

基於機器學習和假設檢定的動態股票交易策略

黃士峰^{1†} 朱苡雯²

¹ 國立中央大學統計研究所

² 國立高雄大學統計學研究所

摘 要

本文提出了一種基於非監督式學習的動態股票交易策略，動態股票交易策略包括三個步驟：第一步為收集每日股票價格和相關特徵，如：每日移動平均價格，交易量，波動率，RSI 等，並通過非監督式學習方法，輪廓係數和假設檢定了解每個時間點的狀態。第二步為在滾動視窗框架下使用訓練數據求解每個狀態的兩個最佳標準，確定接續的 20 個交易日之最佳停利點 (Optimal Take-Profit threshold, 簡記為 OTTP) 和最佳停損點 (Optimal Stop-Loss threshold, 簡記為 OTSL)。第三步是經由組合數個不同的機器學習方法辨識每個測試數據的所屬狀態，並藉由與該狀態相對應的 OTTP 和 OTSL，對每一測試數據所接續的 20 個交易日進行動態交易。我們將所提出的方法應用於 2012 年 1 月至 2022 年 3 月的 13 支台灣股票進行實證研究，數值結果表明在考慮交易成本的情形下，與買入持有策略和文獻上的現有策略相比，本文所提的投資方法具有較為穩健的投資績效表現。

關鍵詞：假設檢定、機器學習、最佳停利點、最佳停損點。

JEL classification: C58, G11.

[†]通訊作者：黃士峰
E-mail: huangsf@ncu.edu.tw

1. 緒論

本文結合機學習與統計檢定的技術，提出一套獲益績效穩健的「投資輔助決策系統」，並以台灣股票市場的資料為例進行實證分析，深入探討機器學習與統計檢定在所提出的投資策略中，所扮演的關鍵角色與其在投資績效的影響。眾所周知，預測股價未來走勢是一項困難的挑戰，但卻又是一項每一位投資人進行投資決策時不可或缺的預測資訊 (Tsay 2010; Huang and Lin 2018; Huang *et al.* 2020; Lai *et al.* 2025)。從統計的角度看，既然是預測就需要面臨預測誤差，如何有效地降低以及管理預測誤差，便成為重要的課題。在降低預測誤差（或提升預測精確度）方面，適當地應用可收集到的解釋變數建立合乎資料動態行為的統計模型，一直是統計學家與資料科學家們努力的目標，在這個方向上，機器學習技術可協助自大量可能的解釋變數中擷取有用資訊。另一方面，本文亦發現適當地應用統計檢定方法則可以在建立投資策略時有效地降低以及管理預測誤差。

由於經濟狀態的動態改變是影響股價走勢的重要因素之一，因此，本文提出透過收集每一個交易日各式各樣可能影響股價走勢之潛在因子（或稱為特徵）的資料，並應用非監督式學習（unsupervised learning）對每一交易日進行聚類分析（clustering analysis），每一個聚類或集群（cluster）則可視為受同一經濟狀態的影響，其股價走勢應該相當接近，因而每一集群中大部分的交易日應該都適合相同的操作（做多、做空、或不進場）模式，對於少部分不符合的交易日，則可應用停損、停利機制，進行風險控管。在實務操作上，設定停損是為了在趨勢判斷錯誤時及時止損，防止虧損增加，而設定停利點是為了確保能及時獲利了結，防止因趨勢突然改變，造成獲利受損 (Wu and Chung 2019)。為了反映不同的經濟狀態對於投資策略的影響，我們亦提出對於每一個集群應分別估計其最佳停損或停利點的方法，同時應用適當的統計檢定方法，在訓練集資料中辨識出適合每一集群的操作策略。在測試集資料中，當收集到一筆新的特徵資料時，則透過分類（classification）程序，觀察新資料與訓練集資料中哪一個集群中的資料行為較為相似，便將其分配到該群，並以該群的投資策略（如：最佳停利點、最佳停損點、操作策略等）進場設立投資部位。

本文所提出的「投資輔助決策系統」旨在協助投資人進行每日股票的投資決策，如：做多、做空、不操作。該系統的幾項重要程序如下：首先將資料以 Wang *et al.* (2017) 所提出的非監督式學習方法—基於多核的相似性學習 (Single-cell Interpretation via Multi-kernel Learning, 簡記為 SIMLR)，對訓練集資料進行聚類分析，並針對聚

類分析後的各集群分別設定每日最佳停利點和最佳停損點並檢定各群適合的操作方式，接著利用隨機森林 (Random Forest)、極限梯度提升 (eXtreme Gradient Boosting, 簡記為 XGBoost)、與光梯度提昇機 (LightGBM, 簡記為 LGBM) 進行測試集資料的類別辨識。此外，為了合理分配手上現有資金同時做好風險的控管，以便使整個投資過程更加穩定，無須擔心資金匱乏的目標，我們參考 Syu *et al.* (2021) 的方法，若當日有進場訊號，則該次操作至多維持 20 天，若 20 天之中的累積報酬皆無抵達所設定的最佳停損停利點，則在第 20 天時會強制平倉，如此便能更加靈活的使用手上僅有的資金。在實證研究中我們使用 2012 年 1 月至 2022 年 3 月台灣各產業總共 13 支股票的資訊 (收盤價、交易量、RSI、K-value、移動平均、波動率) 進行分析。在操作整個投資時會將資金切成 10 (或 20) 份，若測試資料出現進場訊號，則該日使用 1/10 (或 1/20) 份資金，此操作也能適時降低風險。最終結果使用年化報酬率和最大回撤以及夏普率做為投資績效的比較基準。實證分析的結果可發現相較於文獻上不同的投資策略，本文提出的投資方法在確保獲得一定程度的報酬下，所承擔的風險幾乎都最低，故所提出的投資輔助決策系統偏向於建議獲益績效穩健的投資策略。

本文後續章節安排如下，第二節回顧相關文獻，主要針對本文所使用到文獻上的相關方法，進行簡要的介紹。第三節介紹所提出之投資輔助決策系統的建立流程，主要包含如何利用 Wang *et al.* (2017) 提出的 SIMLR 對資料進行聚類分析，以及如何對聚類分析後的各集群設定最佳停損停利點，再利用威爾卡森檢定決定各群之操作策略，最後再利用 3 個機器學習的方式對測試集資料進行分類。第四節為實證研究，介紹一系列的投資流程及評估指標，並比較不同策略的投資績效。第五節為結論與討論，部分圖表則置於附錄中。

2. 文獻回顧

節針對本研究所採用的 3 個主要元素：聚類方法、動態停損停利、與分類 (classification) 方法，簡介相關的文獻。

2.1 基於多核的相似性學習 (SIMLR)

當資料存在噪音以及缺失值時，傳統的無監督式學習演算法，例如：K-平均演算法 (K-means)、階層式分群法 (hierarchical clustering) 等，其聚類效果往往不盡理想，針對此一問題，Wang *et al.* (2017) 提出一個利用多核 (multiple kernels) 方式學習樣本之間的距離度量以便建構相似矩陣，再進行降維聚類的演算法，簡稱為 SIMLR

演算法。SIMLR 計算樣本間相似性時所採用的目標函數如下：

$$\min_{S, L, w} \sum_{i, j} D(x_i, x_j) S_{ij} + \beta \|S\|_F^2 + \gamma \text{tr}(L^T (I_N - S) L) + \rho \sum_l w_l \log w_l. \quad (1)$$

對所有的 (i, j) 需滿足 $D(x_i, x_j) = \sum_l w_l D_l(x_i, x_j)$ 、 $w_l \geq 0$ 、 $\sum_l w_l = 1$ 、 $L^T L = I_C$ 、 $\sum_j S_{ij} = 1$ 與 $S_{ij} \geq 0$ ，其中 S_{ij} 表示樣本之間的相似性， x_i 表示樣本 i 所有特徵形成的向量， $D_l(x_i, x_j)$ 表示第 l 種度量樣本 i 和 j 距離的單核 (single kernel) 量測方法， $D(x_i, x_j)$ 定義為 $D_l(x_i, x_j)$ 的加權和，其中權重為 w_l ，用以表示樣本 i 和 j 的多核距離， $\|S\|_F$ 為 S 的 Frobenius 範數， L 為輔助強制降低 S 秩 (rank) 的低維矩陣 (low-dimensional matrix)， β 與 γ 為非負調整 (tuning) 參數， I_a 表示維度為 a 的單位矩陣。在 (1) 式中，第一項主要反應當樣本間的距離很大時，則相似性應該很小；第二項為正則化項，可避免學習到的相似性太接近單位矩陣；第三項則透過改變 S 的秩，以達到控制聚類數量的效果，換言之，假設輸入的資料能夠分成 C 個聚類，則 S 最終應該要是秩為 C 的相似矩陣；第四項的想法與熵 (entropy) 的概念相似，用以表示隨機變量不確定性，能判斷訊息量是否充足，並控制未知訊息。

許多生醫領域文獻的實證研究發現，SIMLR 具有良好的聚類表現 (Ramazzotti et al. 2018; Scavuzzo et al. 2018; Wang et al. 2018; Vamathevan et al. 2019; Dara et al. 2022; Yuan et al. 2024)。因此，本文提出在收集諸多影響股價走勢的潛在特徵後，將 SIMLR 應用於不同交易日的經濟狀態 (如：漲、跌、大漲、大跌、持平...等) 聚類分析。

2.2 具有自適應關聯規則的投資組合系統

Syu et al. (2021) 提出了一個結合關聯規則 (association) 和停損停利點設定的投資組合系統，透過分析「所觀測到的技術指標」與「多頭部位率先抵達停利點」之間的關聯，建立投資策略。流程如圖 1。

