

輔仁大學金融與國際企業學系

新聞性，你敢信？運用新聞不確定性與新聞情緒指標預測臺灣加權指數漲跌

指導老師：蔡麗茹博士

組員姓名：康韡瀚

卓群祐

陳宇軒

陳爾俊

黃冠儒

郭東輝

中華民國一百一十年五月

中文摘要

為了建構出能在股市中獲利的投資策略，投資人往往需要對股市價格有一定程度上的預測能力。以往，投資人通常採取使用以股票歷史價格資料所編制而成的技術指標或總體經濟指標預測及評價股價。然而，在網路發達的現代，財經新聞在投資人的行為中常扮演很重要的角色，為了量化臺灣大眾媒體對投資人的影響，本文除了採用傳統預測及研究股價時常用到的總體經濟指標如匯率、利率、國際指數等常見的十六個總經變數以外，也參考 Baker et al(2016)和黃裕烈(2020)等人編列經濟政策不確定性指標的方法自行編列新聞性指標以及新聞情緒指標以及 Google 搜尋趨勢對臺灣加權指數建立模型進行預測，以期能夠得到更好的預測精準度。在研究方法上，為了同時考慮線性及非線性關係，本文統計上採取傳統線性迴歸模型以及非線性模型如門檻迴歸模型以盡可能完備的掌握在不同的市場情況下財經新聞對股價的影響。更重要的是，我們也採用了三種機器學習的方法來掌握總體經濟指標或財經新聞性等因素對股價的漲跌影響可能有複雜，且非可以直接觀察到的關係。

統計結果發現，自選變數在不同結構點中，恐慌指數具有加乘或抵消之卓著貢獻，值得一提的是，政策性指標於中長期對於股市影響有限，在不同結構點中大多為不顯著之結果。機器學習發現，參考傳統預測臺灣加權指數論文使用變數的基本模型的 F1-score 皆低於新增新聞性的變數的模型，其中又以新聞情緒指標模型的 F1-score 最高，代表加入新聞性變數後模型對於臺灣加權指數的預測性顯著提升。

關鍵字：EPU 指標、情緒指標、機器學習、門檻、臺灣加權指數預測

ABSTRACT

In order to construct an investment strategy that can make a profit in the stock market, investors often need to have a certain degree of predictive ability about stock market prices. In the past, investors usually used technical indicators or microeconomic indicators compiled from historical stock prices to predict and evaluate stock prices. However, in the modern era of the Internet, financial news often plays an important role in investor behavior. In order to quantify the impact of Taiwan's mass media on investors, this research uses traditional forecasts and research stock prices commonly used in the overall economic indicators, such as exchange rates, interest rates, international indices in 16 common macroeconomic variables. So as to obtain better predicted accuracy, we also refer to Baker et al and Huang Yulie's method of compiling economic policy uncertainty indicator to self-compiled news-based indicator, news sentiment indicator and Google search trend to establish a model to predict Taiwan weighed index. In terms of research methods, in order to consider both linear and non-linear relationships, this research uses traditional linear regression models and non-linear models such as threshold regression models to fully grasp the impact of financial news on stock prices under different market conditions. More importantly, we have also adopted three machine learning methods to grasp the influence of factors such as overall economic indicators or financial news on the rise and fall of stock prices, which may have complex and non-observable relationships.

Statistical results found that the self-selected variables in different structural points, the panic index has an outstanding contribution of multifying or offsetting. It is worth mentioning that the policy indicator has limited impact on the stock market in the medium and long term, and mostly has insignificant results in different structural points. Machine learning found that F1-score of the basic model refers to the using variables in traditional forecasting Taiwan weighed index papers is lower than the model of newly-added news variables. Among them, news sentiment index model has the highest F1-score, which means that after adding news variables, the model's predictability of Taiwan weighed index improve significantly.

Key words: EPU indicator, Sentiment indicator, Machine learning, Threshold, Taiwan weighed index forecast

誌謝詞

時光飛逝，轉眼間，一年的專題研究已進入尾聲。

在大二升大三時，我們對專題懵懂無知，也不知道想要研究的方向，直到與老師互相討論與溝通，才慢慢確定了方向，也才著手開始閱讀文獻、蒐集資料、設計適當的研究方法，到最後不斷的改善實證結果，在這一年中，組員難免有些許的爭吵，不過都是希望能完成一篇專業與有說服力的論文，組員們都發揮自己的專長，將各自擅長的領域應用到論文中，使得這個團隊往更好的目標邁進。彼此都十分珍惜有個能夠一起完成論文的機會，也感謝這段時光，有彼此的互相陪伴及勉勵。

本文幸蒙 蔡麗茹教授這一年的用心指導，不僅耐心的回答我們遇到的種種問題，也幫我們排解許多的研究障礙，在老師教導論文的過程中，我們從老師身上學到嚴謹的態度。儘管老師的學校事務已忙得不可開交，卻還是盡心盡力的教導，全心全意的投入我們的論文，並犧牲許多的下班時間來指導學生論文的的方向及細節，常常到了凌晨還回答我們的問題，而且老師給我們的回覆都能順利解決問題。在老師的諄諄教悔下，我們才能順利完成此篇論文，也才對政策不確定性有更深入的了解及體悟。

在每次的討論中，本組都能從老師身上學到許多知識，不僅是論文上的知識，也培養學生往後踏入社會的做事態度，師恩浩翰，無以回報，在此謹向恩師致上最高的敬意！

最後要感謝專題論文中曾經幫助我們的同學們，給予了本組許多建議，也幫助我們在撰寫論文的過程中能更加順利。雖然要兼顧課業及專題有些許的辛苦，付出了許多時間及辛勞，不過這些將成為大學生活中最具意義且寶貴的回憶，未來雖然存在著許多的不確定性，但這個經歷一定能帶領我們向未來更邁進一步。

