncome) + w_2 \cdot PR(EPS)$$

此步驟確保入選標的在基本獲利能力上具備相對優勢。

3. **第三階段：AI 預測基礎分**，四大構面特徵因子（基本、籌碼、技術、評價）輸入 3.4.3 訓練長度 3.4.2 節獲利最大化之參數優化後之 XGBoost 模型。以 3.4.4 模型得到成長與價值兩個清單，並針對每一檔個股產出兩項分數：

(1) 價值基礎分

(2) 動態基礎分

4. **第四階段：條件式風格加權**，3.4.5 對不同選股模組，啟動特定風格之動態加分機制：

(1) 成長模組（Growth Bonus）：若個股 AI 基礎分排名前 10% 且具備營收與 EPS 雙重成長趨勢，給予額外加權，以捕捉強勢股的主升段動能。

(2) 價值模組 (Yield Bonus)：若個股預估殖利率落於 5% 至 7.5% 之「甜蜜區間 (Sweet Spot)」，給予額外加權，以鎖定高股息且股價穩健之標的。

5. **第五階段：風險懲罰與限制**，為落實風險控管，系統最後執行 3.4.6 節負向扣分與 3.4.7 節濾網機制：

(1) **估值修正**：依據 3.4.6 節之均值回歸邏輯，針對 PE 過高或殖利率異常之個股進行分數懲罰。

(2) **產業集中度懲罰**：依據 3.4.7 節之規範，針對非主流產業中排名次順位之個股進行扣分。

(3) **流動性濾網**：對於日交易均量過低之個股進行剔除。

6. **第六階段：最終排序與權重分配**，本階段旨在整合前述各階段之分析結果，建構最終投資組合。流程如下：

(1) **複合分數計算**：將上述各階段(如基本面、技術面、機器學習預測等)之評分加總，分別計算出 Q2 與 Q3 預測名單的「最終複合分數」(Final Composite Score)。

(2) **排序與篩選**：依據複合分數由高至低排序，分別從 Q2 預測名單與 Q3 預測名單中，各選取排名前 10 檔個股，共計 20 檔標的。

(3) **資金配置**：將入選之標的構成最終投資組合，並依據預定之資金比例(如等權重、市值加權或分數加權)進行權重分配，以完成部位建置。

第六節 投資組合建構

投資策略首先配合財報公布時程的動態投資組合建構循環。流程始於 8/15 Q2 財報公布，系統依據前述複合分數篩選出「Q2 目標清單」，並於 9/30 前完成 Q2 成長股 (Top 10) 的建倉。隨後於 11/15 待 Q3 財報公布，重複篩選步驟產出「Q3 目標清單」，於 12/31 前完成 Q3 成長股的資金配置。

此策略具備跨年度循環與風控機制，持倉延續至次年，於 8/31 執行負面篩選，賣出 EPS 轉負的標的以汰弱留強；最終於 8/15 銜接新一輪財報，重新啟動投資名單生成的循環，確保組合持續優化。

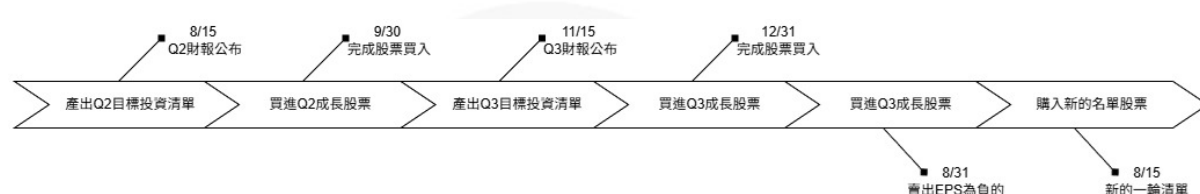


圖 4 交易時間軸

表 8 預估 2026 年 Q2 成長選股名單

Code	Name	季度	綜合評分	動能基礎	價值基礎	動能加分	加分	淨利 PR	EPS_PR
6409	旭隼	Q2	54.15	4.42	7.85	20	25	62	96.3
2327	國巨*	Q3	50.53	4.05	19.41	20	23.66	88	92.6
3017	奇鎡	Q4	39.42	1.14	3.54	20	22.12	82.4	89.8
6196	帆宣	Q1	33.88	0.89	8.91	20	11.32	36.1	70.4
2368	金像電	Q2	32.26	0.93	5.07	20	15.43	71.3	77.8
2612	中航	Q3	31.72	1.17	19.25	20	0	22.2	42.6
2383	台光電	Q4	25.18	1.22	0	20	24.69	80.6	94.4
6669	緯穎	Q1	24.07	0.61	13.43	10	25	93.5	100
2330	台積電	Q2	23.79	0.58	10.6	10	25	100	97.2

Code	Name	進場價格	預估 EPS	預估淨利	預估本益比	自比價	預估股利	預估殖利率
6409	旭隼	951	52.09	4553137	18.26	0.47	39.89	4.19
2327	國巨*	151	34.6	19945966	4.36	0.33	3.1	2.05
3017	奇鎡	1045	30.63	13064101	34.12	1.76	19.01	1.82

6196	帆宣	263	11.04	2170181	23.82	1.41	7.56	2.87
2368	金像電	469	12.99	6324478	36.1	1.66	6.91	1.47
2612	中航	54.3	6.6	1300923	8.23	0.55	4	7.36
2383	台光電	1315	35.01	12103163	37.56	2.16	23.22	1.77
6669	緯穎	3220	190.73	35290952	16.88	1.31	138	4.29
2330	台積電	1255	56.31	1457775143	22.29	1.29	23.68	1.89

表 9 預估 2026 年 Q3 價值選股名單

Code	Name	產業類別	篩選類型	綜合評分	動能基礎分	價值基礎分	殖利率加分	產業扣分
2603	長榮	15 (Tier 2)	Q3	73.21	1.03	28.66	3.17	0
6670	復盛應用	37 (Tier 3)	Q3	67.17	0.7	21.92	7.92	0
6414	樺漢	25 (Tier 1)	Q3	64.99	0.69	22.98	4.57	0
5871	中租-KY	20 (Tier 3)	Q3	64.65	0.36	25.45	3.57	0
2891	中信金	17 (Tier 2)	Q3	63.82	0.37	22.06	4.99	0
3592	瑞鼎	24 (Tier 1)	Q3	63.46	1.08	19.47	9.13	0
2385	群光	28 (Tier 1)	Q3	61.95	0.86	23.74	8.84	0
3617	碩天	31 (Tier 1)	Q3	61.11	0.77	21.9	4.54	0
5225	東科-KY	31 (Tier 1)	Q3	60.34	0.58	23.64	7.11	0

Code	Name	EPS_淨利 加分	淨利 PR 值	EPS_PR 值	預估 EPS	預估淨利	預估本益比	自比價(倍)	預估股利(元)
2603	長榮	25	99.3	97.8	41.93	93168208	4.42	0.88	20.24
6670	復盛應用	24.98	58.3	95	24.06	3622705	10.12	1.2	15.73
6414	樺漢	24.58	74.1	94.2	23.68	8148040	11.91	0.99	13.5
5871	中租-KY	22.18	89.9	74.8	10.93	20175116	9.29	1.02	4.35
2891	中信金	25	97.8	26.6	3.75	75794099	11.39	1.1	2.13
3592	瑞鼎	22.98	23.7	91.4	19.6	1486575	11.33	0.75	15.68
2385	群光	15.79	78.4	78.4	11.33	9516493	11.17	1.01	8.76
3617	碩天	22.18	30.2	89.9	18.76	1775632	11.14	0.88	9.96
5225	東科-KY	16.59	11.5	79.9	11.72	912718	8.58	0.83	6.09

第七節 實證設計與評估指標

本研究旨在驗證「機器學習選股策略」之有效性，透過建構投資組合並與市場主流標竿進行績效對比。實證過程涵蓋短期年度績效與長期累積績效之驗證，並採用多重金融指標與統計檢定方法，以全面評估本策略在超額報酬 (Alpha)、股息收益及風險控管上的表現。

1. 績效比較基準與回測期間

為確保績效評估之客觀性與代表性，本研究選取以下兩類市場主流 ETF 作為對照標竿 (Benchmarks)：

- (1) 高股息型 ETF (High-Dividend ETFs)：選取 元大高股息 (0056) 與 復華台灣科技優息 (00929)，旨在驗證本策略在篩選高殖利率個股及股息收益上的競爭力。
- (2) 市值型 ETF (Market-Cap ETFs)：選取 元大台灣 50 (0050) 與 富邦台 50 (006208)，旨在衡量本策略是否具備超越大盤指數之超額報酬能力。

在評估範疇上，本研究劃分兩大時間維度，以檢視策略的即時效果與長期延伸效益：

- **短期持有效益 (年度績效)**：設定每年 8 月 30 日為年度結算日，計算單一年度之區間含息報酬率，以評估策略在不同市場環境下的年度表現。

- **長期持有效益（累積績效）：**設定回測截止日為 2025 年 8 月 30 日，計算全期間之累積含息報酬率，以檢視策略在完整市場週期內的長期資本增值能力與下檔風險防禦能力。

2. 投資組合風動態權重法計算方法

由於本研究之投資策略具有「分批進場」之特性（即 A 梯次於 8 月中旬進場，B 梯次於 11 月中旬進場），若直接採用全年度平均法計算波動率與回撤，將導致前期（僅持有 A 梯次時）的風險被尚未持有的 B 梯次數據稀釋，產生「前視偏誤（Look-ahead Bias）」。

為精確反映投資人實際持有過程中所承受的風險，本研究針對投資組合之日報酬率序列（Daily Return Series）採用「動態權重法」進行構建，具體演算法如下：

為精確反映投資人實際持有過程中所承受的風險，本研究針對投資組合之日報酬率序列（Daily Return Series）採用「動態權重法」進行構建，具體演算法如下

將年度持有期間劃分為兩個獨立階段，定義每日投資組合報酬率 $R_{p,t}$ ：

(1) **總報酬率（Total Return）：**衡量投資組合在特定期間內的整體資產增長幅度。第一階段（單一梯次持有期， $T_{start} \leq t < T_{mid}$ ）：僅持有 A 梯次成分股（ N_A 檔）。此階段之組合日報酬率為 A 梯次成分股報酬率之平均。