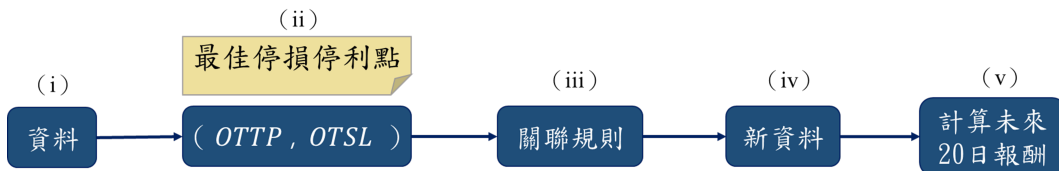


圖 1：自適應的平仓機制流程圖。

由於 Syu *et al.* (2021) 採用關聯規則 (Agrawal *et al.* 1993) 進行分析，因此在資料預處理時需對所觀測到的技術指標 (即圖 1 中的 (i) 資料與 (iv) 新資料)，包括相對強弱指數 (Relative Strength Index, 簡記為 RSI)、隨機指標中的快線 (簡記為 K 值)、成交量、收盤價斜率、近 5 日移動平均價斜率、近 5 日波動斜率等，進行離散化。其中由於 RSI 和 K 值是 $[0, 1]$ 中的連續值，因此將每個值離散為 4 個不相交的類別： $[0, 0.25)$ 、 $[0.25, 0.50)$ 、 $[0.50, 0.75)$ 和 $[0.75, 1)$ ，成交量指標也被離散化為 4 個不相交的類別： $[-\infty, 0.5)$ 、 $[0.5, 1)$ 、 $[1, 1.5)$ 以及 $[1.5, \infty)$ ，而收盤價斜率和近 5 日移動平均價斜率以及近 5 日波動斜率離散為 2 個類別： $[-\infty, 0)$ 、 $(0, \infty)$ 。若所觀測到的資料屬於該類別記為 1，否則記為 0。因此，每一個交易日共有 18 個離散型特徵。另一方面，在感興趣事件 (即「多頭部位率先抵達停利點」) 的標記方面，在給定每日的未來 20 天之最佳停利點 (OpTimal Take-Profit threshold, 簡記為 OTTP) 及最佳停損點 (OpTimal Stop-Loss threshold, 簡記為 OTSL) 下，若未來 20 天先抵達 OTTP，則該日標記為 1，並停利出場；若先抵達 OTSL，則標記為 -1，並停損出場；此外則標記為 0，並於第 20 天平倉出場。

基於上述的每一交易日標記與 18 個離散型特徵，從任一給定的時間 t 之最近 250 筆歷史交易資料的每日 18 個離散型特徵集合中，找出導致時間 s 時感興趣事件 $Y_s = \{\text{時間}s\text{-之未來 20 天率先抵達 OTTP}\}$ 發生的關聯規則， $s = t - 1, \dots, t - 250$ 。令 $Rule_t = \{X_s : X_s \rightarrow Y_s, s \in \{t - 1, \dots, t - 250\}\}$ 表示時間 t 時所對應的關聯規則集合，若時間 t 時所觀測到新資料的離散型特徵屬於 $Rule_t$ ，則建議當日進場做多。

在計算每一交易日之未來 20 天的 OTTP 與 OTSL 方面，Syu *et al.* (2021) 提出一基於過去 250 天累積報酬的計算方法，簡述如下。設定停利點 (Take-Profit threshold, 簡記為 TP) 之目的為確保獲利，設立停損點 (Stop-Loss threshold, 簡記為 SL) 則是為了避免更大的損失，然而在不同的時間點，每支股票的特徵不盡相同，因此可設定動態 TP 及 TSL 配合每日不同的股票特徵。令 $RR_{t-j,t-1,TP,SL}$ 為在第 $t-j$ 天買進含有 TP 及 SL 設定的股票收益率，且至少在第 $t-1$ 天平倉，則第 t 天的 $OTTP_t$ 及 $OTSL_t$ 為：

$$\arg \max_{TP,SL} \sum_{j=1}^{250} RR_{t-j,t-1,TP,SL},$$

其中 TP 可以設定為 3% 或 5% 或 8% 或 10% 或 15% 或 20%，SL 可以設定為 -3% 或 -5% 或 -8% 或 -10%，計算在各種 TP 及 SL 設定下的總報酬，並挑選最大值

的組合做為第 t 天的 $OTTP_t$ 及 $OTSL_t$ 。此方法藉由歷史資料判斷未來的 $OTTP$ 及 $OTSL$ ，該設定方式更加彈性，隨著過去 250 筆資料不同，定出相對應適合的值。

將 Syu et al. (2021) 所提出上述的方法應用在實際台股資料，由於其過程中將連續資料離散化的步驟會損失部分可用資訊，且若採用 SIMLR 分類法是對高維且多型態的特徵進行聚類效果較佳，因此本文採用保留資料原本的型態。此外，本文進一步提出在設定 $OTTP$ 及 $OTSL$ 時依照各個時間點的經濟狀態（漲、跌、大漲...）做分群後，再依各群設定最佳停利停損點，以便挑選出具有隨時間與經濟狀態動態調整優點的 $OTTP$ 及 $OTSL$ 。

2.3 分類模型

有別於 Syu et al. (2021) 採用關聯規則對新資料進行是否進場做多的建議，本文應用以下 3 種機器學習的分類方法，對新資料進場做多、進場做空、或是不進場等 3 種可能的操作，進行建議：

1. 隨機森林 (Random Forest)：隨機森林的概念是使用引導聚集 (bagging) 及隨機挑選特徵的方式所產生出的演算法，由於是透過引導聚集算法 (bootstrap aggregating) 抽樣，因此有機會不挑到有噪聲 (noise) 的資料，可以平衡誤差，降低模型的不穩定性 (Breiman 2001)。
2. 極限梯度提升 (eXtreme Gradient Boosting, 簡記為 XGBoost)：XGBoost 是一種可以大規模並行提升樹 (boosting tree) 的工具，由於並行和分佈式計算，因此 XGBoost 的優勢在於執行速度快速且靈活，也能夠進行多元分類，在工業、生物及醫學的大數據方面有也有不錯的研究結果 (Chen and Guestrin 2016)。
3. 光梯度提昇機 (LightGBM, 簡記為 LGBM)：LGBM 主要利用決策樹為基底不斷迭代訓練，最終取得最佳的模型。由於該演算法進行了優化使得訓練速度變快，並且有效降低消耗的資源，此方法減少了內存消耗並加快了訓練過程，此方法也能進行多元分類，在工業、生物及醫學的大數據方面有不錯的研究結果 (Ke et al. 2017)。

3. 研究方法與流程

3.1 資料介紹

本文考慮在不同的經濟環境下，以台灣股票市場的成分股為例，預測其未來股價趨勢並探討所提出操作策略的投資績效，所採用的股票包含鋼鐵業的中鋼、被動元件產業的國巨、光電業的大立光、面板產業的群創，電腦週邊設備業的廣達和研華、金融業的國泰金和永豐金、半導體產業中晶圓代工的台積電、記憶體族群的南亞科、IC設計族群的瑞昱、聯發科以及普誠等台股各產業 13 支股票。收集每一檔個股自 2012 年 1 月至 2022 年 3 月間的每日收盤價（簡記為 C ）、相對強弱指數（Relative Strength Index，簡記為 RSI）、隨機指標（Stochastic Oscillator，簡記為 KD）中的快線（K 值）與慢線（D 值）、當日交易量（簡記為 V ）、近 5 日移動平均（moving average，簡記為 MA）、近 5 日價格波動（簡記為 VOL）等技術指標進行分析，其中近 5 日價格波動定義為近 5 日收盤價之標準差。其中 RSI、K 值、與 D 值的計算公式簡述如下：

$$RSI_t = \frac{U_{t,s}}{(U_{t,s} + |D_{t,s}|)},$$

其中 RSI_t 為第 t 天的 RSI， $U_{t,s}$ 為前 s 天中所有價格差為正的平均數， $D_{t,s}$ 為前 s 天中所有價格差為負的平均數。此外，每日 K 值的計算乃基於當日的未成熟隨機值（Raw Stochastic Value，簡記為 RSV），其計算公式如下：

$$RSV_t = \frac{P_t - L_n}{H_n - L_n} \times 100,$$

其中 RSV_t 為第 t 天的未成熟隨機值， P_t 為第 t 天的收盤價， L_n 為近 n 日中的最低收盤價， H_n 為近 n 日中的最高收盤價。根據 RSV 可得到以下的 K 值與 D 值定義：

$$\begin{aligned} K_{v_t} &= K_{v_{t-1}} \times \frac{2}{3} + RSV_t \times \frac{1}{3}, \\ D_{v_t} &= D_{v_{t-1}} \times \frac{2}{3} + K_{v_t} \times \frac{1}{3}, \end{aligned}$$

其中 K_{v_t} 與 D_{v_t} 分別為第 t 天的 K 值與 D 值。

將以上得技術指標進行差分可進一步得到： $\Delta RSI_t = RSI_t - RSI_{t-1}$ 、 $\Delta K_{v_t} = K_{v_t} - K_{v_{t-1}}$ 、 $\Delta D_{v_t} = D_{v_t} - D_{v_{t-1}}$ 、 $\Delta VOL_t = VOL_t - VOL_{t-1}$ 、 $\Delta V_t = V_t - V_{t-1}$ 、 $\Delta MV_t = MV_t - MV_{t-1}$ ，以及 $\Delta C_t = C_t - C_{t-1}$ 等指標，令 $\theta_t = (\Delta RSI_t, \Delta K_{v_t}, \Delta D_{v_t}, \Delta VOL_t, \Delta V_t, \Delta MV_t, \Delta C_t)$ 。本文採用將最近 20 個交易日的標準化 θ_t 視為構

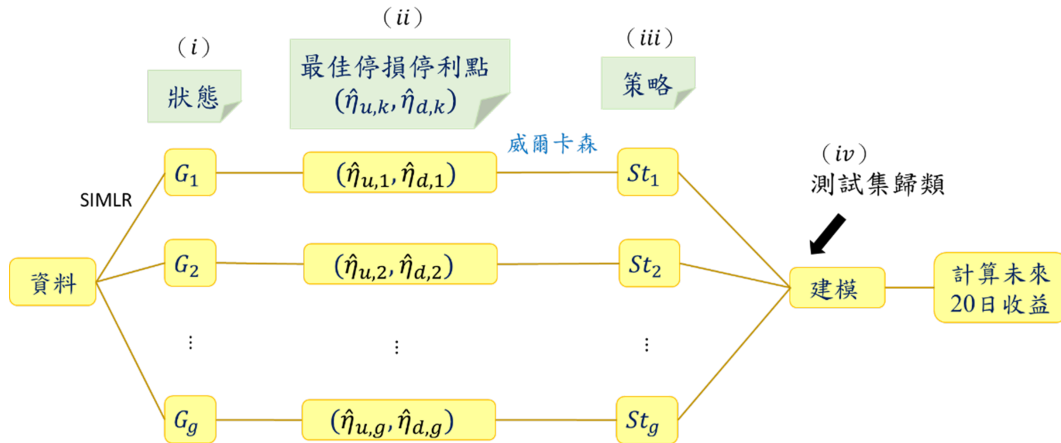


圖 2：輔助決策系統流程圖。

成訓練集所需的特徵集合，以時間 t 為例，為了消除各項特徵因單位不同所造成的影響，我們對 θ_t 中的每一個元素分別進行標準化，同時將標準化後的結果記為 \mathbf{a}_t 。特別地，在本研究所實施的實證過程中，我們發現無論是否採用 ΔDv_t 這項指標，並不影響接下來 3.2 節步驟 (i) 的聚類分析結果，可能的原因為 D 值是慢線，其變化速度較 K 值緩慢，經過差分後 ΔDv_t 的變化不如 ΔKv_t 明顯，故 ΔDv_t 在採用 20 個交易日的資料進行聚類分析時，其貢獻被其他特徵所取代。因此，以下的介紹將移除 ΔDv_t 這項指標以簡化敘述，從而使得在第 t 天的特徵集合 $\{\mathbf{a}_{t-19}, \mathbf{a}_{t-18}, \dots, \mathbf{a}_t\}$ 中，共包含 120 個標準化後的特徵。