康韡瀚、卓群祐、陳宇軒、陳爾俊、黃冠儒、郭東輝 謹誌
輔仁大學金融與國際企業學系
中華民國一一零年五月

目錄

中文摘要.....	i
ABSTRACT.....	ii
誌謝詞.....	iii
目錄.....	iv
表目錄.....	v
圖目錄.....	vi
第一章 緒論.....	1
第一節 研究動機與目的.....	1
第二節 研究流程.....	3
第二章 文獻探討.....	4
第一節 總體經濟指標對股價關聯之相關文獻.....	4
第二節 新聞性指標建構及應用之相關文獻.....	6
第三章 研究方法.....	10
第一節 爬蟲變數處理方法.....	10
第二節 統計方法.....	14
第三節 機器學習方法.....	16
第四章 實證結果分析.....	21
第一節 樣本說明.....	21
第二節 變數基本分析(傳統多變量迴歸).....	23
第三節 多元迴歸結果.....	28
第四節 機器學習結果.....	39
第五章 結論與建議.....	51
第一節 結論.....	51
第二節 後續研究建議.....	51
參考文獻.....	53
附錄.....	55

表目錄

表 3-1 新聞內文斷詞範例	12
表 3-2 NTUSD 正負向詞舉例	13
表 3-3 混淆矩陣範例表	18
表 4-1 使用變數分類說明表	22
表 4-2 基本敘述統計量表	27
表 4-3 基本變數篩檢過程表	28
表 4-4 交乘項迴歸結果	32
表 4-5 門檻迴歸結果(新聞性經濟指標)	34
表 4-6 門檻迴歸結果(新聞性政策指標)	36
表 4-7 門檻迴歸結果(GOOGLE 搜尋趨勢)	37
表 4-8 模型使用變數表	39
表 4-9 模型使用變數表	40
表 4-10 經濟變數模型績效表	41
表 4-11 經濟變數模型特徵比重表	42
表 4-12 新聞情緒模型績效表	43
表 4-13 新聞情緒模型特徵比重表	44
表 4-14 GOOGLE 搜尋趨勢模型績效表	45
表 4-15 GOOGLE 搜尋趨勢模型特徵比重表	46
表 4-16 新聞性模型表	47
表 4-17 新聞性模型特徵比重表	48
表 4-18 各模型績效表	49

圖目錄

圖 1-1 研究流程圖	3
圖 2-1 恐慌指數和 EPU 指標.....	7
圖 3-1 工商時報新聞範例（擷取自工商時報中時新聞網）	11
圖 3-2 決策樹元素表	16
圖 4-1 統計樣本說明圖.....	21
圖 4-2 機器學習樣本說明圖.....	21
圖 4-3 各變數時序表.....	24
圖 4-4 各模型績效比較圖	39
圖 4-5 各模型績效比較圖	49
圖 4-6 模型預測比較圖.....	50

第一章 緒論

第一節 研究動機與目的

股票為經濟的領先指標，常被稱為國家經濟的櫥窗。且臺灣股市高度發達，截至 2021 年 4 月為止，台股累積開戶人數更是高達一千一百五十四萬，平均每兩人就有一人參與股市的投資。因此，可以看出在臺灣的金融市場活動中，股市乃不可或缺的一環。Thaler and Ganse(2015)提到，在過往，由於大部分經濟理論皆有兩種最基本的假設：效率市場以及理性的行為，金融市場的價格被認為不會受到不理性的行為所影響，且市場價格的變動符合隨機漫步理論而無法預測。然而，由於現實世界投資人的決策會受到市場消息及個人情緒的影響，投資人不可能完全保持理性。並且因內線交易的存在，各國的股票市場也不遵循效率市場理論中資訊充分披露的條件。因此市場的報酬率有別於效率市場應呈現的常態分佈，反而會因為黑天鵝事件的發生，投資人的追漲殺跌等原因呈現出一種厚尾分佈，也就是市場大漲大跌的機率比常態分佈高的現象。所以預測股價的趨勢並從中獲利便有了可能性。

為了在蓬勃發展的股市中獲利，在過去，投資人通常是利用技術分析指標如隨機指標(Stochastic Oscillator，又名 KD 指標)、相對強弱指標(Relative Strength Index，又名 RSI 指標)、平滑異同移動平均線指標(Moving Average Convergence Divergence，又名 MACD 指標)、乖離率(Bias Rate)或總體經濟指標如各國匯率、利率、國際指數、恐慌指數、領先指標等預測及評價股價。然而，僅透過股價歷史資訊及總體經濟指標來預測股價的方法並沒有考慮到市場消息以及投資人情緒的影響力。由於財經新聞可反映出世界經濟與政策的不確定性，為了盡可能提高預測股價的準確度，我們認為大眾媒體發布的財經相關新聞也必須納入股價預測的考量中。在 1980 年代，網路還不普及時，財經新聞就已常常影響投資者的決策行為，財經新聞可透過自身媒體影響力主導股價波動，影響散戶的投資決策，甚至還發生過華爾街日報財經新聞記者勞勃·福斯特·威南斯(Robert Foster Winans)，利用自身媒體影響力操縱股價的醜聞。現今網路通訊技術發達，投資人能接受到的消息來源較廣，因此便較難發生單一專欄記者透過自身媒體影響力，影響股價的事件¹。但是各新聞媒體所發出的消息仍會一定程度的造成經濟政策的不確定性進而影響股價。然而，以往學者如 Bernanke(1980)、Bloom(2007)對這種