(2) 第二階段（完整梯次持有期， $T_{mid} \leq t \leq T_{end}$ ）：持有 A 梯次與 B 梯次成分股（共 N_{A+B} 檔）。此階段之組合日報酬率為所有成分股報酬率之平均。

數學公式表示如下：

$$R_{p,t} = \begin{cases} \frac{1}{N_A} \sum_{i \in S_A} r_{i,t}, & \text{當 } T_{start} \leq t < T_{mid} \\ \frac{1}{N_A + N_B} \sum_{j \in S_A \cup S_B} r_{j,t}, & \text{當 } T_{mid} \leq t \leq T_{end} \end{cases}$$

表 10 動態欄位定義表

符號	定義	備註
T_{start}	A 梯次進場日	通常為 8 月 15 日
T_{mid}	B 梯次進場日	通常為 11 月 15 日
S_A, S_B	股票集合	分別代表 A、B 梯次選出的股票代號
N_A, N_B	股票檔數	兩梯次各自包含的成分股總數
$r_{i,t}$	個股日報酬率	股票 i 在 t 日的變動百分比

(3) **淨值曲線串接 (NAV Stitching)**：將上述兩個階段的日報酬率序列進行串接，模擬出資金隨時間變動的標準化淨值曲線 $\$V_t$ 。假設初始淨值 $V_0 = 1$ ：

$$V_t = V_{\{t-1\}} \times (1 + R_{p,t})$$

(4) 以修正後淨值曲線 V_t ，重新計算關鍵指標

- **年化波動率 (Annualized Volatility)**：衡量淨值波動的劇烈程度。
- **夏普比率 (Sharpe Ratio)**：衡量每承受單位總風險所能換取的超額報酬。
- **最大回撤 (MDD)**：反映在實際持倉過程中，可能面臨的最大淨值跌幅。

3. 績效評估指標

本研究採用以下六項量化指標進行績效評量。所有報酬率計算皆採「總報酬 (Total Return)」概念，即包含現金股息再投入之含息報酬率，且假設資金為年度單筆投入：

(1) **總報酬率 (Total Return)**：衡量投資組合在特定期間內的整體資產增長幅度。

$$TR = \frac{V_{end} - V_{start}}{V_{start}} \times 100\%$$

(2) **年化報酬率 (Annualized Return)**：將總報酬率換算為年度複利成長率，便於不同時間長度的績效比較。

$$CAGR = \left[(1 + TR)^{\frac{1}{n}} - 1 \right] \times 100\%$$

(3) **首年股息率 (Dividend Yield)**：衡量投資組合初期的現金流回報能力。

$$DY = \frac{\sum_{i \in S} D_i}{V_{start}} \times 100\%$$

(4) **年化波動率 (Annualized Volatility)**：衡量報酬率的標準差，作為投資組合總體風險之代理指標。

$$\sigma_{ann} = \sqrt{252} \times \sqrt{\frac{1}{N-1} \sum_{t=1}^N (R_{p,t} - \bar{R}_p)^2}$$

(5) **夏普比率 (Sharpe Ratio)**：衡量每承擔一單位總風險所能獲得的超額報酬 (Risk-adjusted Return)。

$$S = \frac{CAGR - R_f}{\sigma_{ann}}$$

(6) **最大回撤 (Maximum Drawdown, MDD)**：衡量資產淨值從高峰跌至谷底的最大跌幅，評估策略之極端風險與抗跌能力。

$$MDD = \max \left(\frac{V_{peak} - V_t}{V_{peak}} \right) \times 100\%$$

第四章 實證結果分析

第一節 預測模型之歷年績效驗證

本節旨在驗證本研究所提出之「動態 EPS 雙軌預測模型」在實務上的預測效能與穩定性。整體實證分析分為兩個層次：首先檢驗 Q2 年中預測的早期鎖定能力，其次探討 Q3 季末修正後的資訊收斂效果。

表 11 測說明與結果

驗證類型	關鍵預測點	真實數據點	核心指標	數據對比（結果）
Q2 預測 （表 12）	Q2 預估 EPS/ 市場前 20% 排名	Q4 實際 EPS 排名	命中率（Hit Rate）/ 預測偏差（%）	命中率 80.88%， 偏差 +2.58%。
Q3 預測 （表 13）	Q3 預估 EPS/ 市場前 20% 排名	Q4 實際 EPS 排名	命中率（Hit Rate）/ 預測偏差（%）	命中率 86.15%， 偏差 +2.97%。
結論論述：命中率從 80.88% 提升到 86.15%，展現了預測的收斂性，證明模型能有效利用新增的 Q3 資訊。同時，偏差的縮小也證明了預測的穩健性。				

1. Q2 預測效力分析

表 4-1 呈現模型於年中（Q2 財報公布後）進行預測的歷年結果，重點評估其對「年度最終 EPS 前 20% 個股（Actual EPS Top 20%）」之捕捉能力與預測偏差。

表 12 Q2 預測模型驗證

年度	Q4 前 20%	Q2 預測數	準確率%	預測偏差%
2015	184	143	77.3	-1.1
2016	187	146	77.66	-2.24
2017	190	151	79.06	3.2
2018	193	150	77.32	4.62
2019	196	160	81.22	-1.99
2020	200	161	80.1	13.91
2021	201	170	83.33	6.95
2022	202	166	81.77	-1.74
2023	211	171	80.66	-5.9

2024	212	182	85.85	10.14
------	-----	-----	-------	-------

$$\text{準確率} = \frac{\text{成功命中的數量(Hit)}}{\text{模型選出的總數(Total Picks)}}$$

(1) 高度的市場防禦力

實證結果顯示，模型於 Q2 時點的歷年平均命中率為 **80.88%**。此結果意謂投資人僅需依據半年報資訊，搭配本研究所提出之雙軌篩選機制（EPS 排名 ∪ 淨利排名），提前找出年底 EPS 前 20% 的公司。

(2) 保守穩健的預測偏差

模型於 Q2 時點的平均預測偏差為 **+2.58%**，代表實際 EPS 普遍略高於模型預估值。特別是在 2020 年（+13.91%）與 2024 年（+10.14%）等景氣擴張年，模型傾向給予較保守的預估，這種「保守偏誤」有助於投資人建立安全邊際。

2. Q3 預測效力分析

隨著 Q3 財報資訊揭露，模型得以整合更完整的年度獲利資訊，其預測效力進一步提升，相關結果如表 4-3 所示。

表 13 Q3 預測模型驗證

年度	Q4 前 20%	Q3 預測數	準確率%	預測偏差%
2015	184	149	80.54	-2.11
2016	187	155	82.45	4.22
2017	190	159	83.25	4.6
2018	193	166	85.57	2.07
2019	196	172	87.31	-0.62
2020	200	173	86.07	11.87
2021	201	180	88.24	5.64
2022	202	183	90.15	-4.1

2023	211	185	87.26	2.26
2024	212	182	85.85	5.83

(1) **資訊收斂效應**：實證結果顯示，隨著 Q3 財務資訊的納入，模型的平均命中率從 Q2 的 80.88% 顯著提升至 86.15%。特別是在 2022 年，命中率一舉突破 90.15%。這證實了模型具備優異的「動態修正能力」，能隨時間推移，提高名單預測準確率。

(2) **穩健性驗證**：即便在加入更多資訊以追求更高命中率的過程中，模型的平均預測偏差仍穩定維持在 +2.97%，與 Q2 時點 (+2.58%) 展現了高度的一致性。這顯示模型並未因資訊量增加而產生過度擬合 (Overfitting) 或激進偏誤，始終維持穩健且略微保守的預估風格。這種穩定性對於機構法人或長期投資者而言，是建構投資組合時極為關鍵的風險控管指標。

3. 實證小結：預測的可信度與穩定性

綜合 Q2 與 Q3 的歷年驗證結果，本研究確認了動態 EPS 預測模型的兩大核心價值：

(1) **高可信度**：無論是在年中 (Q2) 或季末 (Q3)，模型對於最終進入前 20% 領先群個股的命中率均維持在 8 成以上，顯示排名指標具有極高的市場代表性。

(2) **時效性與準確性的平衡**：Q2 預測提供了極佳的布局先機 (Early Detection)，而 Q3 則透過資訊吸收完成精準修正 (Convergence)。對投資者而言，本模型的預測結果不論處於哪一個階段，皆能提供穩定且具安全邊際的个股決策支持。

第二節 成長因子的微調效應與最佳參數確立

本研究進一步探討引入「成長動能因子(Slope)」對模型的具體影響。根據 Grid Search 參數最佳化實驗（參見表 14），本研究確立 10% 為兼顧穩定性與成長性的最佳權重配置。

1. 最佳參數點 ($w = 10\%$):

實驗發現，當成長權重設定為 10% 時，綜合分數達到最高 (171.81)。在此權重下，模型成功地將平均預測偏差從 $w = 0\%$ 時的 +0.91% 完美校正至 +0.01%。這證明了 $w = 10\%$ 的微量成長因子能有效消除模型的系統性偏差，使預測達到更好的狀態。

2. 邊際效益遞減與成長陷阱：

一旦成長權重超過 20%，覆蓋率開始急劇崩塌，且預測偏差亦快速惡化 ($w = 50\%$ 時偏差升至 28.33%)。此現象再次支持了理論，即在篩選 Top 20% 績優股時，獲利規模 (Scale) 的權重必須遠高於 獲利速度 (Speed)。

表 14 基礎與加權預測比較

驗證類型	關鍵預測點	真實數據點	核心指標	數據對比 (結果)
優化基礎 ($w = 0\%$)	Q2 純 EPS 排名	Q4 實際 EPS 排名	預測偏差 (%) / 綜合分數	偏差 +0.91%， 綜合分數 170.40。
優化目標 ($w = 10\%$)	Q2 (90% EPS + 10% Slope) 綜合分	Q4 實際 EPS 排名	預測偏差 (%) / 綜合分數	偏差 +0.01%， 綜合分數 171.81。
名單置換 (表 15)	$w = 0\%$ 名單 vs. $w = 10\%$ 名單	N/A	名單替換率 (%)	平均替換率約 4.7%， 重疊率 95.3%。