3.2 流程

圖 2 為所提出之輔助決策系統的流程圖，共包含四個主要步驟。首先將特徵資料整理並切割出訓練集，透過前三個步驟進行模型訓練，包括 (i) 對各個時間點的經濟狀態進行聚類分析，分成 G_1, \dots, G_g 個群，(ii) 計算各群做多及做空時的最佳停利點 (以 $\hat{\eta}_{u,k}$ 表示) 及最佳停損點 (以 $\hat{\eta}_{d,k}$ 表示)，以及 (iii) 透過統計檢定方法且結合步驟 (ii) 中最佳停利與停損點的設定，判斷每一群適合的操作策略。第四步驟為當得到新資料時，利用 2.3 節所介紹的 3 個分類方法，判斷新資料屬於 $\{G_1, \dots, G_g\}$ 當中的哪一個群，從而根據該群在步驟 (iii) 中所得到的建議進行操作。上述流程的細節如下：

步驟 (i)：經濟行為狀態分類

將上述 3.1 節所提到的多維特徵透過 SIMLR 對經濟趨勢狀態進行聚類分析。假

設將時間區間 $[T - N + 1, T]$ 中的資料分為 g 個群，記為 G_1, \dots, G_g ，每一個聚類可能的操作策略為做多、做空、與不進行任何操作等三種可能，同時以輪廓係數法 (silhouette coefficient) 決定聚類數量。此步驟主要的目的為將多維特徵值較為接近的時間點集中在同一個群，接下來再透過以下步驟，決定各群之最佳停利點及最佳停損點。

步驟 (ii)：設定最佳停利停損點

對每一經濟趨勢狀態聚類 G_k ，令 $\eta_{u,k}$ 與 $\eta_{d,k}$ 分別代表該聚類的停損與停利點，由於我們的目標是求出未來 20 個交易日內，報酬會先抵達 $\eta_{u,k}$ 或是 $\eta_{d,k}$ 或是皆無，因此令

$$P_{t,t'} = \begin{cases} K_{1t} \{S_{t'}(1 - C_S) - S_t(1 + C_B)\}, & \text{若為長部位} \\ K_{2t} \{S_t(1 - C_S) - S_{t'}(1 + C_B)\}, & \text{若為短部位} \end{cases}$$

表示時間 t 時所設立的投資部位，在考慮交易手續費之下，於時間 t' 時平倉的實質收益，其中 C_B 與 C_S 分別為買入與賣出的手續費， $K_{1,t}$ (或 $K_{2,t}$) 代表在時間 t 時偵測到買入 (或賣出) 訊號時，買入 (或賣出) 的股數。其次，令

$$R_{t,t'} = \begin{cases} \frac{P_{t,t'}}{K_{1t}S_t(1 + C_B)}, & \text{若為長部位} \\ \frac{P_{t,t'}}{K_{2t}S_t(1 - C_S)}, & \text{若為短部位} \end{cases}$$

表示時間 t 時所設立的投資部位，在考慮交易手續費之下，於時間 t' 時平倉的實質報酬率 (realized return)。令 $\eta_{u,k} > 0$ 與 $\eta_{d,k} < 0$ 分別代表停利與停損點，若在時間 t 時建立一買入或賣出部位 $t_u = \min_{\{t':t'>t\}} R_{t,t'} = \eta_{u,k}$ 與 $t_d = \min_{\{t':t'>t\}} R_{t,t'} = \eta_{d,k}$ 分別代表該部位停利或停損時的時間點。

接下來，我們提出透過以下的目標函數，進行 $\eta_{u,k}$ 與 $\eta_{d,k}$ 的估計：

$$(\hat{\eta}_{u,k}, \hat{\eta}_{d,k}) = \arg \max_{(\eta_{u,k}, \eta_{d,k}) \in \Theta} \sum_{t \in G_k} R_{t,t'}(\eta_{u,k}, \eta_{d,k}) I_{\{t' \leq T\}},$$

其中

$$R_{t,t'}(\hat{\eta}_{u,k}, \hat{\eta}_{d,k}) = \begin{cases} \eta_{u,k}, & \text{若 } t' = t_u < \min(t_d, t + 20) \\ \eta_{d,k}, & \text{若 } t' = t_d < \min(t_u, t + 20) \\ R_{t,t+20}, & \text{若 } t' = t + 20 < \min(t_u, t_d) \end{cases},$$

t_u 與 t_d 分別為停利與停損的時間點，且任一部位至多維持 20 個交易日， $I_{\{t' \leq T\}}$ 表示無論是否達到停利或停損條件，所有未平倉部位皆須在時間 T 時平倉。經此步驟計算能夠得到各群做多及做空的 $\hat{\eta}_{u,k}$ 與 $\hat{\eta}_{d,k}$ ，也能將各群中每一筆資料在做多時所產生的收益記做向量 \vec{R}_k^u ，而在做空時所產生的收益記做向量 \vec{R}_k^d ，下一步可結合此資訊，透過檢定方法判斷適合各群的操作策略。

步驟 (iii)：判斷該群操作手法

本文採用威爾卡森檢定 (Wilcoxon test；Wilcoxon 1992) 對各群適合的投資策略進行判斷。具體做法為：使用 G_k 群的 \vec{R}_k^u 以及 \vec{R}_k^d 透過以下檢定程序決定該群適合的投資策略。首先考慮以下的假設檢定問題：

$$H_0 : x_q^u \leq 0 \quad \text{vs} \quad H_1 : x_q^u > 0$$

其中 x_q^u 為 \vec{R}_k^u 的第 q 百分位數，若檢定結果顯著，即代表有顯著證據支持該群若是做多，若 $q = 40\%$ ，則有超過 60% 的報酬率皆為正，故判斷該群適合進行多頭操作。直觀上，在上述檢定問題中，較小的 q 設定值將導致不易拒絕 H_0 ，從而可能錯過許多進場投資的機會；反之，設定較大的 q 值則會降低本步驟的影響力，無法達到本步驟降低進場投資風險的設立初衷。

若上述的檢定結果不顯著，則考慮另一個方向的假設檢定問題：

$$H_0 : x_q^d \geq 0 \quad \text{vs} \quad H_1 : x_q^d < 0,$$

其中 x_q^d 為 \vec{R}_k^d 的第 q 百分位數，若此時檢定結果顯著，即代表有顯著證據支持該群適合進行空頭操作。倘若以上兩種假設檢定問題的檢定結果皆不顯著，則判斷該聚類不應適合進場進行任何操作。

步驟 (iv)：建立預測模型

當觀測到一筆新的測試資料時，我們提出使用 Random Forest、XGBoost、和 LGBM 等三種分類器，分別記為第 $i = 1, 2, 3$ 個分類器，分別判斷測試資料屬於訓練集中的哪一個群，並透過給予每一個分類器不同的權重進行加權合，得到該筆測試資料最終的歸類結果。上述權重的計算方式如下：

$$w_i = \frac{Acc_i}{\sum_{j=1}^3 Acc_j}, \quad i = 1, 2, 3,$$

其中 Acc_i 表示第 i 個分類器在訓練集資料的分類正確率 (accuracy)， $i = 1, 2, 3$ 。此

外，我們亦能計算第 i 個分類器將一筆測試資料歸類到第 j 個群的機率，將此機率記為 P_{ij} ，其中 $P_{ij} \geq 0$ ， $\sum_{j=1}^g P_{ij} = 1$ ， $i = 1, 2, 3$ 。結合 w_i 與 P_{ij} ，計算新資料被分到第 j 群之機率：

$$\hat{P}_j = \sum_{i=1}^3 w_i P_{ij}, j = 1, \dots, g。$$

最後將該測試資料歸類給第 $\hat{j} = \arg \max_{j=1, \dots, g} \hat{P}_j$ 群，接下來便可採用該群的操作建議進行投資。

4. 實證研究

4.1 操作流程

本文使用的訓練集從第 1 天開始累積，第一次訓練集為第 1 日~750 日，則第 751 日~1000 日做為測試集進行分類用以決定各日的操作及 $\hat{\eta}_{u,k}$ 和 $\hat{\eta}_{d,k}$ ，第二次訓練集為第 251 日~1000 日，則第 1001 日~1250 日做為測試集，以此類推，從 2012 年 1 月至 2022 年 3 月此操作總共進行了 7 次，如圖 3 所示。分別計算測試集中每個交易日在未來 20 天中，報酬會先高於 $\hat{\eta}_{u,k}$ 或是低於 $\hat{\eta}_{d,k}$ ，又或是皆未抵達，並計算出累積報酬以觀察投資績效。

假設初始資金為 V_0 ，本文的操作方法為將資金切成 M_0 等份，若當日經判斷後有進場訊號，則每次進場使用的資金量固定為 $u = \lfloor \frac{V_0}{M_0} \rfloor$ ， $\lfloor a \rfloor$ 表示小於或等於 a 的最大整數，若第 t 天判斷進場做多，則當日可買 $K_{1,t} = \lfloor u / \{S_t(1 + C_B)\} \rfloor \times I_{\{B,t\}}$ 股，其中 C_B 表示購買股票的手續費， S_t 表示第 t 天的收盤價， $I_{\{B,t\}}$ 表示第 t 天的買入信