¹ 參考自本文作者康韡瀚個人部落格 <https://hugo-kang569.medium.com/>

經濟政策的不確定性的研究多僅停留在文字性的敘述上，一直到 Baker、Bloom and Davis(2016)三位學者提出一個衡量不確定性的指標-經濟政策不確定性指標(Economic Policy and Uncertainty，以下簡稱 EPU 指標)，經濟政策的不確定性才有個具有公信力的指標。然而，已編列的 EPU 指標為月資料的形式，無法配合股市每日的漲跌，因此，為了更精準的預測股市每日的漲跌，本文除了在以往文獻進行預測臺灣加權指數常使用的總體經濟指標如匯率、利率、國際指數等變數的基礎之上，也參考 EPU 指標編制的方式，透過在財經新聞上對關鍵字使用爬蟲技術編制新聞性經濟指標、新聞性政治指標、新聞性不確定性指標(以下簡稱三個指標為新聞性指標)代表特定關鍵字的新聞熱度。此外，我們也將上述爬蟲取得的資料進行情緒分析，以同時考慮大眾媒體對經濟政策不確定性的看法偏向正向或是負向。最後，由於上述兩種指標反映的主要是來自大眾媒體的看法，因此我們也利用類似的方法編制同時期、同地區的 Google 搜尋趨勢，以同時囊括投資人對本文所選關鍵字關注的熱度並將上述三種新聞性指標與總體經濟指標建立模型以期得到對臺灣加權指數良好的預測能力。

在研究方法方面，2017 年 AlphaGo 戰勝中國世界圍棋冠軍柯潔證明人工智能在處理大量資料運算的領域已全面超越人腦，人工智能逐漸成為現今社會最夯的話題。機器學習為人工智能的一種，由於其卓越的處理能力，已經被普遍運用於金融科技領域，因此為了盡可能精準的利用上述提及的變數預測股市價格，本文採用機器學習進行分析預測，但單單依靠機器學習只能得到各指標對於預測的重要性，無法如傳統統計方式般可更直觀的看出指標影響台股的方向，所以我們同時採用三種機器學習的演算法及傳統統計方法中的多元線性迴歸及門檻迴歸，希望能在兼顧預測精準度的同時也能判斷各個指標在不同結構點時對台股的影响方向。

第二節 研究流程

本文主要為探討新聞性與臺灣加權指數的關係，共分五章，第一章為緒論，第二章為針對本文所採用總體經濟指標對股價關聯與新聞性指標建構及應用等方面的研究作回顧，第三章為研究方法，而本文除了以傳統統計方法作預測外，也利用機器學習來比較績效，第四章為研究實證結果，第五章則為本文的研究結論與建議，研究流程如圖 1-1。

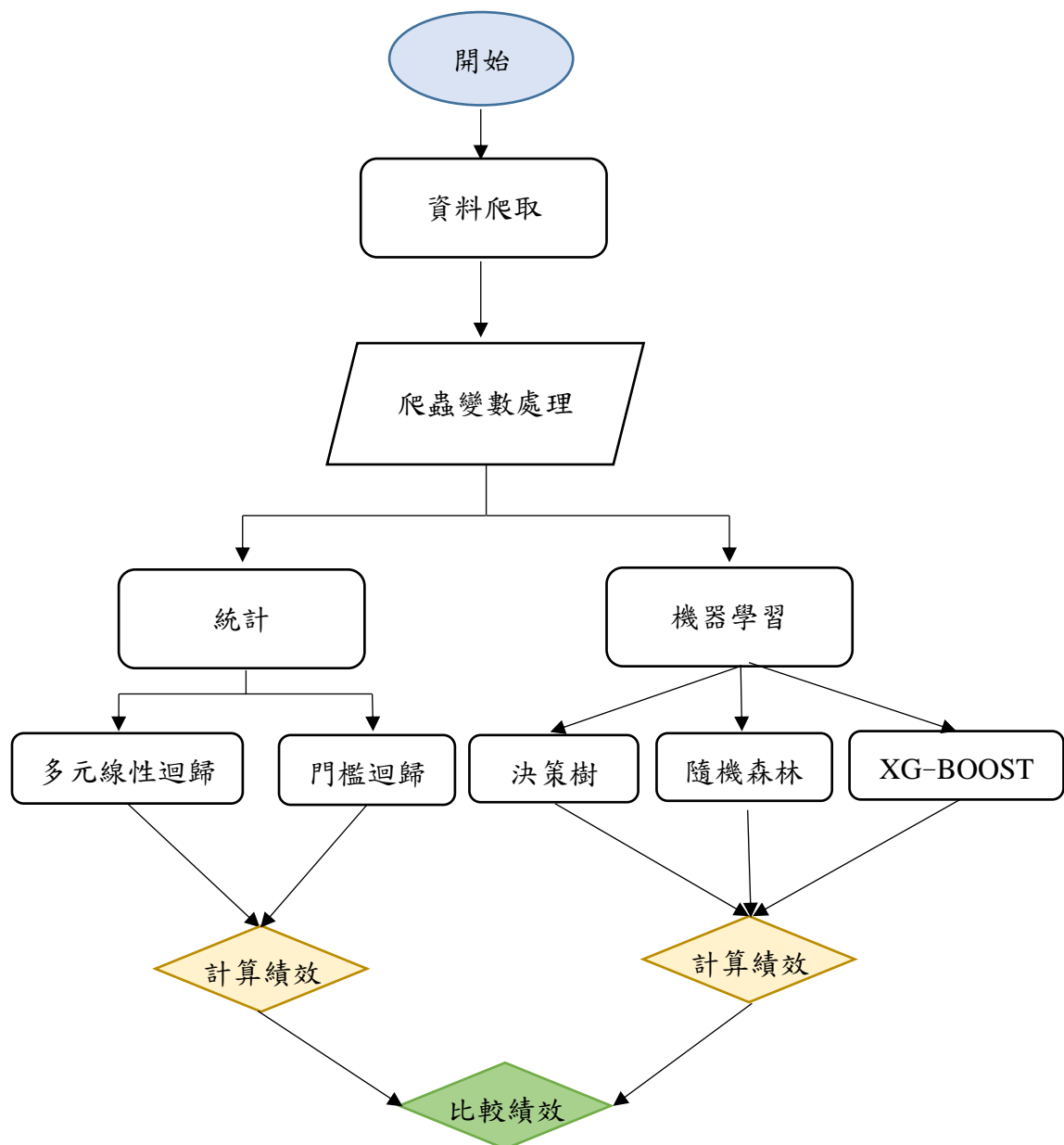


圖 1-1 研究流程圖

第二章 文獻探討

財經新聞對資產價格波動的影響不可忽略，若能掌握新聞消息的脈動，則有機會擬定獲利更佳的投資策略。然而截至目前為止，已編製的相關指標多半為月資料，無法即時反映出股市每日的漲跌。我們希望嘗試藉由量化每日臺灣新聞文字的情緒及訊息來增加對臺灣加權指數每日漲跌預測的精準性。