3. 微調的價值名單替換率分析：

策略的高度穩定性：歷年平均名單重疊率高達 95.3%。這證明了參數優化並未動搖模型的基礎，獲利規模 始終是選股的核心。

4. 微調的價值：

歷年平均名單重疊率高達 95.3%。這證明了參數優化並未動搖模型的基礎，獲利規模 始終是選股的核心，平均每年僅替換約 4.7% 的名單（平均每年踢出 8.6 檔）。這少數的置換，將原本在 $w = 0\%$ 名單中、但獲利正在劇烈衰退的 個股（如 2015 年的美律、2016 年的上銀）剔除，並替換為成長動能強勁的轉 機股。正是這 4.7% 的關鍵修正，成功消除了系統性偏差，實現了模型的精準 化。

表 15 歷年優化名單重疊率與替換率統計

年度	選股總數	重疊數量	重疊率%	踢出數量	名單替換率%
2015	184	177	96.2	7	3.8
2016	187	180	96.26	7	3.74
2017	190	180	94.74	10	5.26
2018	193	179	92.75	14	7.25
2019	196	184	93.88	12	6.12
2020	200	190	95	10	5
2021	201	190	94.53	11	5.47
2022	202	190	94.06	12	5.94
2023	211	198	93.84	13	6.16
2024	212	201	94.81	11	5.19

5. 實證小結：

最後實證結果確立 $w = 10\%$ 為兼顧預測穩定性與成長動能篩選之最佳參數配置。透過此 10% 成長因子之調節，模型成功將平均預測偏差精準校正至 $+0.01\%$ ，在秉持以獲利規模為核心邏輯的前提下，充分發揮了關鍵名單汰弱留強之優化功能。由圖 4 動態流向來看，模型於 Q2 初步選股階段篩選之 212 檔標的中，至 Q3 仍有 196 檔穩定留存，最終更有高達 172 檔個股成功躋身年度最終 Top 20% 領先群，換算命中率穩定維持在 8 成以上。名單的年度高連續性與預測精準度，確切實證了本研究模型指標在不同資訊揭露時點均具備極高之穩定性與市場代表性。

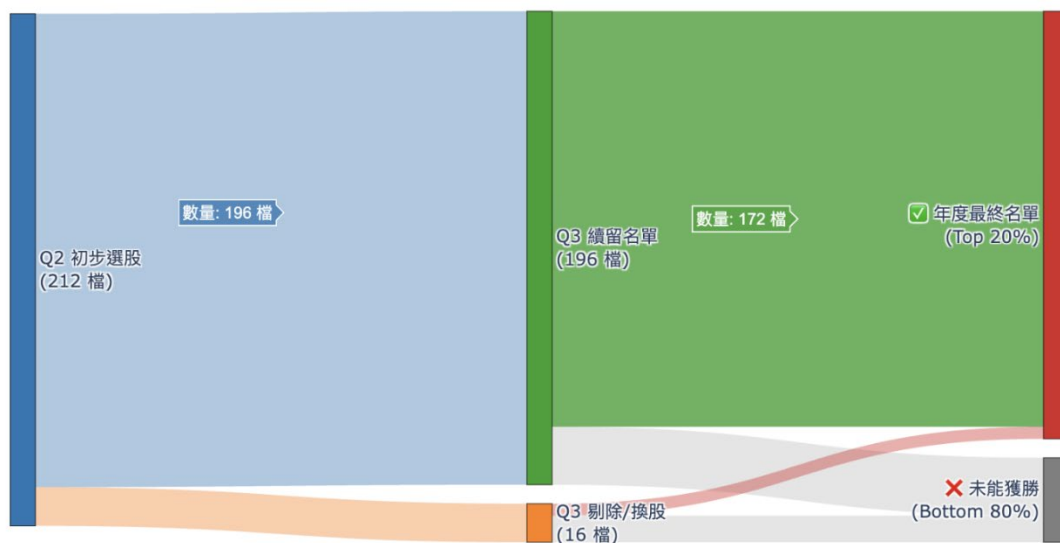


圖 5 Q2 至 Q3 選股名單置換到 Top 20% 流向圖

第三節 參數優化目標之實證分析

本研究旨在探討機器學習在金融預測中，目標函數 (Objective Function) 的選擇對實際投資績效之影響。為此，本節構建了對照實驗，將回測區間依據財報公告與進場時點拆解為 Q2 (8 月中進場) 與 Q3 (11 月中進場) 兩大投資窗口，並對比 A 組 (RMSE 最小化，代表貝葉斯優化最優) 與 B 組 (獲利最大化，代表本研究自訂參數) 之績效差異。

1. 獲利導向優化之優越性驗證

實證結果 (如表 16 所示) 顯示，雖然 A 組 (RMSE 最佳化) 在預測精準度上略勝一籌，但 B 組 (獲利最大化) 在所有報酬率指標上均全面優於 A 組，驗證了針對獲利目標進行自訂參數優化的優越性。

- (1) **關鍵數據：**在 2017 年至 2025 年的觀測區間中，B 組在多個關鍵年份展現出顯著的超額報酬。例如在 2023 年與 2024 年，B 組分別繳出 47.80% 與 39.72% 的平均報酬，大幅領先 A 組 (貝葉斯優化) 的 38.77% 與 27.61%，在市場環境較為險峻的 2018 年，A 組呈現負報酬 (-4.33%)，而 B 組仍能維持 5.71% 的正成長；同樣在 2019 年市場回檔時，B 組的跌幅 (-3.56%) 亦小於 A 組 (-5.93%)。
- 預測精準度：**A 組的平均 RMSE 為 2.4546，略低於 B 組的 2.4567 (差異僅為 0.0021)。這說明 A 組在數學統計意義上更貼近原始數據波動，但此微幅領先並未轉化為實質投資優勢，**獲利表現：**在關鍵季節報酬率中，B 組展現出顯著優勢。特別是在 Q3 Top 10 標的選擇上，B 組繳出 25.99% 的平均報酬，大幅領先 A 組的 19.11%，超額報酬達 6.88%，**穩定性與廣度：**當觀測範圍擴大至 Top 20 時，B 組在 Q2 與 Q3 依然分別維持 22.51% 與 22.16% 的穩定水準，

持續領先 A 組。這顯示 B 組的參數設定在名單擴張後，仍具備高度的績效穩定性。

- (2) **學術意涵與模型確立：**此結果有力地支持了本研究的核心假設：「統計預測誤差 (RMSE) 的極小化，並不同於投資獲利的極大化」。在金融市場的高度雜訊環境下，過度追求低 RMSE 可能導致模型陷入過度擬合 (Overfitting) 於非理性的微小波動中，相反地，B 組直接以「獲利結果」為導向進行參數調整，雖然犧牲了極微小的統計精準度，卻能更有效地捕捉具備趨勢爆發力的特徵標的。這種「以終為始」的優化邏輯，更符合實務投資決策的需求。因此，本研究後續章節將統一採用 B 組參數作為回測與模型架構的核心基礎。



表 16 參數優化目標之績效總表

評估維度	指標	A 組 (RMSE 最佳化)	B 組 (獲利最大化)	差異	勝出者
預測精準度	平均 RMSE	2.4546	2.4567	0.0021	A 組 (預測較準)
關鍵季節獲利	Q2 Top 10 報酬率	23.34%	25.81%	2.47%	B 組 (績效領先)
	Q2 Top 20 報酬率	22.42%	22.51%	0.09%	B 組 (微幅領先)
回測穩定性	Q3 Top 10 報酬率	19.11%	25.99%	6.88%	B 組 (顯著優勢)
	Q3 Top 20 報酬率	20.03%	22.16%	2.13%	B 組 (穩定擴張)
核心假設驗證	核心假設驗證	模型誤差極小化	名單擴張仍保穩定	-	B 組 (綜合勝出)