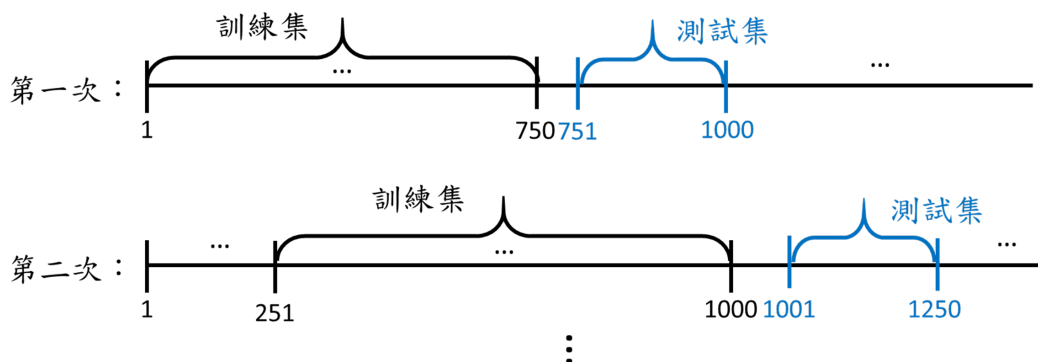


圖 3：訓練集和測試集天數示意圖。

號，若 $I_{\{B,t\}} = 1$ 表示第 t 天存在進場買股票的訊號， $I_{\{B,t\}} = 0$ 表示第 t 天不進場買股票。若第 t 天判斷進場做空，則當日可賣 $K_{2,t} = \lfloor u/\{S_t(1 - C_S)\} \rfloor$ 股，其中 C_S 表示賣出股票的手續費， $I_{\{S,t\}}$ 表示第 t 天的賣出信號，若 $I_{\{S,t\}} = 1$ 表示第 t 天存在進場賣股票的訊號， $I_{\{S,t\}} = 0$ 表示第 t 天不進場賣股票， $I_{\{B,t\}}$ 和 $I_{\{S,t\}}$ 只會等於 0 或 1。假設 \tilde{V}_t 表示在時間 t 的總資產，由未實現收益 (V_t^U) 及已實現收益 (V_t) 相加得到，亦即由 $\tilde{V}_t = V_t^U + V_t$ 得到時間 t 的總資產。我們使用 N_t 表示第 t 天結束時持有的股票數量，由此可推得 $V_t^U = N_t S_t$ ，綜觀上述 \tilde{V}_t 的計算方式可更新為 $\tilde{V}_t = N_t S_t + V_t$ ，由於交易機制是按日設計，因此我們忽略了無風險利息的影響。

設在第 t 天進場做多，當日的累積股票數量為 $N_t = N_{t-1} + K_{1,t}$ ，當日的累積已實現收益為 $V_t = V_{t-1} - K_{1,t} S_t (1 + C_B)$ 。相對的，若第 t 天進場做空，可得 $N_t = N_{t-1} - K_{2,t}$ 和 $V_t = V_{t-1} + K_{2,t} S_t (1 - C_S)$ 。總結上述針對做多及做空計算 N_t 和 V_t 的式子，若第 t 天有進場訊號，即 $I_{\{B,t\}} = 1$ 或 $I_{\{S,t\}} = 1$ ，可得到下式：

$$\begin{aligned} N_t &= N_{t-1} + (K_{1,t} I_{\{B,t\}} - K_{2,t} I_{\{S,t\}}), \\ V_t &= V_{t-1} - \{K_{1,t} (1 + C_B) I_{\{B,t\}} - K_{2,t} (1 - C_S) I_{\{S,t\}}\} S_t. \end{aligned} \quad (2)$$

若第 t 天沒有進場訊號，即 $I_{\{B,t\}} = I_{\{S,t\}} = 0$ ，但我們在第 $t - j$ 天， $j = 1, \dots, 20$ ，有進場操作，且正好在第 t 天平倉，可得下式：

$$\begin{aligned} N_t &= N_{t-1} - \sum_{j=1}^{20} (K_{1,t-j} - K_{2,t-j}) I_{t-j,t}^{close}, \\ V_t &= V_{t-1} + \sum_{j=1}^{20} \{K_{1,t-j} (1 - C_S) - K_{2,t-j} (1 + C_B)\} S_t I_{t-j,t}^{close}, \end{aligned} \quad (3)$$

其中 $I_{t-j,t}^{close}$ 表示假設在第 s 天進場操作並在 $t > s$ 時平倉的指標函數。

根據 (2) 及 (3) 式，可將總收益 (V_t^T) 可整合成 $\tilde{V}_t = N_t S_t + V_t$ ，其中

$$\begin{aligned} N_t &= N_{t-1} + (K_{1,t} I_{\{B,t\}} - K_{2,t} I_{\{S,t\}}) - \sum_{j=1}^{20} (K_{1,t-j} - K_{2,t-j}) I_{t-j,t}^{close}, \\ V_t &= V_{t-1} - \{K_{1,t} (1 + C_B) I_{\{B,t\}} - K_{2,t} (1 - C_S) I_{\{S,t\}}\} S_t \\ &\quad + \sum_{j=1}^{20} \{K_{1,t-j} (1 - C_S) - K_{2,t-j} (1 + C_B)\} S_t I_{t-j,t}^{close}, \end{aligned} \quad (4)$$

(4) 式右側第二項表示新建立新頭寸的交易成本，第三項表示在第 t 天平倉的影響。

4.2 評估指標

評估各種投資方法的好壞可以由穩定性和承受風險以及收益效果判斷，本文使用 3 種評斷方式，說明如下：

1. 第 t 天的每日、每週、每月、每季、每年對數報酬，並將各個報酬年化，說明如下：

$$\text{每日年化報酬率：} Annual\ return_{day} = 250 \times \log \left(\frac{Asset_t}{Asset_{t-1}} \right)$$

$$\text{每週年化報酬率：} Annual\ return_{week} = \frac{250}{5} \sum_{i=1}^5 \log \left(\frac{Asset_{t-i+1}}{Asset_{t-i}} \right)$$

$$\text{每月年化報酬率：} Annual\ return_{month} = \frac{250}{20} \sum_{i=1}^{20} \log \left(\frac{Asset_{t-i+1}}{Asset_{t-i}} \right)$$

$$\text{每季年化報酬率：} Annual\ return_{quarter} = \frac{250}{60} \sum_{i=1}^{60} \log \left(\frac{Asset_{t-i+1}}{Asset_{t-i}} \right)$$

$$\text{每年年化報酬率：} Annual\ return_{year} = \sum_{i=1}^{250} \log \left(\frac{Asset_{t-i+1}}{Asset_{t-i}} \right)$$

其中 $Asset_t$ 表示第 t 天的總資產，數值越高表示收益越高。

2. 最大回撤 (Maximum drawdown, 簡記為 MDD; [Magdon-Ismail and Atiya 2004](#))，此指標藉由計算最大相對跌幅評斷風險：

$$\max_{i,j, i < j} \frac{A_i - A_j}{A_i},$$

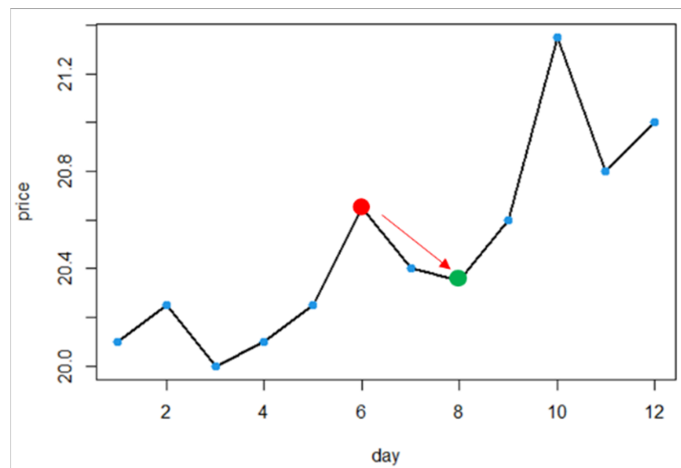


圖 4：MDD 示意圖。

其中 A_i 表示第 i 天的資產，而 j 表示使用者希望觀察未來 j 天內的 MDD，MDD 越小表示未來可能的損失量越少。圖 4 所示，假設希望求得紅點 6 天內的 MDD，則綠點是在未來這段區間中的最小值，將此 2 點的值帶入 $\frac{A_i - A_j}{A_i}$ 中，其中紅點的值為 A_i ，綠點的值為 A_j ，經計算便能求得紅點的 MDD。

3. 夏普比率 (Sharpe ratio; Sharpe 1998)，此指標在衡量風險及收益之間的關係，每承擔一單位的風險時，能夠產生多少超額報酬，數值越高越好：

$$\text{Sharpe Ratio} = \frac{\text{Annual return} - r_f}{\sigma \times \sqrt{250}},$$

其中 r_f 為無風險利率，此處使用政府 10 年期公債， σ 為每日報酬率所算出的標準差，250 主要是將波動率年化，且 250 表示一年中的平均交易日。特別地，本文採用前述所介紹的第 t 天的每日、每週、每月、每季、每年對數報酬，並分別將其年化後，計算各個不同時間長度下之報酬率所對應的夏普比率。

4.3 預測表現評估

本文所提出的方法 (Grouping and Band-Wise Investment Strategy, 簡記為 GBWI) 在步驟 (iii) 引入統計檢定方法協助判斷各群適合的投資方向，我們在實證研究中測試過許多不同的 q 設定值，結果呈現當 $q = 40\%$ 時，所提出的方法具有較穩健的投資績效表現。因此，以下將呈現 $q = 40\%$ 的實證結果，意即設定該群 (無論做多或做空) 在訓練集資料中至少需有 60% 的資料具有正報酬，我們期待若該群通過檢定，當測試集資料被分類至該群時，進場操作的建議能有超過 50% 的勝率。換言之，步驟 (iii) 的檢定步驟有助於保障 GBWI 在建議進場操作時的勝率，進而保障 GBWI 建議進場時的投資績效。

除了圖 3 所介紹的 250 個交易日的更新頻率以外，表 1 進一步列出 GBWI 在不同更新頻率並且將資金皆分為 20 份的情況下之年化報酬率，由左至右的更新頻率分別為 60、125、250 個交易日，紅色的字表示效果最佳，我們可以觀察到投資績效並未隨著越常更新模型而遞增。此外，表 2 為 GBWI 在不同更新頻率下的 MDD 比較表，我們也發現到並不是更新頻率越高，GBWI 所承受風險就越低。上述觀察可能的解釋為 GBWI 需要在訓練集中學習到足夠多元的經濟型態，才能在測試集中有較好的投資績效表現，若頻繁地更新模型，將導致前後兩期訓練集中的資料所包含的經濟型態過於相似，造成訓練效果差異不大而未能顯著提升測試資料的投資績效。在我們的實證