為了驗證本文編制的新聞性指標是否能在以往文獻股價預測的基礎之上增加預測的精準度，故本文選取總體經濟指標與股價做為基礎模型，再逐一新增本文編制的三個新聞性指標進行比較。本章首先介紹常用於預測股價之總體經濟指標對股價關聯的相關文獻，同時考量到僅運用總體經濟指標來預測股價的方式忽略了新聞性以及新聞情緒的影響力，故第二節介紹新聞性指標建構及應用的相關文獻。

第一節 總體經濟指標對股價關聯之相關文獻

一般而言，常見被學者²用於進行與臺灣股價關聯性影響研究的總體經濟指標可分為與貨幣市場有關的總經指標如匯率、利率、貨幣供給、物價指數，與產出及就業有關的總經指標如失業率、景氣訊號燈、景氣指標、進出口貿易總額，以及常見的金融市場標的如大宗物資與國際指數等等。匯率是兩國貨幣間的兌換比率，以出口導向的臺灣的匯率為例，若匯率走高，即臺幣貶值，會導致臺灣的出口企業每收到一枚外幣所能兌換的臺幣變多，能因此賺取更多的利潤進而導致股價的上漲。反之若匯率走低則有益於進口企業的獲利。翁小蘅(2009)以向量自我迴歸模型探討匯率、利率及加權股價報酬率三個變數間的關聯性。其選擇美金對臺灣匯率、商業本票次級市場 90 天期定盤利率、加權股價指數之日報酬率進行分析。實證結果發現，匯率變動率對股價報酬率具有單向的因果關係，利率變動率對匯率變動率具有單向的因果關係。且就長期累積效果而言，匯率變動率和利率變動率與股價報酬率具有負向關係。謝文馨(2008)以多元迴歸模式探討股價指數與領先指標和景氣同時指標包含項目及匯率、利率、黃金價格、原油價格、納斯達克指數等十二項總體經濟因素之間的關係。實證結果為：匯率與股價指數有顯著的負向關係以及利率與股價指數有顯著的正向關係。然而，吳怡慧(2011)的

² 翁小蘅、謝文馨、吳怡慧、楊怡俐…等學者。

研究結果卻與上述兩位學者得出了不同的結論，其以向量自我迴歸模型探討臺灣加權股價指數與領先指標指數、同時指標指數、貨幣供給、消費者物價指數、利率及匯率等總體經濟變數之關聯性。實證結果則發現匯率與臺灣股價指數為正向關係。另外，楊怡俐(2009)發現，股價與匯率的關係也不總是匯率單方面的影響股價。其以向量自我迴歸模型探討總體經濟變數與股票報酬率間的關聯性，以上市股票報酬率、進出口貿易總額、失業率、利率及匯率等變數分析，其實證結果為：股票報酬率為匯率之領先因子，股價之漲跌會影響匯率之升貶，因此，在短期可利用股票報酬率來做為匯率升貶之判斷工具。

與實體產出面有關的總經指標通常由一國政府統計並公布數據，如失業率、景氣指標等等。這些指標可以反映國內的經濟概況，進而影響到投資人的決策及股價。吳怡慧(2011)發現領先指標指數與臺灣股價指數為正向關係，謝文馨(2008)則提到製造業存貨指數與臺灣股價指數有顯著的負向關係。總經指標也可以用來預測臺灣股價指數：魏楚蕎(2015)透過 K-最近鄰近法、倒傳遞類神經、線性迴歸、支持向量機此四種資料探勘技術及灰預測來預測臺灣加權股價指數，選取美元 30 天期匯率、基本放款利率、臺灣地區消費者物價總指數、存款貨幣、準貨幣、M1A、M1B、M2、臺灣痛苦指數、景氣領先指標綜合指數對臺灣加權股價指數進行預測，實證結果主要為：當股市處於盤整期間時，以匯率、利率、消費者物價指數和貨幣供給類變數組成的模型適合用來預測臺灣股票加權指數。

在資金高度流動且現代通信技術發達的現代，金融機構及大型公司間跨國投資的繁榮導致了各國股市間相互連動的現象。股價指數可以反映一國股市的概貌，因此國際指數通常與臺灣股價指數有關聯性。陳建州(2005)以向量自我迴歸模型探討投資台股的擇時策略，其選擇臺灣加權股價指數、標普 500、景氣對策信號、匯率、利率、M1b 相對 M2 增幅、實質國內生產毛額、週轉率等變數進行分析。其實證結果為：當期股價除了受當期 S&P500 股價指數的影響外，尚受到前一期股價指數、M1b 相對 M2 增幅與週轉率的影響。謝文馨(2008)也發現那斯達克指數與股價指數有顯著的正向關係。國際指數在預測臺灣股價指數時也有不錯的表現：王鼎宏(2016)應用機器學習的方法選取匯率、利率、國際指數、蘋果股價及加權指數的成交量、本益比、股價淨值比對臺灣加權股價指數進行預測，其發現由那斯達克指數、道瓊工業指數、臺幣對美元匯率、大盤股價、股價淨值比這五項數據所組成的模型有最佳的預測能力，且最為穩定。另外，他同時也發現近三年來蘋果股價的預測能力逐漸提高。