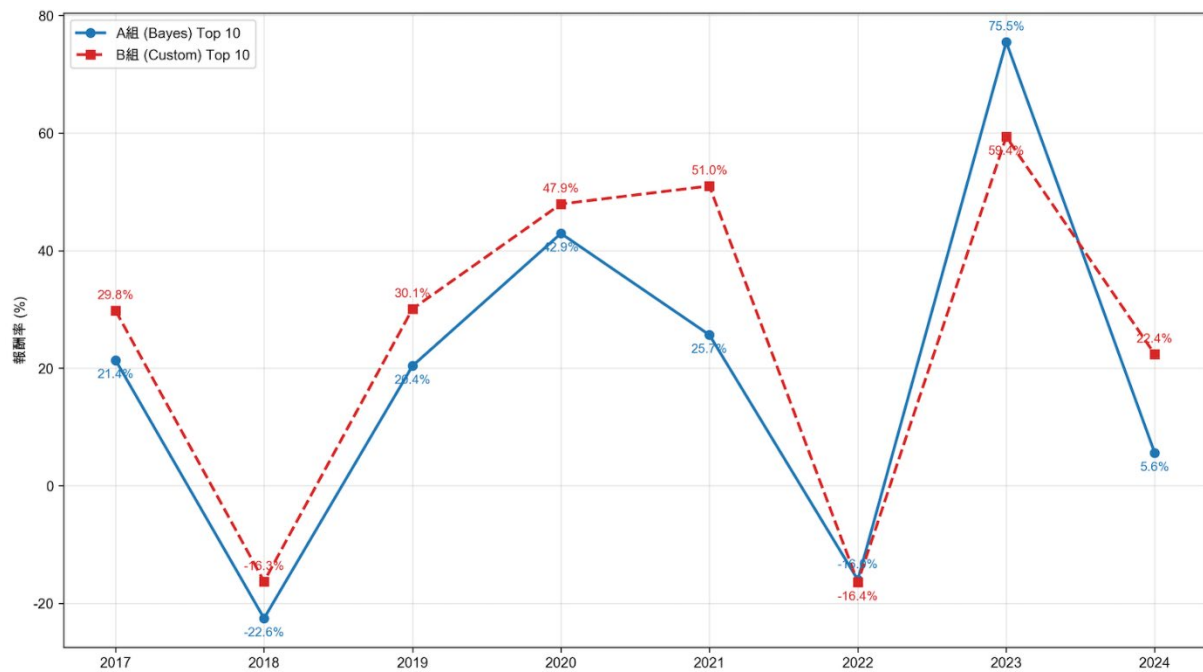


圖 6 獲利最大化比較

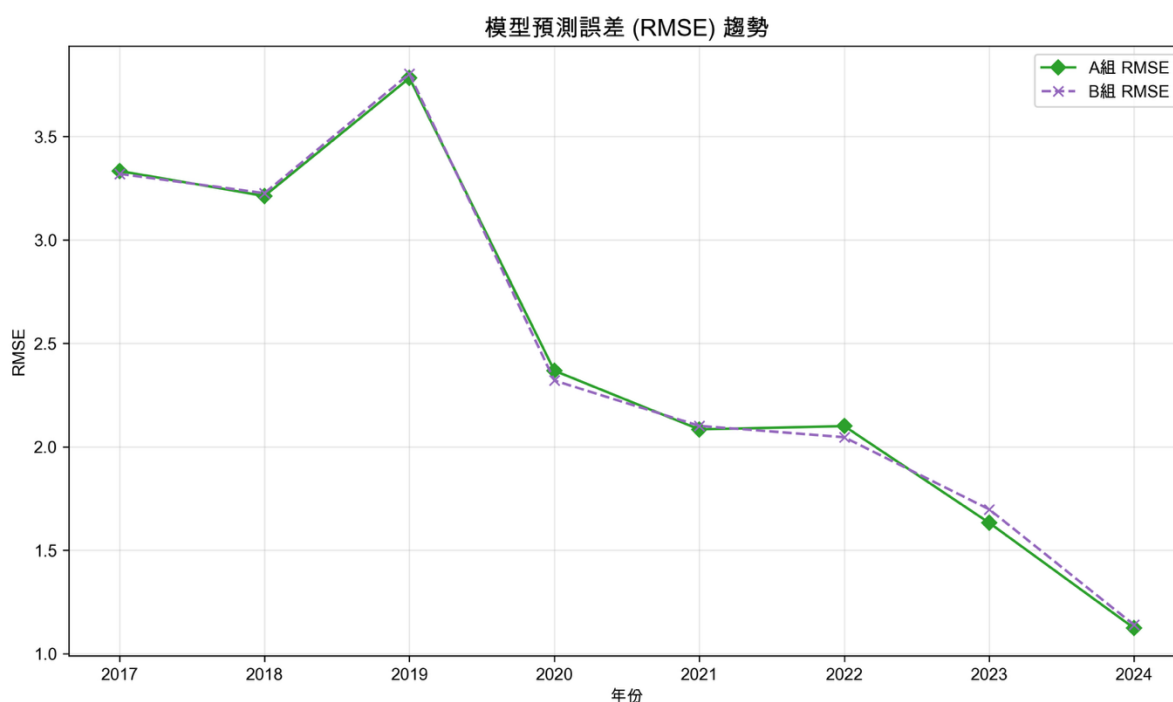


圖 7 RMSE 比較

2. 季節性效應

透過拆解 Q2 與 Q3 之績效表現，本研究發現顯著的「季節性結構差異」：

- (1) **現象觀察：** 無論是 A 組或 B 組，Q2 名單的預測準確度與獲利爆發力普遍優於 Q3 名單。這反映了 8 月中旬（Q2 財報公布後）市場資訊含量最高，且適逢除權息行情尾聲與隔年佈局初期，股價反應最為劇烈。
- (2) **開啟後續研究：** 此一發現凸顯了「單一靜態策略」的侷限性。由於 Q3（11 月）面臨財報空窗期與年底作帳行情的不同市場動力，顯示策略必須具備**動態調整機制**。這為本研究第三章所提出的「動態投資組合建構」與「雙軌風格分離機制」提供了實證支持：針對不同季節的市場特性，模型需要不同的適應性調整，而非一體適用。

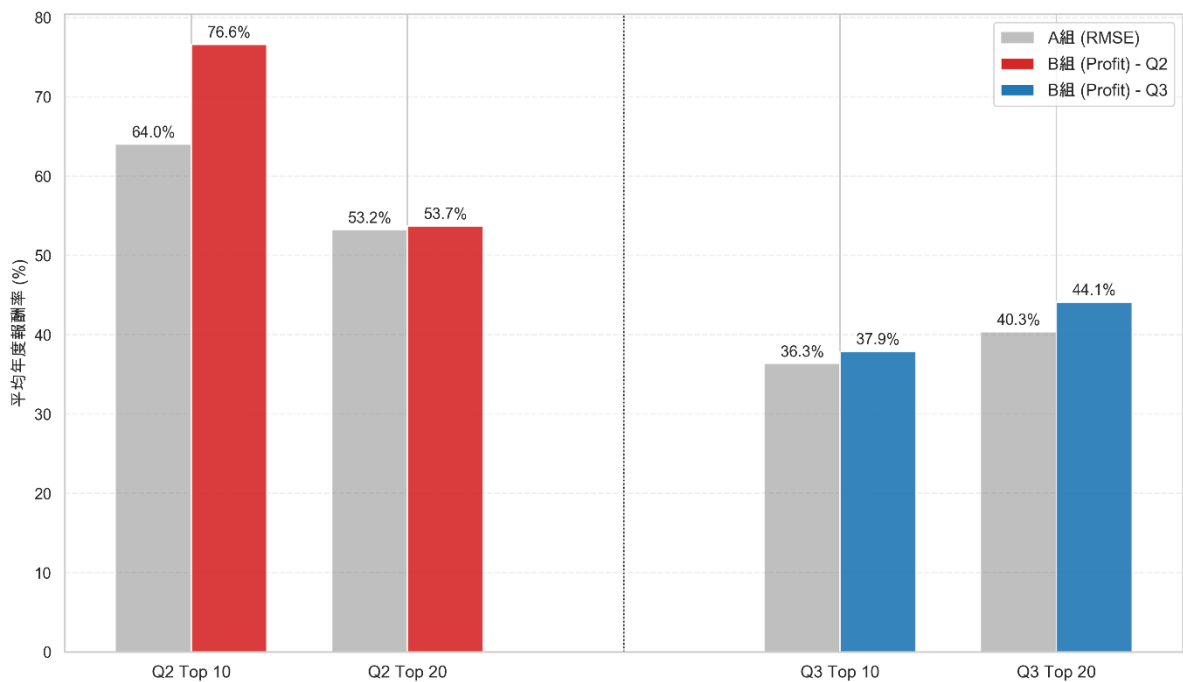


圖 8 Q2、Q3 獲利狀況

3. 實證小結：

本節實證結果顯示，以『獲利最大化』為目標函數之 B 組參數，在識別具備超額報酬潛力的個股上展現顯著優勢。即便將選股範圍擴大至 Top 20，其資產組合之勝率（含金量）仍維持穩定，這為後續章節導入『基本面濾網』、『產業風險控管』與『流動性篩選』提供了充分的理論支撐與實證基礎。此外，針對 Q2 與 Q3 選股產出之差異，將於後續章節進行深入探討與成因分析。

第四節 模型結構與訓練視窗敏感度檢定

本章透過敏感度檢定與特徵重要性分析，成功驗證針對不同屬性的股票池採用差異化的訓練策略，能顯著提升模型的預測效力與邏輯合理性。

1. 研究動機與實驗設計

實驗數據表明，單一時間視窗無法同時滿足成長與價值的預測需求。本研究確立了以下配置：

- (1) **時序敏感度檢定**：針對成長與價值兩類模型，分別測試 3 年與 5 年滾動訓練視窗，觀察其對回測報酬與最大回撤的影響。
- (2) **雙軌分流驗證**：透過特徵重要性分析，檢視單一混合模型是否存在特徵稀釋問題，以證成雙軌模型的必要性。

2. 訓練視窗的時序敏感度分析

實證結果顯示，不同投資風格的因子對「時間長度」的需求存在顯著差異。強行統一訓練視窗將導致模型效能的次佳化，採用「非對稱視窗」策略能顯著提升預測力。

- (1) **成長型模型 (Growth Model)：短期動能的捕捉優勢**：針對成長股，實驗數據表明 3 年視窗 ($W = 3$) 為最佳參數配置

- **多頭爆發力**：在 2024 年的多頭行情中，3 年視窗模型展現了更高的捕捉效率，回測報酬率達 36.52%，優於 5 年視窗的 35.10%。
- **空頭防禦力**：在 2022 年空頭市場中，3 年視窗因能快速反應趨勢反轉，表現抗跌性 (11.16% vs -3.58%)。

- **實際辯證：** 成長股的本質往往伴隨著技術典範的轉移(如從 PC 週期轉向 AI 伺服器週期)。數據顯示，超過 3 年的歷史數據對於預測高成長個股往往存在「資訊衰減」現象。5 年前的市場結構數據非但無法提供預測力，反而成為雜訊。因此，縮短視窗至 3 年，是在「捕捉最新動能」與「統計顯著性」之間取得的最佳平衡。

(2) **價值型模型 (Value Model)：均值回歸的週期驗證：**針對價值股，實
驗數據顯示 5 年視窗 ($W = 5$) 具備更優異的穩定性。

- **下檔保護：** 在 2022 年熊市中，5 年視窗模型將最大回撤控制在 -12.88%，優於 3 年視窗的 -16.14%。
- **選股穩定度：** 該模型對華碩、力成等成熟型企業的評分變異度較低，未因短期市場波動而頻繁進出。
- **理論辯證 (週期性驗證)：** 價值投資的核心邏輯建立於「均值回歸 (Mean Reversion)」之上。為了區分「暫時被錯殺的優質股」與「基本面惡化的價值陷阱」，模型必須跨越至少一個完整的景氣循環 (Business Cycle)。較長的數據長度 (5 年) 讓模型得以學習企業在不同經濟環境下的獲利韌性與配息能力，從而過濾掉短期獲利爆發的假象。

3. 雙軌機制的特徵互斥性驗證

除了時間維度，本研究進一步深入模型內部的決策邏輯。透過比較 XGBoost 模型的特徵重要性排名，我們發現成長與價值模型的驅動因子存在**高度互斥性**。此結果強有力地反駁了「通用單一模型」的可行性。

表 17 特徵與決策邏輯驗證

特徵維度	成長模型排名	價值模型排名	決策邏輯衝突解析
PE_Relative_Ratio (自比價)	第 1 名(14.4%)	第 5 名 (6.2%)	擇時 vs. 品質：成長模型極度依賴「歷史相對低點」來捕捉動能轉折；價值模型因已預篩選低 PE，故權重自動轉移至基本面品質。
Revenue_YoY (營收動能)	第 7 名 (4.8%)	未入榜	爆發 vs. 穩定：成長模型需營收爆發支撐高估值；價值模型則對短期營收波動不敏感，更看重長期現金流。
Est_EPS (預估 EPS)	未入榜	第 2 名 (9.0%)	夢想 vs. 現實：價值模型高度依賴分析師的未來獲利預估 (Forward EPS)，視其為股價修復 (Re-rating) 的核心催化劑。
Volatility_3m (波動率)	第 10 名 (正向)	第 9 名 (負向)	活潑 vs. 穩健：成長模型偏好高波動帶來的價差空間；價值模型則懲罰高波動（特別是配息不穩定的標的）。
IndustryCode (產業)	第 2 名	第 1 名	共同驗證：兩者皆認同「選對賽道 (Beta)」的重要性，證明模型具備正確的宏觀判斷能力。