表 1：不同更新頻率之年化報酬比較表。

更新頻率		60	125	250	
大立光	1 天	MEAN	0.024	0.037	0.043
		SD	0.062	0.077	0.077
		SP	0.312	0.421	0.505
	5 天 (週)	MEAN	0.024	0.037	0.043
		SD	0.06	0.071	0.068
		SP	0.322	0.456	0.574
	20 天 (月)	MEAN	0.024	0.037	0.044
		SD	0.061	0.075	0.069
		SP	0.322	0.439	0.569
	60 天 (季)	MEAN	0.026	0.038	0.045
		SD	0.059	0.078	0.07
		SP	0.361	0.434	0.574
	250 天 (年)	MEAN	0.027	0.047	0.050
		SD	0.055	0.088	0.072
		SP	0.408	0.479	0.637
台積電	1 天	MEAN	-0.005	0.017	0.010
		SD	0.075	0.066	0.056
		SP	-0.125	0.186	0.100
	5 天 (週)	MEAN	-0.004	0.018	0.011
		SD	0.07	0.064	0.055
		SP	-0.114	0.212	0.123
	20 天 (月)	MEAN	-0.008	0.014	0.008
		SD	0.06	0.058	0.047
		SP	-0.217	0.168	0.074
	60 天 (季)	MEAN	-0.014	0.010	0.004
		SD	0.053	0.050	0.041
		SP	-0.349	0.112	-0.002
	250 天 (年)	MEAN	-0.009	0.018	0.013
		SD	0.041	0.048	0.040
		SP	-0.32	0.285	0.210

研究中發現不同股票在不同更新頻率的表現不一，經觀察後發現更新頻率為 250 個交易日的效果較為穩定，因此，為了簡化敘述，以下將以更新頻率為 250 個交易日進行討論。

此外，[附錄 C](#) 提供所考慮的 13 檔股票在研究期間（自 2012 年 1 月至 2022 年 3 月）的股價時間序列圖，從圖中可發現，每一檔個股每日收盤價的時間序列圖走勢都不太一樣，且大部分的個股在研究期間皆有明顯的波段急遽上升與下降的時間區間，表示在研究期間中，大部分個股皆會遭遇經濟狀態對其有利或不利的情况。因此，如何有效地捕捉急遽上升或向下的波段進行正確的進場操作，或是規避錯誤的操作，將影響最終的投資績效。

表 2：不同更新頻率之 MDD 比較表。

更新頻率			60	125	250
大立光	5 天 (週)	MEAN	0.002	0.003	0.003
		SD	0.006	0.006	0.007
	20 天 (月)	MEAN	0.006	0.007	0.007
		SD	0.011	0.011	0.012
60 天 (季)	MEAN	0.013	0.013	0.014	
	SD	0.018	0.019	0.021	
250 天 (年)	MEAN	0.024	0.032	0.032	
	SD	0.026	0.041	0.035	
台積電	5 天 (週)	MEAN	0.004	0.004	0.003
		SD	0.007	0.006	0.006
	20 天 (月)	MEAN	0.009	0.009	0.007
		SD	0.013	0.011	0.010
60 天 (季)	MEAN	0.018	0.016	0.012	
	SD	0.020	0.017	0.016	
250 天 (年)	MEAN	0.037	0.026	0.024	
	SD	0.028	0.025	0.024	

4.4 投資績效比較

為了調查 GBWI 的投資收益，本文亦與 Syu *et al.* (2021) 提出的投資策略（此處以 SYWH 代稱），以及買進持有（Buy and Hold，簡記為 B&H）策略進行投資績效比較，其中 B&H 屬於長期投資策略，會忽略短期金融市場的波動，專注於長期成長與收益率，即第一天進場，最後一天出場，且只針對多頭操作。

表 3 呈現各個投資策略的年化報酬，特別地，我們挑選大立光、台積電、與普誠等 3 家公司的股票進行討論，其餘結果則放置附錄 A。表 3 中，「年化報酬」該欄由上至下分別為每日、每週、每月、每季、與每年的平均年化報酬，表頭中 proposed_20 為本文提出的交易策略且將資金分為 20 份，proposed_10 為本文提出的交易策略且將資金分為 10 份，B&H 為買進持有，SYWH_20 為 Syu *et al.* (2021) 提出的方法且將資金分為 20 份，SYWH_10 為 Syu *et al.* (2021) 提出的方法且將資金分為 10 份，MEAN 表示平均值，SD 表示標準差，SP 為夏普指數，而紅色表示效果最佳，綠色表示效果次好，藍色表示為第 3 名。

由表 3 能夠發現 GBWI 在大立光的年化報酬率和夏普率優於 B&H 及 SYWH，主因是這間公司在研究期間的股價波動大，而 GBWI 多考慮到空頭操作，使整個投資系統更全面，因此效果優於另外 2 種（只考慮多頭）的操作方法。然而，GBWI 在台積電的效果反而沒有另外 2 種方法好，主要原因是這支股票在研究期間的整體趨勢幾

表 3：年化報酬比較表。

年化報酬		proposed_20	proposed_10	B&H	SYWH_20	SYWH_10	
大立光	1 天	MEAN	0.043	0.066	-0.017	-0.055	-0.082
		SD	0.077	0.131	0.431	0.110	0.183
		SP	0.505	0.472	-0.050	-0.541	-0.472
	5 天 (週)	MEAN	0.043	0.066	-0.019	-0.055	-0.083
		SD	0.068	0.115	0.395	0.099	0.165
		SP	0.574	0.538	-0.058	-0.602	-0.528
	20 天 (月)	MEAN	0.044	0.067	-0.014	-0.050	-0.075
		SD	0.069	0.115	0.386	0.096	0.161
		SP	0.569	0.543	-0.049	-0.570	-0.492
	60 天 (季)	MEAN	0.045	0.068	-0.031	-0.050	-0.075
		SD	0.07	0.116	0.402	0.111	0.179
		SP	0.574	0.550	-0.088	-0.494	-0.442
	250 天 (年)	MEAN	0.050	0.077	-0.007	-0.061	-0.093
		SD	0.072	0.116	0.342	0.118	0.190
		SP	0.637	0.631	-0.032	-0.553	-0.510
台積電	1 天	MEAN	0.010	0.013	0.229	0.035	0.061
		SD	0.056	0.091	0.241	0.080	0.125
		SP	0.100	0.089	0.934	0.378	0.457
	5 天 (週)	MEAN	0.011	0.014	0.232	0.034	0.061
		SD	0.055	0.091	0.231	0.081	0.124
		SP	0.123	0.103	0.984	0.371	0.454
	20 天 (月)	MEAN	0.008	0.009	0.223	0.033	0.057
		SD	0.047	0.080	0.225	0.082	0.124
		SP	0.074	0.058	0.969	0.344	0.424
	60 天 (季)	MEAN	0.004	0.004	0.217	0.031	0.054
		SD	0.041	0.072	0.226	0.083	0.122
		SP	-0.002	-0.009	0.942	0.322	0.412
	250 天 (年)	MEAN	0.013	0.020	0.248	0.044	0.072
		SD	0.040	0.065	0.195	0.072	0.107
		SP	0.210	0.234	1.252	0.553	0.626
年化報酬		GBWI_20	GBWI_10	B&H	SYWH_20	SYWH_10	
普誠	1 天	MEAN	0.043	0.064	0.156	0.021	0.237
		SD	0.052	0.084	0.485	0.157	0.035
		SP	0.747	0.712	0.313	0.108	0.130
	5 天 (週)	MEAN	0.043	0.065	0.163	0.022	0.036
		SD	0.046	0.076	0.515	0.156	0.234
		SP	0.851	0.793	0.308	0.109	0.133
	20 天 (月)	MEAN	0.043	0.063	0.164	0.024	0.041
		SD	0.052	0.082	0.548	0.155	0.226
		SP	0.735	0.716	0.291	0.129	0.161
	60 天 (季)	MEAN	0.043	0.063	0.162	0.022	0.037
		SD	0.048	0.074	0.56	0.146	0.206
		SP	0.790	0.794	0.282	0.123	0.156
	250 天 (年)	MEAN	0.046	0.067	0.223	0.029	0.043
		SD	0.051	0.074	0.613	0.104	0.166
		SP	0.816	0.848	0.357	0.233	0.234

表 4：MDD 比較表。

MDD		GBWI_20	GBWI_10	B&H	SYWH_20	SYWH_10	
大立光	5 天 (週)	MEAN	0.003	0.006	0.032	0.006	0.011
		SD	0.007	0.013	0.037	0.012	0.018
	20 天 (月)	MEAN	0.007	0.014	0.071	0.016	0.028
		SD	0.012	0.024	0.067	0.023	0.035
60 天 (季)	MEAN	0.014	0.029	0.133	0.035	0.058	
	SD	0.021	0.043	0.114	0.042	0.061	
250 天 (年)	MEAN	0.032	0.066	0.263	0.100	0.158	
	SD	0.035	0.069	0.161	0.066	0.093	
台積電	5 天 (週)	MEAN	0.003	0.005	0.015	0.004	0.007
		SD	0.006	0.009	0.018	0.008	0.012
	20 天 (月)	MEAN	0.007	0.012	0.033	0.011	0.017
		SD	0.01	0.018	0.033	0.017	0.025
60 天 (季)	MEAN	0.012	0.022	0.054	0.023	0.035	
	SD	0.016	0.028	0.051	0.029	0.041	
250 天 (年)	MEAN	0.024	0.042	0.068	0.042	0.066	
	SD	0.024	0.040	0.071	0.041	0.054	
普誠	5 天 (週)	MEAN	0.001	0.003	0.027	0.009	0.014
		SD	0.003	0.006	0.034	0.014	0.02
	20 天 (月)	MEAN	0.004	0.007	0.066	0.023	0.035
		SD	0.06	0.011	0.062	0.027	0.039
60 天 (季)	MEAN	0.007	0.014	0.130	0.042	0.065	
	SD	0.01	0.018	0.107	0.044	0.06	
250 天 (年)	MEAN	0.015	0.028	0.201	0.074	0.107	
	SD	0.016	0.031	0.152	0.060	0.081	

乎都在上升，由於 GBWI 屬於較保守的投資方法，因此較容易錯失部分進場時機，導致減少部分收益。而從各個方法的 SD 也能發現到 GBWI 的變動幾乎為最低，波動不會過大，如此也能發現此方法比較不會因為突發事件而損失慘重。最後，雖然 GBWI 在普誠的年化報酬低於 B&H 但卻擁有較高的夏普率，這顯示出 B&H 的波動率較高，相對較不穩定。

此外，表 4 呈現上述 3 家公司股票的 MDD 值，其餘結果請參見附錄 B。表 4 中的「MDD」該欄由上至下分別為每週、每月、每季、與每年的平均承擔風險，紅色表示效果最佳，綠色表示效果次好，藍色表示為第 3 名。由表 4 可以發現 GBWI 所承受的風險都最低。

綜合表 3、表 4、與附錄 A、附錄 B 的結果，我們發現在本實證分析的 13 間公司中，GBWI 在其中 10 支股票的報酬大多都能名列前 3，且其 SD 較為穩定，同時所承受的 MDD 風險幾乎是最低。回顧附錄 C 的 13 檔股票在研究期間（自 2012 年 1 月