綜合上述文獻所示，過去研究臺灣加權股價指數與總體經濟關聯性的文獻主要以向量自我迴歸模型、迴歸分析法為研究方法，使用領先指標、同時指標、貨幣供給、利率、匯率、國際指數以及加權指數之前一期價格與成交量等等變數進行分析。因此，本文總體經濟變數方面以投資人的角度出發選取較易取得日資料的人民幣兌美元匯率、歐元兌美元匯率、英鎊兌美元匯率、日圓兌美元匯率、台幣兌美元匯率、臺灣銀行一年期定存利率(下稱臺銀一年定存利率)、道瓊工業指數、上海證券交易所綜合股價指數(下稱上海綜合指數)、標普 500 指數、那斯達克綜合指數、日本經濟平均指數、芝加哥選擇權交易所波動率指數(下稱恐慌指數)、蘋果公司股價以及加權指數的成交量、本益比與股價淨值比並加入我們自行編撰之新聞性變數對臺灣加權股價指數進行預測，期望能夠透過新聞性變數的加入提高對臺灣加權股價指數預測的精準性。

由上述文獻可看出，以往對於股價預測的研究方法以傳統統計方法為主，但在如今金融全球化，金融市場錯綜複雜的當今，許多對股市價格來說十分重要的因子往往不是以線性關係的方式來影響股價。且近年來機器學習蓬勃發展，其演算法如類神經網路、向量機、隨機森林等對股價的預測能力已被證實比起傳統迴歸方法的更為優秀。故本文之研究方法採用機器學習中常見的隨機森林、決策樹以及 XGBoost，同時也嘗試以傳統迴歸方法中的多元線性迴歸及門檻迴歸建立模型並比較傳統迴歸方法與機器學習的績效之差別，希望能提供投資人當作投資臺灣股票市場決策之參考。

第二節 新聞性指標建構及應用之相關文獻

本節將整理並介紹新聞性指標：(1)經濟政策不確定性指標(Economic Policy Uncertainty Index，下稱 EPU 指標)、(2)情緒指標、(3)Google 搜尋趨勢的相關文獻。

一、EPU 指標之相關文獻

在過去，學者為了衡量不確定性建構出 VIX 指標，即是恐慌指數，是因當市場出現「恐慌性拋售」時，跌幅急且令投資者對未來波幅預期上升，所以 VIX 通常在「恐慌市」中急升，此指標是將系統性風險所帶來的不確定性數據化，建構方法則是一系列標準普爾 500 指數期權的價格加權指數，其目的是要用來估計標準普爾 500 指數在未來 30 天的隱含波動性，然而 Baker、Bloom and Davis(2016)

三位學者提出另一個衡量不確定性的指標-經濟政策不確定性指標(Economic Policy and Uncertainty，以下簡稱 EPU)，建構方法分為三個部分，第一部分是藉由通過統計美國「the Wall Street Journal」、「USA Today」、「the Washington Post」、「the New York Time」……等 10 家大型報社中與經濟政策不確定性有關的文章數目來衡量經濟政策的不確定性，第二部分是通過統計每年稅法法條數目的修訂來衡量稅法變動的不確定性，第三部分是通過消費者物價指數（CPI）和聯邦及地方州的政府支出的不一致來建構出不確定性的指標，此 EPU 指標的走勢與 VIX 的走勢具高相關性（圖 2-1），所以在很大程度上，兩者可反映相同的經濟情緒，且 EPU 指標的建構方法相對簡單，可以成為 VIX 以外另一經濟風險指標，也方便各方的學者可以透過 EPU 來增加更多的研究方向。

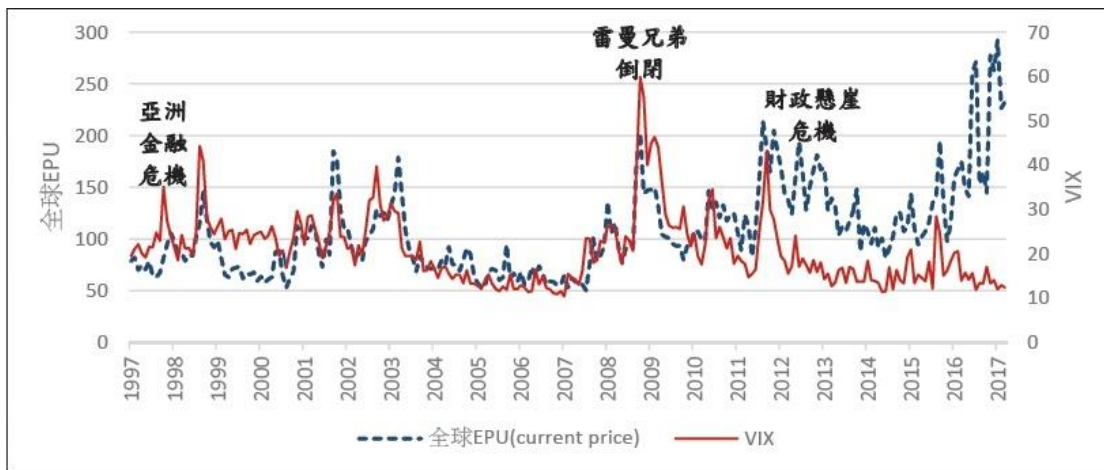


圖 2-1 恐慌指數和 EPU 指標(擷取自彰化銀行商品策畫處報告³)

EPU 指標對多項重大經濟變數都有影響能力：如劉靜玟(2018)發現中國與美國 EPU 指標和上海與深圳證交所的 A、B 股呈顯著的負向關係、Baker et al.(2016)發現 EPU 指標與許多總體經濟變數如就業率和經濟成長以及 SP500 都呈現負向關係，以及 Huang and Luk(2018)發現用中國大陸當地報紙編製而成的 EPU 指標對中國大陸的就業率和產出有負向關係。且根據黃裕烈（2020）的報告指出，臺灣的 EPU 指標對臺灣的工業生產指數、大盤指數等重大經濟變數有顯著的預測以及解釋能力。然而，目前大多已建構的 EPU 指標及上述文章使用的 EPU 指標(<https://www.policyuncertainty.com/>)皆為月資料，難以直接運用在實際的交易決策中。因此，為了更精準的預測及解釋臺灣大盤指數的波動，我們藉由建構日資料的臺灣新聞性指標，以配合股市每日的變動。本文建構的新聞性日指標