(1) **模型特色的確立與互補**：上述特徵分歧證實了雙軌模型已演化出截然不同的決策邏輯，成功避免了單一模型導致的特徵稀釋

- **Growth 模型 (趨勢交易)**：採取「動能 + 擇時」策略。它不畏懼高本益比，但嚴格要求股價處於自身的相對低檔 ($PE_{textRelativeRatio}$)，且具備強勁的營收動能。

- **Value 模型 (基本面交易)**: 採取「低估值 + 品質」策略。
在便宜的股票池中，優先挑選具備未來獲利成長 (Est_{EPS})
與配息穩定性 ($Yield_{Std}$ 低) 的標的。

4. 實證小結：

透過時序敏感度與特徵異質性的雙重實證，成功證成『雙軌非對稱視窗』架構的必要性：在時序維度上，利用成長模型（3 年視窗）的『短記憶性』精準捕捉產業動能轉折，並結合價值模型（5 年視窗）的『長記憶性』跨越景氣循環驗證均值回歸邏輯；在決策維度上，特徵權重的顯著分歧證實了成長與價值策略在因子驅動上具備高度互斥性，透過雙軌分流有效化解了單一模型必然面臨的特徵稀釋難題。



第五節 增加濾網效益分析

在本研究的 AI 原始評分模型之上，疊加「財務硬約束」（如 EPS 成長、股利發放門檻、流動性指標等）作為第二層濾網。實證結果顯示，濾網機制在多數年度能有效剔除具備高 AI 評分但財務韌性不足的標的，進而優化策略組合之風險報酬比。

1. 年度報酬率對比表

透過對比無濾網（原始組合）與有濾網（優化組合）之績效（見表 5-1），可觀察到濾網機制在不同市場週期下展現的增益效果：

表 18 濾網機制對年度與累積報酬之影響對比

年度	模型類型	年度報酬率 (無濾網)	年度報酬率 (有濾網)	累積報酬率 (無濾網)	累積報酬率 (有濾網)	績效增量 (年度)
2017	Growth	19.31%	20.31%	333.36%	466.19%	+1.00%
2017	Value	12.95%	20.33%	144.74%	108.04%	+7.38%
2018	Growth	-13.28%	-7.28%	102.75%	152.54%	+6.00%
2018	Value	-1.1%	2.12%	73.81%	95.12%	+3.22%
2020	Growth	16.31%	36.46%	319.92%	391.04%	+20.15%
2022	Growth	3.19%	6.4%	168.1%	189.72%	+3.21%
2025	Value	23.65%	29.97%	12.19%	30.11%	+6.32%

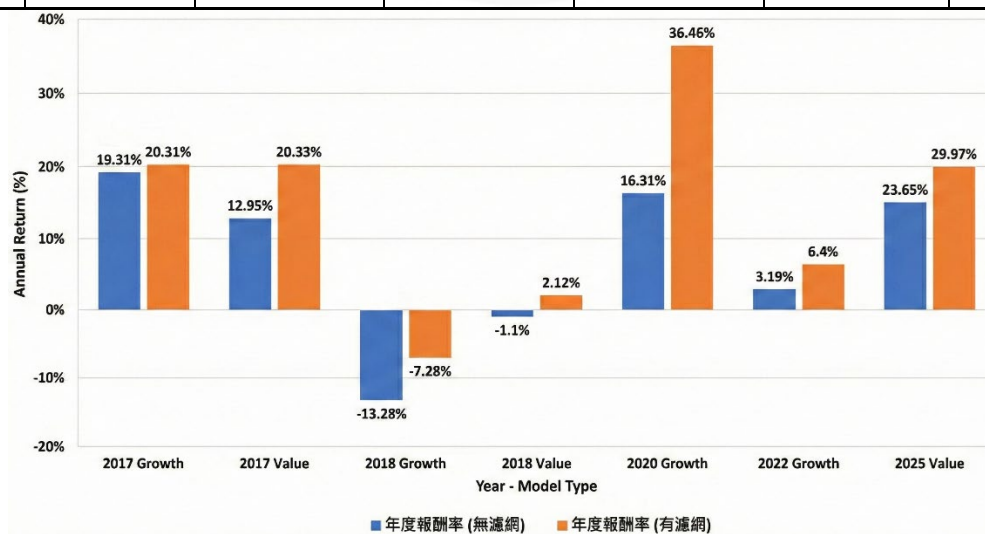


圖 9 有無濾網年度報酬率比較

2. 數據詮釋：風險規避與收益增強：

- (1) **下行風險之防禦：**在市場修正期（如 2018 年系統性風險爆發），濾網機制展現了強大的防禦屬性。Growth 模型在無濾網狀態下回撤達 -13.28%，但在導入財務門檻後，跌幅收窄至 -7.28%。這證實了財務約束能有效剔除「純動能驅動」但缺乏基本面支撐的投機性標的。
- (2) **多頭行情之收益增強：**在 2020 年受寬鬆貨幣驅動的多頭行情中，濾網機制透過剔除「品質瑕疵股」並聚焦於「強勢龍頭標的」，使 Growth 模型年度報酬從 16.31% 翻倍成長至 36.46%。這說明了 AI 分數結合質量指標（Quality Factor）能產生正向的交互作用。

3. 標的汰換之微觀實證觀察

- (1) **2017-2018 年質量升級：**在 Growth 模型中，原始組合傾向選入如 3055(蔚華科)、2062(橋樑) 等中小型標的；經濾網篩選後，組合轉向獲利能力更穩健的權值藍籌，如 3008(大立光)、2330(台積電)。實證顯示，在景氣循環末端，傳統龍頭股的抗跌性顯著優於中小型標的。
- (2) **2020 年排除負貢獻因子：**濾網成功辨識並剔除高 AI 評分但財務指標轉弱的 4943(康控-KY) 與 5225(東科-KY)，改由 2308(台達電) 與 3413(京鼎) 替代。此舉不僅規避了後續的股價修正，更因掌握產業龍頭的回升行情而大幅超越大盤。
- (3) **2025 年價值重估與標的優化：**價值模型透過排除高估值疑慮的 3406(玉晶光)，替換為具備訂單能見度與低估值優勢的 6139(亞翔)，使報酬率由 23.65% 提升至 29.97%，達成「估值安全邊際」與「獲利成長性」的動態平衡。

4. 實證小結：

濾網機制的導入對投資組合的 夏普比率 (Sharpe Ratio) 與 最大回撤 (MDD) 均有顯著改善。其核心價值在於提供了一層「非統計性」的理性約束，緩解了機器學習模型在面對極端市場樣本或結構性轉變時可能產生的過擬合風險。這種「AI 動能偵測 + 財務硬約束」的混合框架，能有效提升策略在實務環境中的可執行性與累積報酬的穩定度。



第六節 投資組合品質實證與選股邏輯一致性檢定

本章節旨在透過財務基本面指標，驗證 AI 模型所建構投資組合之安全性與獲利品質。本研究不僅關注價格動能，更重視標的之營運實質，以確保投資組合在追求高報酬的同時，具備防禦系統性風險。

1. 基本面安全性實證：EPS 與股利發放

透過追蹤歷年組合中「EPS 為負數」及「發放現金股利」之個股數量，可評估模型是否成功避開基本面轉壞之企業。

(1) **極高之獲利勝率**：在九年回測期間，有六個年度實現了投資組合「無虧損企業」的目標。即便是面臨全球升息與通膨壓力的 2022 年，組合中獲利企業占比仍高達 85%。這證明模型能精確識別企業的獲利韌性，有效規避信用風險與財務陷阱。

(2) **現金流量之穩健性**：歷年組合中，發放現金股利之個股幾乎維持在 95% (約 19~20 檔)。穩定的股利政策是企業現金流充沛且財務健全的直接指標。本模型選股具備找出「盈利且願意分配」之個股，為組合提供了強大的價值支撐與下檔保護。

表 19 股利發放與殖利率

年度	持有單位數	有發放股利單位數	年度發放率 (%)	平均股利殖利率 (%)
2017	20	20	100%	6.46%
2018	40	40	100%	5.46%
2019	60	60	100%	6.03%
2020	80	77	96%	5.64%
2021	100	95	95%	6.11%
2022	120	115	96%	6.16%
2023	140	134	96%	5.83%

2024	160	148	93%	7.01%
------	-----	-----	-----	-------

- (3) **績效之廣度**：在 2021 年與 2024 年等多頭年份，年度報酬率為正之個股占比達 90%。這顯示模型不僅能捕捉少數飆股，更能帶領整體投資組合實現普遍性的獲利，展現了極佳的策略廣度。

表 20 投資組合年度品質統計表

年度	EPS 為負 個股數	股利發放 個股數	年度報酬率為正	總持股數 (含重複)
2017	0	20	15	20
2018	0	20	7	20
2019	0	20	12	20
2020	2	20	9	20
2021	0	19	18	20
2022	3	20	10	20
2023	0	19	15	20
2024	1	20	18	20
2025	0	20	12	20

2. 持股延續性分析：核心標的之動態輪動邏輯

對模型傾向於「短期投機」或「價值輪動」，本節針對 180 次持股決策進行「去重複 (Unique Stocks)」統計，藉此檢視選股邏輯的一致性。

- (1) **重複持有率**：統計顯示，九年累計 180 檔持股樣本中，去重複後僅為 92 檔個股，因該年度該股票符合時進行重複買進，以優質股票加碼持有，名單比重會透過年度的方式，類比加權買入的方式。
- (2) **非隨機性之決策特徵**：近 50% 的高重複率證實模型展現了鮮明的「核心優質企業輪動」特徵。模型並非隨機追逐市場熱點，而是反覆辨識出如台積電 (2330)、大聯大 (3702)、華碩 (2357) 等具備長期