至 2022 年 3 月) 的股價時間序列圖, 從報酬率比較表的結果我們推估 GBWI 在有效地捕捉急遽上升或向下的波段進行正確的進場操作方面較為保守, 導致錯過一些進場操作的機會, 但從 MDD 的比較則可得知其在規避錯誤的操作上有優異的表現。因此, GBWI 可視為一個較為穩健的投資策略, 適合風險趨避程度較強的投資人。

5. 結論與討論

本文提出一 GBWI 的投資策略, 該策略同時考慮了許多財經指標, 應用非監督式學習 SIMLR 法對交易日進行分群以協助判斷每一交易日的經濟狀態, 再根據不同經濟狀態分別設立多頭操作與空頭操作的最佳停損點與最佳停利點, 接著透過 Wilcoxon 檢定法判斷各群適合的操作策略, 此步驟亦能夠區分未來趨勢不明顯的族群, 最終利用機器學習的方式進行測試集的分類, 進而決定測試集資料的操作方向。

本文除了考慮多空雙向操作以增加投資多樣性之外, 亦同時採用設立停損點及停利點的策略, 以防止趨勢判斷錯誤或在投資期間發生其他不可測的事件, 雖然可能在持有(賣出)股票的未來 20 天中會因先碰到停損點而提早出場, 錯過未來股價回升(下降)的機會, 但此設定能有效防止損失過多, 使整體投資績效更加穩健。在本文所考慮的 13 檔台灣上市公司股票自 2012 年 1 月至 2022 年 3 月間所進行的實證分析結果顯示, 相較於 B&H 與 SYWH 方法, 所提出的 GBWI 策略所承擔的風險幾乎都是最低, 即便少數公司投資收益為負, 但也因停損和資金的合理分配使得損失得到控制。整體而言所提出的方法的投資獲利較穩定, 適合風險趨避程度較高的投資人採用。未來將朝向其他股票市場或非股票金融商品進行更完整的實證分析, 以測試所提出方法的適用性與穩健性。

然而, 實證結果同時顯示 GBWI 仍無法讓投資本文所考慮的每支股票時, 在各項評估指標都達到最佳。因此, 整個框架中仍有許多地方可以嘗試調整, 例如: 目前本文採用近 5 日收盤價之標準差作為近 5 日波動的代表性統計量, 未來可進一步嘗試將此部分改用其他量測個股或市場波動的統計量, 或有機會能提升投資組合依市場波動自動調整投資權重的效果。此外, 目前的原指標已進行差分處理, 可以將原資料(未經差分)也加入訓練集中, 如此一來便能將一筆資料的特徵增加至 240 個, 獲得更多特徵資訊。此外, 除了透過實證研究探討所提出方法的投資績效以外, 透過模擬股價亦能在較為可控的環境下, 探討所提出之方法在各個環節對於投資績效的影響, 對於未來如何優化與精進所提出之投資策略, 有助於提供更具體的線索與明確的改善方向。

另一個可以調整的部分是關於 SIMLR 分群, 此部分並非只有 SIMLR 能夠嘗試,

目前採用 SIMLR 是因為其具有不錯的分群效果，但每一群卻不一定能夠完全適合單一投資操作（做多或做空），例如：若在 A 交易日進場後，未來會先碰到停利點，但在 B 交易日進場後，未來卻會先碰到停損點，但 SIMLR 卻可能將其分到同一群，使得該群未來趨勢不夠明顯，會判斷不進行操作，若每一群皆產生此現象便會造成測試集資料幾乎都判斷不進場操作，或許會錯失一些進場機會。因此，持續尋找其他非監督式學習的方法，使其不僅能依據資料間的相關性進行分群，並且各群能夠完全適合單一投資操作亦為有趣的研究方向。

在步驟 (iv) 對測試資料的歸類方面，本研究採用三種機器學習的分類器，分別計算每一測試資料的分類結果，再透過加權和的方式得到最後的分類結果，其中的權重則以每一個分類器在訓練集資料中的準確率 (accuracy) 之比例計算。在這一類透過加權和以整合不同模型預測結果的作法方面，文獻上亦有許多不同的權重計算方法 (Kuncheva 2014)，特別地，Buckland *et al.* (1997) 與 Claeskens and Hjort (2003) 提出以模型選擇 (model selection) 領域中的信息準則 (information criterion) 作為計算不同模型在預測時的權重，進一步探討如何將信息準則的概念延伸到機器學習領域中的分類器 (如本文所採用的三種無母數分類器)，以作為決定權重的依據，亦是一個重要且有趣的研究方向。

附錄 A. 年化報酬率及夏普率表格

由上至下分別為瑞昱、南亞科、聯發科、廣達、研華、中鋼、國巨、國泰金、永豐金、群創年化報酬率及夏普率表格。

年化報酬		proposed_20	proposed_10	B&H	SYWH_20	SYWH_10	
瑞昱	1 天	MEAN	0.068	0.115	0.242	0.040	0.071
		SD	0.094	0.153	0.366	0.104	0.161
		SP	0.668	0.723	0.649	0.337	0.410
	5 天 (週)	MEAN	0.067	0.115	0.245	0.040	0.071
		SD	0.092	0.149	0.356	0.103	0.159
		SP	0.687	0.742	0.676	0.345	0.420
	20 天 (月)	MEAN	0.071	0.120	0.253	0.042	0.075
		SD	0.094	0.154	0.353	0.101	0.160
		SP	0.711	0.755	0.702	0.371	0.442
	60 天 (季)	MEAN	0.070	0.120	0.255	0.042	0.076
		SD	0.097	0.16	0.351	0.103	0.163
		SP	0.676	0.720	0.714	0.366	0.438
	250 天 (年)	MEAN	0.083	0.142	0.319	0.055	0.091
		SD	0.072	0.107	0.258	0.100	0.173
		SP	1.093	1.280	1.218	0.503	0.500

年化報酬		proposed_20	proposed_10	B&H	SYWH_20	SYWH_10	
南亞科	1 天	MEAN	0.009	0.015	-0.008	-0.034	-0.035
		SD	0.052	0.098	0.413	0.113	0.199
		SP	0.093	0.104	-0.029	-0.338	-0.197
	5 天 (週)	MEAN	0.009	0.015	-0.002	-0.034	-0.035
		SD	0.047	0.09	0.425	0.104	0.185
		SP	0.106	0.118	-0.015	-0.371	-0.213
	20 天 (月)	MEAN	0.010	0.016	-0.004	-0.035	-0.036
		SD	0.047	0.090	0.437	0.105	0.191
		SP	0.115	0.125	-0.019	-0.375	-0.211
	60 天 (季)	MEAN	0.010	0.017	-0.006	-0.037	-0.038
		SD	0.046	0.089	0.436	0.103	0.178
		SP	0.127	0.137	-0.025	-0.396	-0.240
	250 天 (年)	MEAN	0.012	0.020	0.068	-0.019	-0.017
		SD	0.038	0.075	0.370	0.079	0.135
		SP	0.196	0.203	0.173	-0.291	-0.162
聯發科	1 天	MEAN	0.008	0.020	0.116	0.047	0.056
		SD	0.046	0.087	0.368	0.103	0.172
		SP	0.078	0.181	0.302	0.416	0.297
	5 天 (週)	MEAN	0.008	0.021	0.119	0.048	0.056
		SD	0.047	0.088	0.362	0.1	0.163
		SP	0.082	0.186	0.317	0.435	0.318
	20 天 (月)	MEAN	0.009	0.022	0.122	0.049	0.058
		SD	0.042	0.078	0.356	0.100	0.155
		SP	0.110	0.221	0.330	0.446	0.346
	60 天 (季)	MEAN	0.006	0.016	0.127	0.049	0.059
		SD	0.042	0.079	0.379	0.108	0.163
		SP	0.029	0.150	0.322	0.415	0.331
	250 天 (年)	MEAN	0.004	0.011	0.180	0.050	0.059
		SD	0.022	0.05	0.407	0.114	0.168
		SP	-0.041	0.132	0.432	0.397	0.327
廣達	1 天	MEAN	0.003	0.005	0.025	-0.023	-0.031
		SD	0.015	0.028	0.243	0.069	0.112
		SP	-0.092	0.009	0.086	-0.391	-0.317
	5 天 (週)	MEAN	0.003	0.005	0.027	-0.023	-0.031
		SD	0.013	0.024	0.236	0.063	0.107
		SP	-0.107	0.010	0.097	-0.427	-0.333
	20 天 (月)	MEAN	0.003	0.005	0.025	-0.023	-0.031
		SD	0.010	0.017	0.230	0.065	0.108
		SP	-0.140	0.016	0.088	-0.422	-0.332
	60 天 (季)	MEAN	0.003	0.005	0.020	-0.023	-0.032
		SD	0.009	0.017	0.221	0.054	0.094
		SP	-0.126	0.037	0.069	-0.514	-0.391
	250 天 (年)	MEAN	0.003	0.005	0.047	-0.025	-0.034
		SD	0.010	0.018	0.206	0.040	0.070
		SP	-0.095	0.043	0.208	-0.742	-0.545

年化報酬		proposed_20	proposed_10	B&H	SYWH_20	SYWH_10	
研華	1 天	MEAN	-0.009	0.013	0.072	0.001	0.013
		SD	0.102	0.129	0.274	0.078	0.128
		SP	-0.131	0.070	0.248	-0.05	0.064
	5 天 (週)	MEAN	-0.01	0.013	0.074	0.000	0.012
		SD	0.095	0.116	0.241	0.071	0.115
		SP	-0.147	0.070	0.289	-0.061	0.065
	20 天 (月)	MEAN	-0.01	0.012	0.073	0.000	0.012
		SD	0.09	0.109	0.221	0.073	0.115
		SP	-0.157	0.072	0.312	-0.057	0.069
	60 天 (季)	MEAN	-0.015	0.006	0.067	-0.001	0.011
		SD	0.077	0.088	0.194	0.066	0.106
		SP	-0.255	0.015	0.324	-0.076	0.061
	250 天 (年)	MEAN	-0.009	0.006	0.079	0.001	0.014
		SD	0.052	0.051	0.164	0.038	0.068
		SP	-0.259	0.032	0.453	-0.099	0.138
中鋼	1 天	MEAN	0.002	0.005	0.036	0.017	0.026
		SD	0.022	0.043	0.226	0.044	0.079
		SP	-0.095	0.020	0.138	0.290	0.272
	5 天 (週)	MEAN	0.003	0.007	0.039	0.017	0.026
		SD	0.021	0.041	0.237	0.044	0.079
		SP	-0.055	0.061	0.144	0.292	0.272
	20 天 (月)	MEAN	0.004	0.008	0.041	0.018	0.026
		SD	0.019	0.037	0.248	0.043	0.074
		SP	-0.016	0.106	0.145	0.306	0.295
	60 天 (季)	MEAN	0.004	0.008	0.041	0.016	0.024
		SD	0.016	0.031	0.243	0.039	0.067
		SP	-0.036	0.110	0.151	0.299	0.292
	250 天 (年)	MEAN	0.004	0.007	0.069	0.018	0.028
		SD	0.013	0.027	0.232	0.026	0.046
		SP	-0.053	0.111	0.278	0.544	0.501
國巨	1 天	MEAN	0.079	0.132	0.330	0.029	0.054
		SD	0.148	0.224	0.516	0.109	0.188
		SP	0.502	0.569	0.632	0.227	0.266
	5 天 (週)	MEAN	0.079	0.132	0.333	0.03	0.056
		SD	0.15	0.229	0.532	0.108	0.182
		SP	0.495	0.557	0.618	0.236	0.281
	20 天 (月)	MEAN	0.073	0.121	0.325	0.029	0.054
		SD	0.162	0.251	0.567	0.109	0.181
		SP	0.423	0.464	0.565	0.227	0.274
	60 天 (季)	MEAN	0.067	0.110	0.309	0.024	0.045
		SD	0.196	0.302	0.673	0.129	0.209
		SP	0.319	0.351	0.452	0.152	0.194
	250 天 (年)	MEAN	0.102	0.160	0.381	0.046	0.076
		SD	0.161	0.266	0.772	0.153	0.244
		SP	0.602	0.586	0.488	0.272	0.294