³ 圖片來源 https://www.bankchb.com/chb_2a_resource/leap_do/gallery/1498456331860_427cf.pdf

參考 Banker et al(2016)及黃裕烈(2020)的方法將關鍵字分為三種類別編制而成，其中黃裕烈(2020)編列臺灣月 EPU 指標的公式為：

$$\frac{\text{當月符合EPU關鍵字的新聞篇數}}{\text{當月新聞總篇數}} \quad (2-1)$$

本文原欲參考此公式編制臺灣日 EPU 指標，惟資料蒐集因當日工商時報的所有新聞篇數無法取得，因此我們斟酌後採由當日符合 EPU 關鍵字的新聞篇數做為衡量臺灣日新聞性的指標。

二、市場情緒指標之相關文獻

市場情緒指標是反應市場上樂觀或悲觀的程度的指標，包括直接調查投資人看法而形成的直接指標，以及透過統計市場交易數據而形成的間接指標。其中間接指標的種類繁多，目前較常見的情緒代理指標有 VIX 恐慌指數、選擇權 P/C ratio、道瓊指數波動率及納斯達克波動率指數等等。

相較於 EPU 指標反映的更多是新聞媒體關注的金融市場趨勢與國家政策，新聞情緒指標則是側重在大眾媒體對景氣和經濟現況的看好與否，當市場呈現多頭趨勢的時候，媒體人在撰寫新聞時也會使用更多正向的字眼並影響到接受到訊息的投資人，反之媒體人則會傾向使用負面的字眼，同樣會在一定程度上影響接受到訊息的投資人之決策判斷。因此，本文認為在考慮新聞性時也有必要將新聞情緒指標納入考量。

大眾媒體不僅影響到一般的投資人，Shiller(2002)提到，大眾媒體對非專業者的影響大於專業的投資人，因此即使是專業投資人也會受新聞媒體的影響而做出不理性的投資決策。此影響可能會進一步造成資產價格發生不可預期的波動。

而以財經新聞所建構的情緒指標也被證實對股市價格有顯著的影響：Baker and Wurgler(2006)發現當以財經新聞建構之投資人情緒指標很高時，股票對於觀望中的投資人及樂觀主義者會更加有吸引力，但對套利者的吸引力則是下降。且小型股票、沒有配發股利的股票、波動性高的股票及經營狀況較差公司股票之報酬率會下降。投資人情緒指標較低時，上述幾種股票能顯著的擁有超額收益。詹佑維(2019)也發現以財經新聞建構之投資人情緒指標在正向時會使隔日加權指數之報酬變動率上升，但對於期貨報酬率之波動程度卻較為平穩。

常見的情緒代理變數如恐慌指數、賣權買權未平倉比、信心指數、外資情緒指標等也被發現具有對股市的多空研判能力：江秉倫(2020)發現利用九個情緒代理變數進行主成分分析合成之情緒指標可作為單一指標判斷市場是否處於不理性上漲或下跌，當投資人過度樂觀或悲觀時常常使股市偏離內在價值。李盈屏(2016)則發現外資情緒指標對台股報酬率及期貨報酬率有顯著的正向關係等等。

在進行情緒分析的時候，通常分為兩種方法：字典法以及機器學習法。字典法將分類好正面及負面的字詞整理成一本字典，將文本與該字典進行對比，出現在字典中的正面詞彙為正向，出現在字典中的反面詞彙則為負向。機器學習法則是主要以已經標記好的資料進行分類學習，最後再利用分類進行分析。本文採字典法進行情緒分析。

三、 Google trend(下稱 Google 搜尋趨勢)之相關文獻

Google 搜尋趨勢，舊稱 Google insights for search，為 Google 在 2008 年時提出的一項免費服務，可以透過輸入關鍵字來進行該關鍵字指定地、時間、所屬領域、資料來源範圍的時間序列及大眾關注的熱度分析。有別於以財經新聞編制的新聞性指標，Aouadi(2013)發現 Google 搜尋趨勢和投資人的關注成高度相關。也就是說由於 Google 搜尋趨勢反應的是在指定時間內使用 Google Chrome 搜尋該關鍵字對比最高點的相對熱度，Google 搜尋趨勢可以反應除媒體人外社會大眾對關鍵字的關心程度，甚至透過設定指定地及所屬領域可以指定關鍵字的相關領域以及知道趨勢主要集中在哪些地區，因此本文納入 Google 趨勢以完善新聞性所遺漏的投資人情緒。

四、 小結

由上述文獻可知，EPU 指標以及市場情緒指標以及 Google 搜尋趨勢對股價確實有一定的影響能力及預測能力，因此本文在匯率、利率、國際指數等十六個總體經濟指標的基礎上納入本文根據 EPU 指標的方式編列而成的新聞性指標、新聞情緒指標以及 Google 搜尋趨勢，並採用機器學習中的隨機森林、決策樹與 XGBoost 以及傳統迴歸方法中的多元線性迴歸與門檻迴歸對臺灣加權指數進行預測並比較兩種方法的預測精準度。

第三章 研究方法

本章共分三節，第一節是資料抓取與處理方法，抓取方式為使用爬蟲技術，獲取新聞資料，編纂本文新增變數，第二節是統計方法，使用多變量迴歸分析法以及門檻迴歸，試圖探討本文第二章第一節所彙整之變數對臺灣加權指數的影響，及其影響的方向為何，並藉此預測臺灣加權指數漲跌，第三節是機器學習方法，使用的模型為決策樹、隨機森林及 XGBoost，預測臺灣加權指數漲跌方向。