競爭優勢的標的。當此類標的符合 AI 動能評價且處於合理價值區間時，模型會重複納入配置。

- (3) **一致性與獲利路徑：**模型圍繞 92 檔核心企業進行動態調整，採取「賣出過熱者、買回修正者」的策略邏輯。這種高度的選股集中度，說明 AI 模型對於台灣市場「高成長」與「高價值」特徵的識別具有極強的連貫性 (Path Dependency)，確保了投資決策的穩定性。

3. 實證小結：

綜合實證數據得出結論：模型經多因子篩選建構「高純度獲利」特徵組合，有效排除虧損企業確保安全性；個股近全數發放股利，保證營運實質，於多頭市場展現極高勝率與獲利性；近 50% 重複持有率證明邏輯非隨機，而是精準抓牢具競爭力的核心企業，呈現高度穩定性。這種結合被動穩健與主動超額報酬，「圍繞核心優質標的動態輪動」的策略，是維持長期績效領先之關鍵。

第七節 策略績效總覽與標竿比較

本章將本研究建構之 AI 選股策略組合（以下簡稱「本組合」）與台灣市場具代表性之指數股票型基金（ETF）進行橫向績效對比。對照標竿包括市值型代表（元大台灣 50, 0050）及高股息代表（元大高股息, 0056），藉以比較三者的表現。

1. 報酬率維度：

實證數據顯示，本組合在多數年份展現出顯著的獲取超額報酬能力：

- (1) **多頭爆發力：** 在 2024 與 2025 年，本組合分別實現了 41.61% 與 48.49% 的高額回報，顯著超越同期的 0050（48.64% 與 18.80%）與 0056。特別是在 2025 年，當市場風格轉向時，本組合的報酬率高出市值型標竿近 30 個百分點。
- (2) **風格切換韌性：** 在 2022 年全體市場修正之際，本組合報酬率僅輕微修正 -2.66%，展現了優於 0050（-9.42%）與 0056（-6.91%）的抗跌性。這說明 AI 模型能透過因子分流，在空頭環境中有效識別具備基本面支撐的防禦性標的。

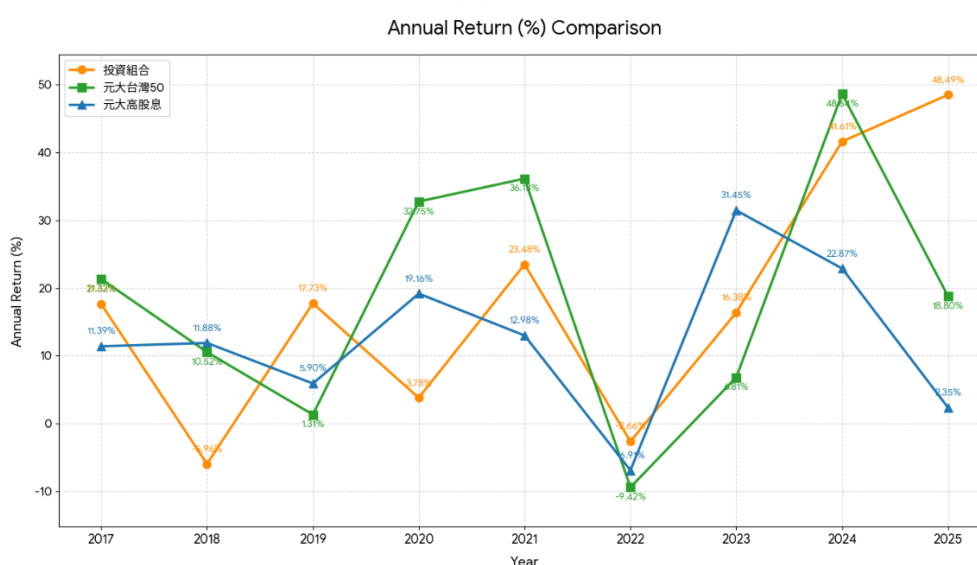


圖 10 年度報酬率比較

2. 夏普值維度：

夏普值 (Sharpe Ratio) 衡量的是每承受一單位風險所能獲得的超額回報。

本組合在效率表現上具有顯著優勢：

(1) **效率的一致性：** 在 2024 年 (1.85) 與 2025 年 (1.68)，本組合的夏普值均維持在高檔，顯示其報酬的獲取並非單純依賴擴大風險敞口，而是具備極高的單位風險貢獻。

(2) **極端環境下的表現：** 2022 年市場環境惡劣，0050 與 0056 的夏普值均轉為大幅負值(-0.72 與 -0.89)，而本組合仍能維持在 -0.21，證明在下行風險管理上，AI 篩選機制能比單一市值型或高股息型策略更有效地保護資本效率。

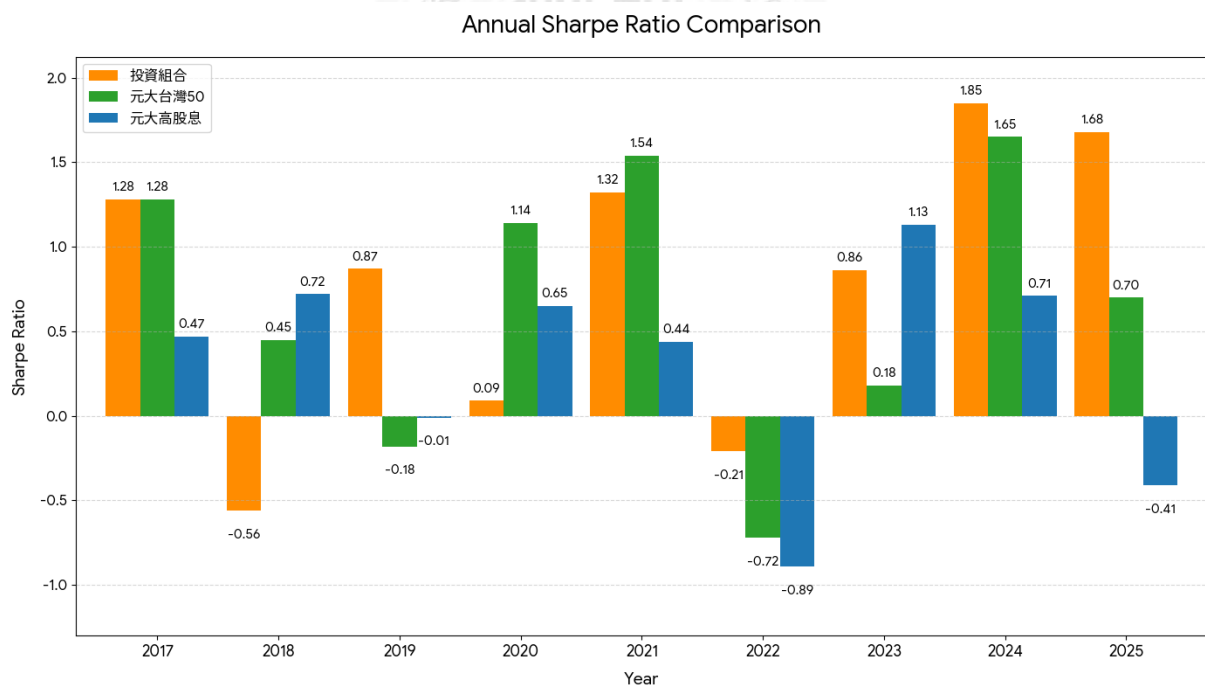


圖 11 夏普值 (Sharpe Ratio) 比較

3. 最大回撤（MDD）維度

最大回撤反映了策略對極端風險的承受能力，本研究觀察到以下特徵：

- (1) **非對稱風險控制：** 在 2023 年，本組合之 MDD 為 -13.07%，顯著優於 0050 的 -20.84%。這顯示在震盪回升行情中，AI 選股策略能透過「財務硬約束」濾網提前規避大幅波動標的。
- (2) **極端事件的挑戰：** 在 2020 年受疫情衝擊的極端情境下，本組合 MDD 為 -36.62%，略高於 0050 的 -30.48%。這反映出當市場發生結構性流動性危機時，以「成長動能」為主軸的 AI 選股策略可能面臨短期劇烈波動。然而，隨後的高報酬回補能力（Recovery）證明了該策略在風險釋放後具備強大的修復動能。