年化報酬		proposed_20	proposed_10	B&H	SYWH_20	SYWH_10	
國泰金	1 天	MEAN	0.010	0.019	0.049	0.026	0.036
		SD	0.015	0.030	0.218	0.047	0.078
		SP	0.359	0.492	0.202	0.471	0.402
	5 天 (週)	MEAN	0.010	0.019	0.050	0.027	0.036
		SD	0.016	0.031	0.215	0.047	0.077
		SP	0.350	0.480	0.212	0.474	0.411
	20 天 (月)	MEAN	0.010	0.020	0.044	0.024	0.031
		SD	0.017	0.033	0.203	0.041	0.066
		SP	0.334	0.457	0.196	0.481	0.409
	60 天 (季)	MEAN	0.010	0.020	0.042	0.024	0.031
		SD	0.016	0.032	0.199	0.04	0.065
		SP	0.355	0.483	0.186	0.492	0.405
	250 天 (年)	MEAN	0.012	0.022	0.031	0.024	0.030
		SD	0.017	0.033	0.23	0.045	0.070
		SP	0.416	0.539	0.114	0.425	0.363
永豐金	1 天	MEAN	0.001	0.002	0.050	-0.014	-0.030
		SD	0.009	0.019	0.193	0.042	0.079
		SP	-0.347	-0.111	0.238	-0.427	-0.438
	5 天 (週)	MEAN	0.000	0.001	0.048	-0.014	-0.03
		SD	0.009	0.018	0.195	0.04	0.075
		SP	-0.452	-0.204	0.223	-0.453	-0.464
	20 天 (月)	MEAN	0.000	0.001	0.042	-0.015	-0.033
		SD	0.008	0.016	0.193	0.038	0.074
		SP	-0.522	-0.242	0.195	-0.504	-0.498
	60 天 (季)	MEAN	0.000	0.000	0.041	-0.014	-0.033
		SD	0.009	0.017	0.189	0.039	0.075
		SP	-0.490	-0.229	0.193	-0.491	-0.498
	250 天 (年)	MEAN	0.001	0.002	0.035	-0.014	-0.033
		SD	0.007	0.015	0.184	0.044	0.082
		SP	-0.452	-0.150	0.167	-0.429	-0.453
群創	1 天	MEAN	0.027	0.037	0.002	0.007	-0.024
		SD	0.074	0.094	0.392	0.136	0.208
		SP	0.306	0.352	-0.007	0.017	-0.137
	5 天 (週)	MEAN	0.027	0.037	0.003	0.007	-0.024
		SD	0.076	0.096	0.399	0.134	0.202
		SP	0.298	0.346	-0.003	0.018	-0.140
	20 天 (月)	MEAN	0.027	0.037	0.010	0.007	-0.024
		SD	0.083	0.101	0.433	0.134	0.205
		SP	0.269	0.321	0.014	0.020	-0.138
	60 天 (季)	MEAN	0.028	0.039	0.017	0.009	-0.021
		SD	0.075	0.085	0.428	0.14	0.213
		SP	0.314	0.403	0.028	0.033	-0.119
	250 天 (年)	MEAN	0.027	0.036	0.065	0.021	-0.001
		SD	0.054	0.062	0.451	0.148	0.221
		SP	0.427	0.504	0.134	0.115	-0.023

附錄 B. MDD 比較表格

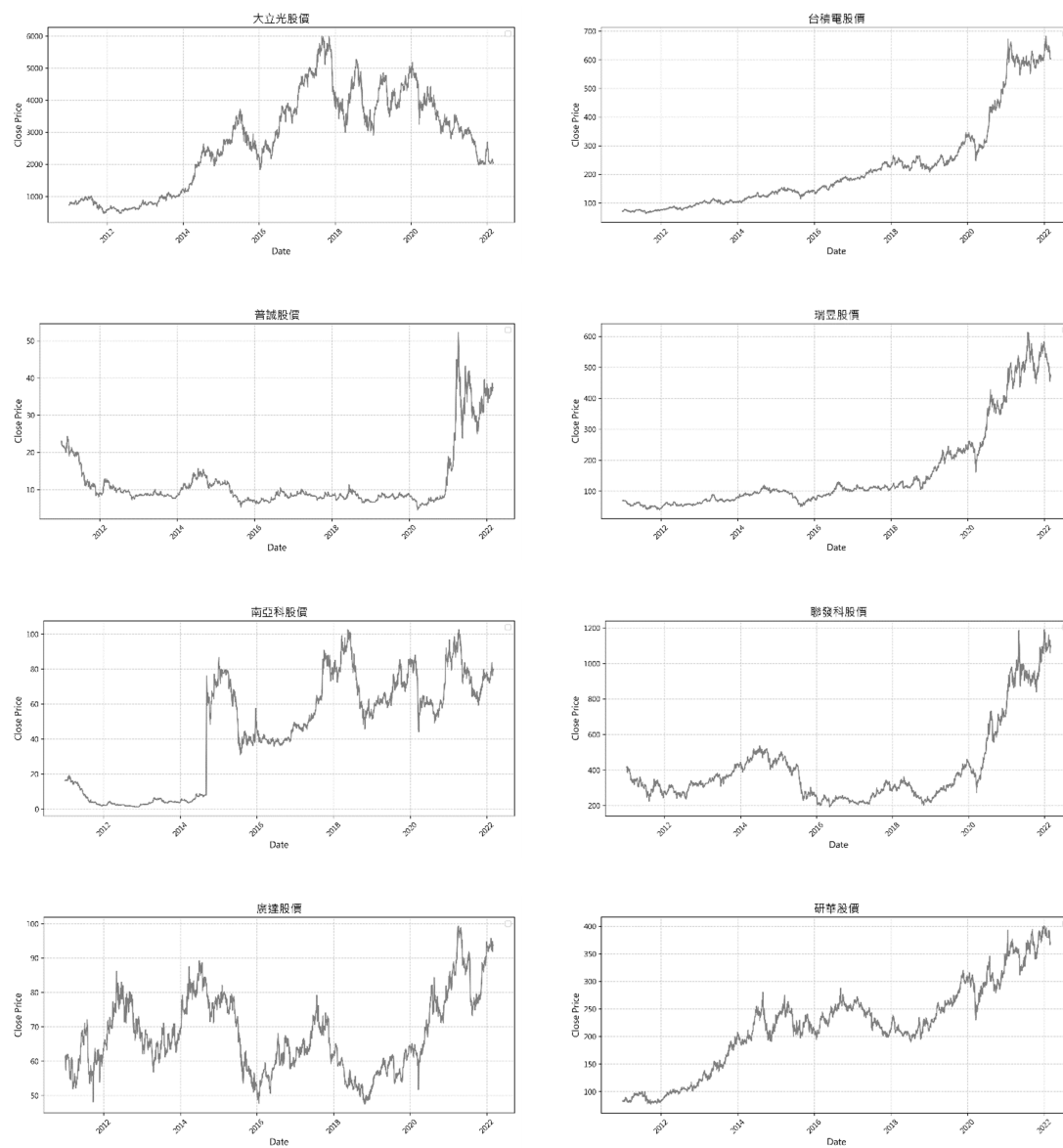
由上至下分別為瑞昱、南亞科、聯發科、廣達、研華、中鋼、國巨、國泰金、永豐金、群創 MDD 比較表格。