第一節 爬蟲變數處理方法

第一章提及 Thaler and Ganse(2015)認為投資人的決策會受到市場消息及個人情緒的影響，投資人不可能完全保持理性。因此本文參考前章提及 EPU 算法，編纂本文新增新聞性變數。⁴

一、新聞性變數

(一) 爬蟲

本文主要探討財經新聞預測股價趨勢，為避免有過多娛樂新聞和八卦新聞影響，並參考李嘉洲(2016)應用深度學習於財經新聞來源對股價趨勢預測之研究，因此選用以財經新聞為主的工商時報作為實驗資料來源。

本文以 Python 中的 BeautifulSoup 套件進行爬取工商時報的資料，Beautifulsoup 是一個 Python 的函式庫模組，可以讓開發者僅需撰寫少量程式碼就可以快速解析網頁的 HTML 碼，從中萃取出使用者所需資料。下圖 3-1 為工商時報中時新聞⁵，再爬取資料時必須先取得該網站網址，並從 HTML 中尋找所需資料的標籤節點利用 BeautifulSoup 套件爬取所需資料，再利用 Python 中的 Pandas 套件將資料區分為日期，標題，網址與內容，轉存為 csv 檔方便日後計算指標，即如圖 3-2，內文欄位為網站爬取的新聞內容。

⁴ 網站來源 <https://hugo-kang569.medium.com/>

⁵ 網站來源 <https://www.chinatimes.com/newspapers/2602?chdtv>

The screenshot shows the China Times website interface. The top navigation bar includes sections like '即時' (Real-time), '政治' (Politics), '言論' (Opinion), '生活' (Life), '娛樂' (Entertainment), '社會' (Society), '財經' (Economy), '科技' (Technology), '寶島' (Treasure Island), '話題' (Topics), '有影' (Photos), '國際' (International), '軍事' (Military), '兩岸' (Cross-strait), '時尚' (Fashion), and '運勢' (Fortune). The main content area is titled '工商時報' (Business Times) and features three main news items:

- 大立光、先進光 冤家變親家**: A financial news article dated 2021/03/10, discussing the relationship between Daxi Optics and Advanced Optics.
- 國房稅 第四戶起課稅2~4%**: A financial news article dated 2021/03/10, reporting on the proposed national housing tax rates.
- 新台幣 躍升今年最強亞幣**: A financial news article dated 2021/03/10, highlighting the New Taiwan Dollar's performance as the strongest Asian currency.

圖 3-1 工商時報新聞範例（擷取自工商時報中時新聞網）

A	B	C	D
日期	新聞標題	網址	內文
2021/3/10	大立光、先進光 冤家變親家	https://www.chinatimes.com/news/2602?chdtv	從敵人變朋友?大立光
2021/3/10	國房稅 第四戶起課稅2~4%	https://www.chinatimes.com/news/2602?chdtv	財政部研議中國房稅
2021/3/10	新台幣 躍升今年最強亞幣	https://www.chinatimes.com/news/2602?chdtv	外資9日擴大匯出，新

圖 3-2 爬蟲範例

（二）JIEBA 斷詞解釋

詞是最小有意義且可以自由使用的語言單位。任何語言處理的系統都必須先能分辨文本中的詞才能進行進一步的處理，例如語言分析或情緒分析。因此中文自動分詞的工作成為了語言處理不可或缺的技術。中文句子中，詞與詞之間，不像英文句子的詞透過空格隔開，因此中文需要透過中文斷詞系統將詞做出分隔。我們使用的斷詞系統：

結巴(JIEBA) 是 Python 語言中的中文斷詞套件⁶，結巴所採用的方法為詞庫式斷詞法，將文件上的資料與詞庫中的詞語進行比對，進行斷詞。詞庫式斷詞法有一個缺點，當自身詞庫量不足，或是文件中有詞不存在於詞庫時，會造成斷詞錯誤，但是結巴可透過增加自定義字典解決問題。

表 3-1 新聞內文斷詞範例

新聞原稿範例為上文提及爬蟲技術所蒐集之新聞資料，使用結巴進行斷詞後，將詞語詞之間分隔，結果列在新聞內文中文斷詞範例。

新聞原稿範例

美國聯準會主席鮑爾（Jerome Powell）1 日出席國會聽證會，他在預先準備的書面證詞表示，自新冠疫情春季爆發以來，Fed 為了穩定信貸市場推出一系列措施，而這些年底到期的緊急貸款計劃協助釋放將近 2 兆美元資金，暗示此時並非退場時機。他亦指出，隨著近日確診病例攀升，美國經濟前景恐陷入「高度不確定性」。(資料來源工商時報中時新聞網)

新聞內文中文斷詞範例

美國 聯準會 主席 鮑爾 () 日 出席 國會 聽證會 他 在 預先 準備 的 書面 證詞 表示 自新 冠 疫情 春季 爆發 以來 為了 穩定 信貸 市場 推出 一 系列 措施 而 這些 年底 到期 的 緊急 貸款 計劃 協助 釋放 將近 兆

(三) 使用字典

本文使用的字典有兩個，第一個為結巴本身的詞庫，其次為本文自行編撰的使用者定義字典，可彌補結巴詞庫量的不足。

二、情緒分析 (SENTIMENT)

第二章提及以前的研究者發現大眾媒體會影響包括專業投資人在內的投資者的投資決策近而導致資產價格波動，因此本文分析大眾媒體在新聞用字的情緒編纂新聞情緒指標。本文以字典法的方式做情緒分析的研究，參考謝鎮宇 (2010) 意見探勘在中文評鑑語料之應用 和李政儒、游基鑫、陳信希(2012)廣義知網詞彙意見極性的預測採用的字典。

⁶ 中文斷詞套件來源 (<https://github.com/APCCLab/jieba-tw>)