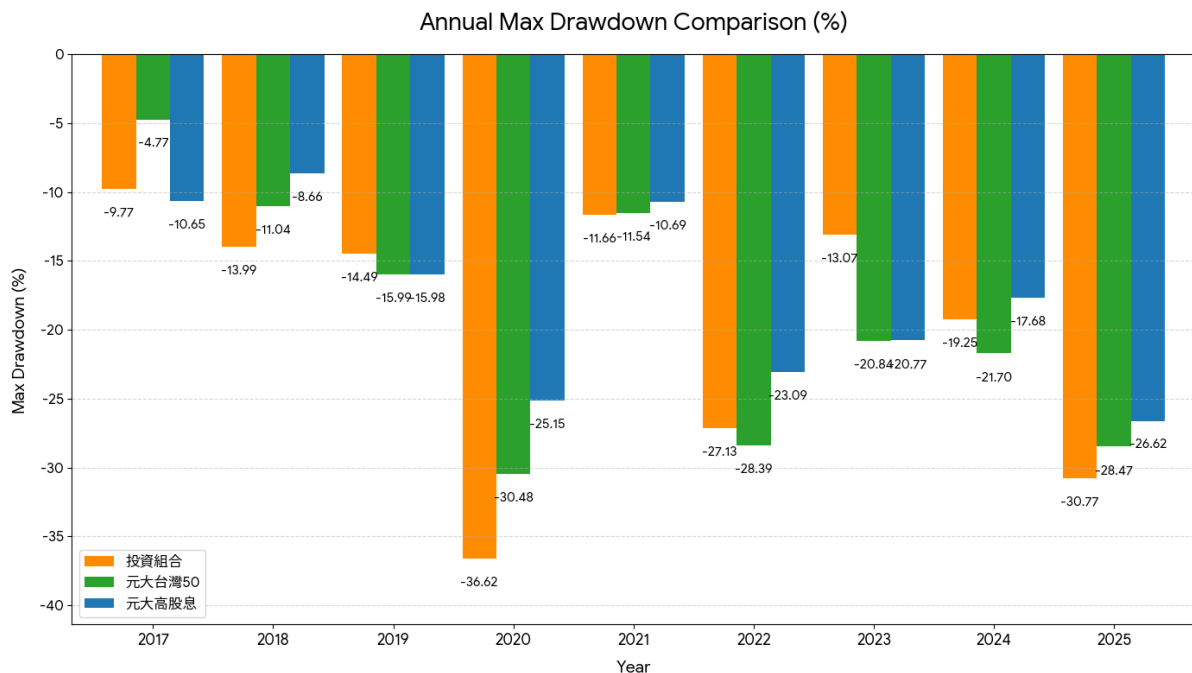


圖 12 最大回撤（MDD）比較

表 21 投資組合與市場標竿之年度績效對照表

	分類	年度報酬率	CAGR	年度最大回撤	年度夏普值	累計報酬率
2017	投資組合	17.63%	9.08%	-9.77%	1.28	246.71%
2017	元大台灣 50	21.32%	14.64%	-4.77%	1.28	243.80%
2017	元大高股息	11.39%	9.47%	-10.65%	0.47	126.48%
2018	投資組合	-5.96%	4.55%	-13.99%	-0.56	74.58%
2018	元大台灣 50	10.52%	14.38%	-11.04%	0.45	194.54%
2018	元大高股息	11.88%	9.79%	-8.66%	0.72	111.81%
2019	投資組合	17.73%	10.30%	-14.49%	0.87	251.86%
2019	元大台灣 50	1.31%	15.90%	-15.99%	-0.18	182.46%
2019	元大高股息	5.90%	10.13%	-15.98%	-0.01	97.20%
2019	國泰股利精選 30	3.43%	9.23%	-10.16%	0.03	86.20%
2019	元大台灣高息低波	6.44%	12.85%	-15.98%	-0.1	134.22%
2020	投資組合	3.78%	11.89%	-36.62%	0.09	178.61%
2020	元大台灣 50	32.75%	19.56%	-30.48%	1.14	194.16%
2020	元大高股息	19.16%	11.58%	-25.15%	0.65	93.86%
2021	投資組合	23.48%	17.84%	-11.66%	1.32	188.24%
2021	元大台灣 50	36.13%	16.73%	-11.54%	1.54	117.77%
2021	元大高股息	12.98%	10.08%	-10.69	0.439	62.12%
2021	國泰永續高股息	23.97%	13.95%	-9.35%	1.16	92.93%
2021	元大台灣高息低波	37.79%	16.12%	-11.30%	1.74	112.18%
2022	投資組合	-2.66%	16.18%	-27.13%	-0.21	115.99%
2022	元大台灣 50	-9.42%	13.62%	-28.39%	-0.72	67.39%
2022	元大高股息	-6.91%	10.18%	-23.09%	-0.89	47.88%
2023	投資組合	16.38%	25.86%	-13.07%	0.86	116.94%
2023	元大台灣 50	6.81%	21.73%	-20.84%	0.18	81.76%
2023	元大高股息	31.45%	16.76%	-20.77%	1.13	60.14%
2023	國泰永續高股息	29.37%	16.03%	-10.91%	1.37	57.13%
2023	元大台灣高息低波	26.01%	16.67%	-13.85%	1.12	59.75%
2024	投資組合	41.61%	40.77%	-19.25%	1.85	112.59%
2024	元大台灣 50	48.64%	31.22%	-21.70%	1.65	74.07%
2024	元大高股息	22.87%	11.72%	-17.68%	0.71	25.37%

2024	富邦台 50	48.95%	33.37%	-21.19%	1.89	79.93%
2024	國泰永續高股息	16.99%	9.48%	-13.26%	0.41	20.30%
2024	群益台灣精選高息	35.87%	15.90%	-17.80%	1.2	35.12%
2024	復華台灣科技優息	31.67%	13.94%	-16.82%	1.18	30.51%
2025	投資組合	48.49%	47.55%	-30.77%	1.68	43.50%
2025	元大台灣 50	18.80%	18.06%	-28.47%	0.7	18.80%
2025	元大高股息	2.35%	4.45%	-26.62%	-0.41	4.62%
2025	元大台灣價值高息	1.77%	2.91%	-	-	3.02%

4. 定期定額投資策略之實證分析：

鑑於實務投資人多採分批投入資金以規避單一進場點之擇時風險，單純之「首年單筆買進持有 (Lump-Sum)」回測雖能呈現複利潛力，但易受初始進場時點之偏差影響。為驗證本模型在真實投資行為下的穩健性，本節採用定期定額 (Dollar-Cost Averaging, DCA) 模型進行動態模擬。

(1) **實驗設計與參數設定：**本實驗模擬投資人於 2017 年至 2025 年期間，於每年年初固定投入資金，藉以評估在資產取得成本隨市場波動變動下，本策略之最終財富累積效率。

- **投入方式：** 每年年初定期定額投入。
- **投入金額：** 每年新台幣 200,000 元。
- **回測區間：** 2017 年至 2025 年 (共 9 期)。
- **總投入成本：** 新台幣 1,800,000 元。

表 22 投資組合定期定額績效與累積損益表

結算年度	年度報酬率	累積資產淨值(萬)	累積報酬率
2017	+17.63%	23.5	17.60%
2018	-5.96%	40.9	2.30%
2019	+17.73%	71.7	19.60%
2020	+3.78%	95.2	19.00%
2021	+23.48%	142.2	42.30%
2022	-2.66%	157.9	31.60%
2023	+16.38%	207.1	47.90%
2024	+41.61%	321.6	101.00%
2025	+48.49%	507.2	181.80%



圖 13 投資組合定期定額績效與累積損益表

(2) 績效對比與歸因分析：

- 實證結果顯示，本策略期末總資產達 507.2 萬元，相較市
值型標竿（0050）之結算金額多出 52.4 萬元，較高股息標
竿（0056）則大幅領先 173.9 萬元。數據證實，即便投資

成本因市場長期走揚而逐年墊高，本策略憑藉選股模型的高成長潛力，仍能展現強大的「成本抵銷能力」。

表 23 逐年定期定額累積報酬率

結算年度	投資組合	元大台灣 50	元大高股息
2017	17.60%	21.30%	11.40%
2018	2.30%	22.30%	18.30%
2019	19.60%	16.40%	18.80%
2020	19.00%	49.10%	35.90%
2021	42.30%	89.50%	45.50%
2022	31.60%	58.20%	28.40%
2023	47.90%	60.10%	63.40%
2024	101.00%	126.8%	91.00%
2025	181.80%	152.7%	85.20%

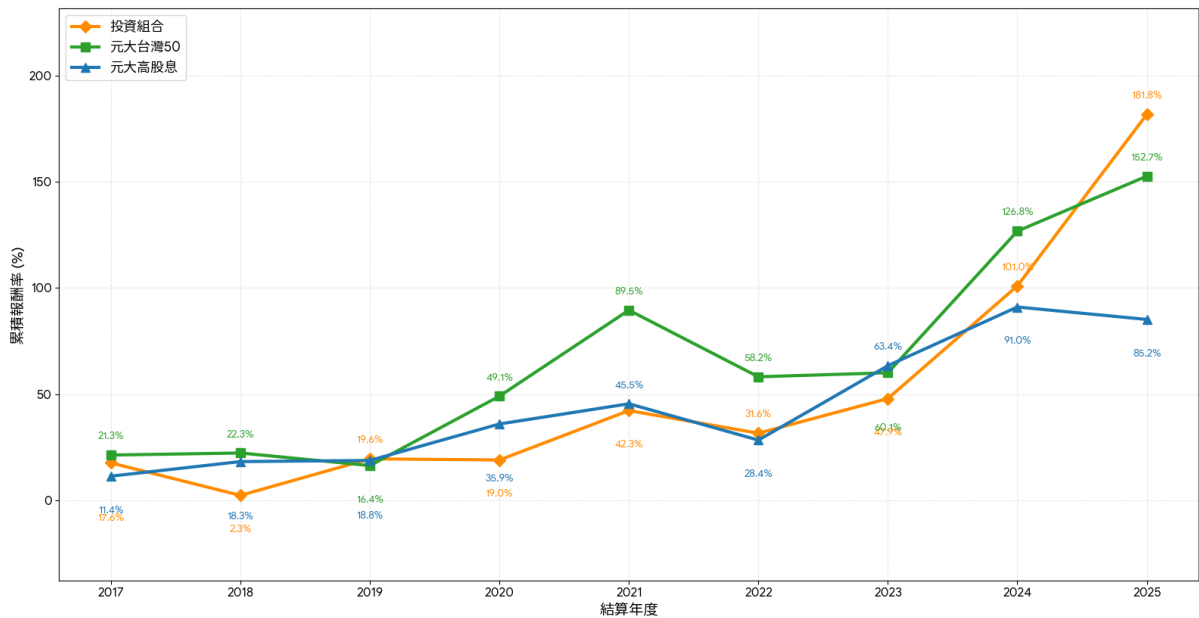


圖 14 定期定額累積損益比較

- 高股息策略之比較：總報酬假訂下高股息 ETF (0056) 在定期定額架構下之總報酬率為 85.18%，顯著落後於本策略的

181.78%，在長週期投資中，「股息發放」若缺乏「資本增值」支撐，將受限於複利速度。本模型透過機器學習篩選兼具營收動能與評價修復能力的個股，其產生的資本利得複利效應遠大於單純領息再投入之效益。

- 市值型策略之比較：相較於 0050 於 2024 至 2025 年受特定權值股驅動的漲幅，本策略在總獲利金額領先約 19.1% (327.2 萬 vs 274.8 萬)，定期定額平滑了進場成本，若策略僅依賴單一年份的偶然獲利，其長期財富累積曲線將趨近於大盤。然而本策略仍能穩定擴大領先幅度，證明模型所辨識在超額報酬具有時序上的持續性。

(3) **停損策略研究：**本研究曾針對主動賣出策略進行回測與比較分析，結果顯示，在整體投資組合中實際需要賣出的股票數量相對有限，約僅占全體樣本之少數（約 180 檔次中僅 12 檔左右）。此外，該等股票多已於回測期間出現較大幅度之帳面虧損（約介於 50% 至 80%），此時即使執行賣出，所能回收之資金規模有限，對整體投資組合之再配置效益不顯著。