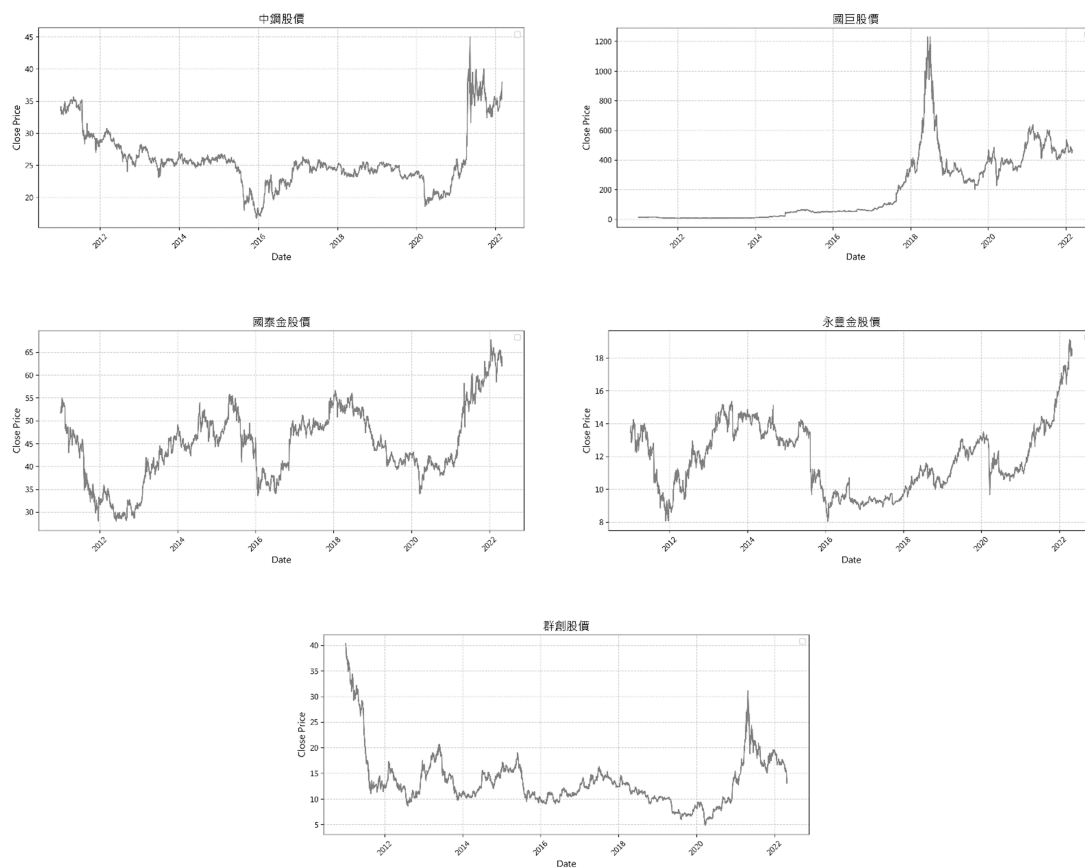
MDD		GBWI_20	GBWI_10	B&H	SYWH_20	SYWH_10	
瑞昱	5 天 (週)	MEAN SD	0.005 0.009	0.007 0.014	0.023 0.03	0.006 0.01	0.009 0.015
	20 天 (月)	MEAN SD	0.010 0.016	0.017 0.025	0.053 0.058	0.014 0.021	0.021 0.031
	60 天 (季)	MEAN SD	0.019 0.028	0.030 0.046	0.099 0.099	0.029 0.035	0.044 0.050
	250 天 (年)	MEAN SD	0.025 0.038	0.042 0.065	0.135 0.154	0.049 0.037	0.079 0.052
南亞科	5 天 (週)	MEAN SD	0.002 0.005	0.003 0.01	0.03 0.037	0.007 0.011	0.012 0.019
	20 天 (月)	MEAN SD	0.004 0.009	0.009 0.018	0.071 0.077	0.018 0.023	0.031 0.038
	60 天 (季)	MEAN SD	0.009 0.015	0.018 0.028	0.132 0.14	0.038 0.04	0.063 0.062
	250 天 (年)	MEAN SD	0.026 0.025	0.051 0.048	0.174 0.201	0.073 0.047	0.122 0.074
聯發科	5 天 (週)	MEAN SD	0.002 0.005	0.003 0.008	0.027 0.032	0.005 0.009	0.009 0.016
	20 天 (月)	MEAN SD	0.004 0.008	0.008 0.014	0.066 0.063	0.012 0.018	0.022 0.031
	60 天 (季)	MEAN SD	0.008 0.012	0.016 0.022	0.131 0.106	0.024 0.032	0.042 0.051
	250 天 (年)	MEAN SD	0.013 0.012	0.026 0.024	0.223 0.176	0.044 0.05	0.08 0.085
廣達	5 天 (週)	MEAN SD	0.000 0.001	0.001 0.003	0.019 0.021	0.003 0.007	0.005 0.012
	20 天 (月)	MEAN SD	0.001 0.002	0.002 0.005	0.046 0.04	0.008 0.017	0.014 0.026
	60 天 (季)	MEAN SD	0.002 0.004	0.004 0.007	0.088 0.064	0.019 0.026	0.030 0.039
	250 天 (年)	MEAN SD	0.005 0.005	0.009 0.009	0.180 0.126	0.051 0.036	0.074 0.052
研華	5 天 (週)	MEAN SD	0.004 0.011	0.005 0.013	0.019 0.023	0.004 0.008	0.007 0.012
	20 天 (月)	MEAN SD	0.010 0.021	0.012 0.024	0.042 0.038	0.010 0.017	0.017 0.025
	60 天 (季)	MEAN SD	0.019 0.037	0.022 0.039	0.074 0.053	0.021 0.027	0.034 0.039
	250 天 (年)	MEAN SD	0.042 0.068	0.042 0.061	0.143 0.071	0.039 0.029	0.062 0.044

MDD		GBWI_20	GBWI_10	B&H	SYWH_20	SYWH_10	
中鋼	5 天 (週)	MEAN SD	0.000 0.002	0.001 0.004	0.013 0.016	0.002 0.004	0.003 0.008
	20 天 (月)	MEAN SD	0.001 0.004	0.002 0.007	0.032 0.037	0.004 0.008	0.007 0.015
	60 天 (季)	MEAN SD	0.002 0.005	0.004 0.009	0.061 0.064	0.007 0.012	0.013 0.023
	250 天 (年)	MEAN SD	0.003 0.003	0.006 0.007	0.117 0.114	0.012 0.016	0.022 0.032
國巨	5 天 (週)	MEAN SD	0.007 0.013	0.011 0.02	0.031 0.043	0.006 0.01	0.01 0.017
	20 天 (月)	MEAN SD	0.018 0.025	0.027 0.038	0.064 0.082	0.015 0.021	0.025 0.033
	60 天 (季)	MEAN SD	0.036 0.051	0.055 0.074	0.108 0.154	0.032 0.04	0.051 0.062
	250 天 (年)	MEAN SD	0.061 0.075	0.095 0.104	0.074 0.093	0.064 0.056	0.101 0.085
國泰金	5 天 (週)	MEAN SD	0.000 0.001	0.001 0.003	0.016 0.019	0.002 0.004	0.004 0.007
	20 天 (月)	MEAN SD	0.001 0.002	0.001 0.005	0.038 0.041	0.005 0.008	0.009 0.014
	60 天 (季)	MEAN SD	0.001 0.003	0.002 0.006	0.070 0.071	0.009 0.011	0.017 0.019
	250 天 (年)	MEAN SD	0.002 0.004	0.004 0.007	0.135 0.126	0.015 0.015	0.029 0.029
永豐金	5 天 (週)	MEAN SD	0.000 0.001	0.000 0.002	0.013 0.02	0.002 0.005	0.004 0.009
	20 天 (月)	MEAN SD	0.000 0.002	0.001 0.004	0.033 0.043	0.006 0.01	0.011 0.02
	60 天 (季)	MEAN SD	0.001 0.003	0.002 0.006	0.066 0.076	0.014 0.018	0.026 0.033
	250 天 (年)	MEAN SD	0.002 0.005	0.005 0.01	0.133 0.141	0.037 0.034	0.069 0.059
群創	5 天 (週)	MEAN SD	0.002 0.006	0.003 0.008	0.024 0.027	0.007 0.012	0.012 0.019
	20 天 (月)	MEAN SD	0.005 0.013	0.007 0.017	0.057 0.054	0.018 0.024	0.030 0.039
	60 天 (季)	MEAN SD	0.010 0.023	0.013 0.029	0.104 0.095	0.037 0.041	0.059 0.067
	250 天 (年)	MEAN SD	0.025 0.038	0.034 0.045	0.195 0.147	0.074 0.062	0.125 0.105

附錄 C. 股票價格時間序列圖

2012 年 1 月至 2022 年 3 月，本文所討論的 13 檔台灣上市公司股票價格時間序列圖。





參考文獻

- [1] Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data. URL: <https://doi.org/10.1145/170036.170072>
- [2] Breiman, L. (2001). Random forests. *Machine Learning*, 45, pages 5-32.
- [3] Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics*, 53(2), pages 603-618.
- [4] Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. URL: <https://doi.org/10.1145/2939672.2939785>

- [5] Claeskens, G., and Hjort, N. L. (2003). The focused information criterion. *Journal of the American Statistical Association*, 98(464), pages 900-916.
- [6] Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., and Ahsan, M. J. (2022). Machine learning in drug discovery: a review. *Artificial Intelligence Review*, 55, pages 1947-1999.
- [7] Huang, S. F., Guo, M., and Chen, M. R. (2020). Stock market trend prediction using a functional time series approach. *Quantitative Finance*, 20(1), pages 69-79.
- [8] Huang, S. F., and Lin, T. Y. (2018). A linearization of the portfolio optimization problem with general risk measures under multivariate conditional heteroskedastic models. *Asia-Pacific Journal of Financial Studies*, 47(3), pages 449-469.
- [9] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Proceedings of the 31st International Conference on Neural Information Processing Systems. URL: <https://dl.acm.org/doi/pdf/10.5555/3294996.3295074>
- [10] Kuncheva, L. I. (2014). *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons.
- [11] Lai, W. T., Chen, R. B., and Huang, S. F. (2025). A modified VAR-deGARCH model for asynchronous multivariate financial time series via variational Bayesian inference. *International Journal of Forecasting*, 41(1), pages 345-360.
- [12] Magdon-Ismail, M., and Atiya, A. F. (2004). Maximum drawdown *Risk Magazine*, 17(10), pages 99-102.
- [13] Ramazzotti, D., Lal, A., Wang, B., Batzoglou, S., and Sidow, A. (2018). Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nature Communications*, 9, 4453.
- [14] Scavuzzo, M. A., Hill, M. C., Chmielowiec, J., Yang, D., Teaw, J., Sheng, K., Kong, Y., Bettini, M., Zong, C., Martin, J. F., and Borowiak, M. (2018). Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. *Nature Communications*, 9, 3356.

- [15] Sharpe, W. F. (1998). The sharpe ratio. *Streetwise-the Best of the Journal of Portfolio Management*, 3, pages 169-185.
- [16] Syu, J. H., Yeh, Y. R., Wu, M. E., and Ho, J. M. (2021). Self-management portfolio system with adaptive association mining: a practical application on Taiwan stock market. *Mathematics*, 9(10), 1093.
- [17] Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons.
- [18] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., and Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6), pages 463-477.
- [19] Wang, B., Ramazzotti, D., De Sano, L., Zhu, J., Pierson, E., and Batzoglou, S. (2018). SIMLR: A tool for large-scale genomic analyses by multi-kernel learning. *Proteomics*, 18(2).
- [20] Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., and Batzoglou, S. (2017). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nature Methods*, 14, pages 414-416.
- [21] Wilcoxon, F. (1992). Individual comparisons by ranking methods. In Kotz, S., Johnson, N.L. (Ed.) *Breakthroughs in Statistics: Methodology and Distribution* (pages 196-202). Springer.
- [22] Wu, M. E., and Chung, W. H. (2019). Empirical evaluations on momentum effects of Taiwan index futures via stop-loss and stop-profit mechanisms. *International Journal of Information Technology & Decision Making*, 18(2), pages 629-648.
- [23] Yuan, Z., Zhao, F., Lin, S., Zhao, Y., Yao, J., Cui, Y., Zhang, X. Y., Zhao, Y. (2024). Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nature Methods*, 21(4), pages 712-722.