(4) **報酬率計算：**本研究之所有績效評估皆採用還原股價（Total Return Price）進行報酬率計算，已完整納入現金股利之影響。相關績效比較基準（包含 ETF 如 0050 與 0056）亦採相同計算方式，以確保各投資標的間之報酬率比較具有一致性與公平性。所呈現之報酬表現為「含息總報酬」，可更真實反映長期投資情境下投資人所能實際獲得之整體報酬。

5. 實證小結：

研究實證結果顯示，相較於市值型（0050）或高股息型（0056）之傳統被動投資指標，本研究建構之 AI 選股模型透過多因子分流架構，不僅在多頭行情下能顯著擴張超額溢酬，更於 2022 年等空頭環境中展現出優異之風險抵禦韌性與下行風險控管。儘管受選股集中度與成長因子高波動特性之影響，導致 2020 年及 2025 年之最大回撤略高於市場標竿組，但其累積報酬率與夏普值之長期領先優勢，足以證成此動態策略在台灣市場具備高度之實戰價值。進一步透過定期定額（DCA）模型驗證發現，本策略不僅適用於單筆配置，對於採取分批投入之長期資產規劃者而言，更是極具效率之財富累積工具；其核心價值在於，選股模型具備之「高成長特性」能有效中和因市場長期走升而逐年墊高之持股溢價成本，且在面對 2020 與 2025 年之劇烈波動時，定期定額機制能發揮平滑持有成本之功效，將短期回撤轉化為低位累積單位數之契機，進而在隨後之市場反彈中釋放更強大之淨值修復力與獲利增長動能。

第五章 研究結論與未來研究方向

第一節 結論

本研究透過 EPS 預測模型，結合機器學習技術(XGBoost)與財務硬約束濾網，針對台灣股市進行為期九年的實證回測。研究結果不僅證實 傳統在 AI 輔助下 在獲利預測上的高精準度，更揭示了量化策略在長期財富累積中的結構性優勢。

1. 績效驗證：主動選股超越被動指數之實證

本研究實證結果顯示，該投資策略在績效表現與穩定性上均展現出顯著優勢：在定期定額（DCA）的實證架構下，本策略期末總資產達 507.2 萬元，較市值型標竿 0050 領先約 52.4 萬元，證實模型所辨識的特徵具備跨越牛熊的一致性，即使在市場高位投入，其選股增益仍能有效抵銷持倉成本；而在與高股息標竿 0056 的對比中，本策略更以 181.78% 的總報酬率完勝對手的 85.18%，強力支持了「總報酬導向」策略在長期複利效應上優於單純「現金流導向」的論點；此外，模型在多頭年份的正報酬個股占比高達 90%，且近 49% 的持股決策呈現核心標的重複選擇，顯示系統具備高度的邏輯連貫性與過濾市場雜訊的能力，能精準鎖定具備長期價值的優質資產，達成超越傳統被動投資的獲利廣度。

2. AI 賦能：兼顧「主動增益」與「被動省力」的安全性：

證實機器學習已成功實現從「統計工具」到「決策大腦」的範式轉換，賦予投資策略卓越的市場適應力。首先，模型展現了優異的動態資訊收斂能力，使預測命中率由 Q2 的 80.88% 顯著躍升至 Q3 的 86.15%；其次，在目標函數的優化上，實證發現針對「投資獲利」進行建模比單純追求「統計誤差（RMSE）最小化」更能產生實質績效，有效協助模型跳脫傳統數據陷阱。此外，透過非對稱架構成功辨識成長因子（3 年短視窗）與價值因子（5 年長視窗）的時序差異，解決

了特徵稀釋 (Feature Dilution) 問題，提升了模型對不同週期邏輯的辨識精度；最終在風險規避強化方面，配合濾網機制於 2018 年等市場回撤期成功收窄跌幅達 6% 以上，充分展現了機器學習在追求超額報酬的過程中，具備強大的防禦邊際與非理性風險過濾能力。

3. 投資策略與方法總結：

最終建構之 選股系統，在操作維度上與被動投資高度兼容，但在績效維度則展現了主動管理的爆發力，其核心邏輯可歸納為「**主動選股引擎，被動紀律執行**」。儘管底層算法複雜，系統卻呈現出高重複持有率與低換股頻率的特徵，使投資者能以類似 0050 的低決策負擔進行長線配置，無需受市場情緒干擾；實證顯示，AI 選股的高成長性與定期定額 (DCA) 的平均成本法具備極佳協同效應，能於多頭市場放大獲利空間，並在 2020 年與 2025 年等回撤期實現低位有效積累，成功平滑波動並驅動資產曲線趨於穩健。最終，本系統實現了從「跟隨市場」到「優化市場」的優化，在保留被動投資「簡單、長線、紀律」優點的同時，透過 AI 智慧篩選與財務硬約束剔除平庸標的，達成了資產增值極大化的戰略目標。

第二節 未來研究方向

本研究之實證基礎已確立「成長型」與「價值型」雙軌模型並行之必要性。後續研究可針對兩者之性格差異，進一步實施更具針對性的架構調整與特徵工程擴張，並深研科學化的獲利出場策略，以優化整體投資組合之穩定性。後續方向如下：

1. 模型分別研究：

後續研究建議可進一步深研「成長型」與「價值型」標的之非對稱優化路徑。在成長型模型的優化上，可嘗試將預測週期由「季」縮短至「月」以捕捉高波動市場中的短期動能，並探討將目標函數調整為「次月超額報酬預測」之效益，藉由提高調倉頻率與汰弱留強之機制，驗證其是否能更敏銳地捕捉趨勢轉折並避開財報空窗期之回檔風險；與此同時，針對價值型模型則建議著重於「均值回歸」特性的開發與「估值陷阱」之識別，透過拉長觀測視窗並結合獲利穩定性因子，釐清區分「市場錯殺」與「營運惡化」標的之關鍵特徵，藉此評估提升長線部位回報率與穩定性之潛力等。

2. 模型特徵工程的研究

在特徵工程的精進方面，未來研究建議可嘗試將維度由靜態財務指標擴張至動態市場偵測，探討納入量價動態、營收加速度與籌碼純度等因子之有效性。具體而言，後續研究可著重於開發「成交量異常變異」特徵並執行量價協同分析，以評估其是否能更精準地辨識趨勢轉折，並進一步驗證過濾無量空拉標的之可行性；同時，建議引入「營收加速度」指標，探討其在捕捉業績主升段初期企業時的預測潛力，並結合「法人持股集中度」數據來釐清爆量後的籌碼流向（如強手換手或籌碼潰散）對於強化策略安全邊際之實質貢獻等。

3. 獲利了結與出場機制之實證研究

後續研究建議可針對不同模型性格開發科學化的退出策略，以完善投資週期的最後環節。針對成長型標的，可進一步探討以「動能轉向」或「乖離率控制」作為動態止盈點之效益，驗證其在面對成長股高波動特質時，對於保護帳面利潤並規避獲利回吐之潛力；同時，針對價值型標的則建議研究以「估值修復至合理區間」為核心的出場準則，評估當股價回歸歷史本益比均值或達成目標價時，執行紀律化獲利了結對提升資金周轉效率與資產再配置價值之實質貢獻等。



附錄



參考文獻

英文文獻

1. Ben-David, I., Franzoni, F., & Moussawi, R. (2014). *Do ETFs increase volatility?* (NBER Working Paper No. 20071). National Bureau of Economic Research.
2. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383 – 417.
<https://doi.org/10.2307/2325486>
3. Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3 – 56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
4. Graham, B., & Dodd, D. L. (1934). *Security analysis: Principles and technique*. New York, NY: McGraw-Hill.
5. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223 – 2273.
<https://doi.org/10.1093/rfs/hhaa009>
6. Wang, J. (2022). Comparison of Random Forest, OLS, and XGBoost in stock return prediction. *Business Management and Economics*, 10(1), 1 – 11.
7. Mintarya, I., et al. (2023). Machine learning applications in stock price, revenue, and dividend prediction: A systematic literature review. *Journal of Financial Technology*, 5(2), 45 – 67.
8. Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, 38(Supplement), 1 – 41.
9. Lin, Z. (2014). High dividend yield stock portfolio versus active management performance. *Journal of Financial Markets*, 17(3), 123 – 145.
10. Njoroge, L., & Nasieku, T. (2021). Revenue growth and firm performance: Empirical evidence from Kenyan manufacturing firms. *International Journal of Business and Management*, 16(4), 89 – 102.

中文文獻

1. 蔡麗雀、蔡佩恂、徐川皓（2017）。價值投資在台灣股市之實證：股利殖利率與 F-SCORE。《修平學報》，35，125 - 144。
2. Wang, S. M. (2008)。台灣高股息 ETF 發行對其成分股流動性影響 [碩士論文，國立臺灣大學]。Airiti Library。
3. 陳鼎玟（2018）。應用機器學習與深度學習預測台股指數 [碩士論文，朝陽科技大學]。華藝線上圖書館。
4. 蕭弘宗（2024）。運用機器學習於資產定價模型之研究—以台灣股市為例 [博士論文，國立中山大學]。臺灣博碩士論文知識加值系統。
5. 謝曜宏、許資法、朱涵宇、林祝興（2019）。使用機器學習的智慧選股策略之研究。收錄於國立金門大學（主編），《NCS 2019 全國計算機會議論文集》（頁 467 - 472）。國立金門大學。
6. 田鉉元（2022）。以隨機森林法建構投資組合績效—以台灣股票市場為例 [碩士論文，國立清華大學]。臺灣博碩士論文知識加值系統。
7. 許雅筑（2024）。使用機器學習選股之投資績效研究—台灣股票市場之實證 [碩士論文，國立政治大學]。臺灣博碩士論文知識加值系統。
8. 白翔文（2012）。月營收資訊對股價報酬之影響 [碩士論文，正修科技大學]。臺灣博碩士論文知識加值系統。
9. 陳佑鎮（2013）。營收、盈餘及營業活動現金流量變動對股價報酬之影響 [碩士論文，東吳大學]。臺灣博碩士論文知識加值系統。
10. 陳佳鳳（2001）。股價報酬率與月營收成長率因果關係之實證研究 [碩士論文，淡江大學]。臺灣博碩士論文知識加值系統。
11. 黃鈺棋（2023）。股票報酬與盈餘市價比、淨值市價比、營收市價比之關係 [碩士論文，淡江大學]。華藝線上圖書館。
12. 吳幸姬、李顯儀（2006）。產業月營收變化與股價報酬的關聯性之研究。《管理科學研究》，3(1)，61 - 74。