（一）字典選取

採用 NTUSD 提供的中文情感詞典(Nation Taiwan University Sentiment Dictionary)⁷，裡面包含了正面以及負面情緒的詞彙集，正面情緒大約有 2800 筆詞彙，負面情緒大約有 8000 筆詞彙，總共大約 10000 多筆詞彙。但是此字典只有正負向兩種極端詞彙，不包含正向程度和負向程度。

表 3- 2 NTUSD 正負向詞舉例

正向字	不凡的、公平的、友善的、引人注目的、出眾的
負向字	不公平的、不友善的、不正直的、劣等的、不好的

(資料來源:NTUSD)

（二）計算指標的方法

藉由上文提及爬蟲技術所蒐集之新聞資料，分析新聞標題及內文中的詞彙，以字典中的正負向詞，計算所選取新聞文中各日新聞情緒指標，計算出當天出現本文所選關鍵字的所有新聞中的正向詞個數減去負面詞個數。

每日情緒指標 = 該目標新聞中所有正向詞次數-該目標新聞中所有負向詞次數

三、Google 搜尋趨勢 (Google Trend)

Google 搜尋趨勢，可以讓我們用國家、搜尋類別和不同時間來了解使用者在 Google 上搜尋哪些關鍵字，關鍵字的搜尋熱度會介於 0 到 100 之間，此數值並不代表真正的搜尋次數，是以最高點做為比較基準，100 代表該字詞的熱門程度在該時間點達到最高峰。

計算指標方式：

我們選用資料處理後所得到字詞，作為 Google trend 的搜尋關鍵字，並選用臺灣地區財經類別。因為 Google trend 提供之長期趨勢為月資料，而短期趨勢為日資料，為獲取長期趨勢之日資料，將採以下方程式作轉換：

$$G_{day} = g_{month} \times \frac{g_{day}}{\sum g_{day}} \quad (3-1)$$

⁷ NTUSD 中文情感詞典來源 <http://nlg.csie.ntu.edu.tw/>

其中， G_{day} 為長期趨勢之日資料， g_{month} 為年趨勢之月資料， g_{day} 為月趨勢之日資料，透過轉換之後將每日每個字詞的 G_{day} 做平均，最後獲取本文的每日 Google Trend 指標。

$$\text{Google 搜尋趨勢指標} = \frac{\sum G_{day}}{\text{字詞個數}} \quad (3-2)$$

四、新聞性指標編制

如本文第一章及第二章所述，我們認為新聞性對臺灣加權指數的影響甚鉅。然而，在網路通訊發達，大眾媒體業蓬勃發展的現代，網路新聞除了財經類別以外，還同時包含娛樂、八卦、社會、人物等與經濟政策不確定性相關性較低的類別。因此，在利用爬蟲技術編制新聞性指標時，若無指定新聞類別，會有不相關的新聞混雜在其中導致新聞性指標對臺灣加權指數的解釋能力下降。因此，本文以工商時報財經新聞建構的新聞性指標作為衡量臺灣新聞性的標準，承接上述爬蟲取得原始資料，透過斷詞方法將爬下的新聞資料進行處理過後本文依照計算指標方式：新聞性日指標 = 當日符合關鍵字的新聞篇數來編制新聞性日指標。接著我們以日期為基準合併臺灣加權股價指數以及關鍵字的資料，整理出 2014/1/2 到 2020/12/31 共 1707 筆資料，並依照使用關鍵字分為政策性、經濟性、不確定性三類。

第二節 統計方法

一、多元線性迴歸分析(Multiple regression analysis)

多元迴歸是探討自變數與應變數之間的關聯性，所建立的迴歸模型，藉此預測研究者感興趣的應變數。本文先以文獻參考變數作為自變數中大部分解釋的因子，因為臺灣加權指數很難單單以本文新增變數全面地解釋，故除了參考文獻中常見的參考變數(如利率、蘋果股價、那斯達克指數、新台幣匯率……等等)以外，本文也加入不同關鍵字統計次數結果而形成的新聞性經濟指標、新聞性不確定性指標、新聞性政策指標、Google 搜尋趨勢與新聞情緒指標。試圖解釋臺灣加權指數。

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t \quad (3-3)$$

y_t :臺灣加權指數

x_{it} :總體經濟指標變數(經濟變數、新聞性經濟指標、新聞性不確定性指標、新聞性政策指標、Google 搜尋趨勢與新聞情緒指標)

β_i :樣本估計系數

ε_t :殘差項

上式迴歸式是以最小平方法估計參數，不僅能從單個變數求解，更能推廣至多個自變數的領域，在本專題研究中，往往很難界定多少個因子才能有效顯著解釋自變數所含相關變數，單單以本文新增變數試圖解釋臺灣股票加權指數，確實不忍卒觀，故本文中，以相關文獻作為自變數試圖增加解釋能力之勝義。為確保建構穩定模型之效率性，所提取之數據為以 2014 年至 2019 年作為預測樣本統計量的資料並搭配 2020 年的資料為預測目標，進而判斷正確率的結果。

而金融性時間序列資料特別需要注意一階或以上的自我相關問題，當資料本身會因為期與期之間因為動量效應而出現自我相關時，則必須要注意是否會影響迴歸本身估計的充分一致有效性，本文使用 Newey West 修正異質變異數與自我相關之方法修正時間序列資料增加穩定性，保持資料在分析過程中有效率之估計。

二、門檻迴歸

門檻迴歸為學者 Hansen 所建構的一門找尋資料異常結構性轉折點的迴歸估計方法，當資料為時間序列資料時，時常會因為國際間與總體經濟發展原因出現不尋常之改變轉折點，為了試圖刻劃不同時點下經濟概況之分析，時常使為門檻檢定作為分析工具。以臺灣實證論文關於此方法使用也不少數，吳貞宜(2007)報告中亦以門檻迴歸作為研究方法研究臺灣股票市場異常報酬之研究，故本文亦以門檻迴歸作為欲探討自選變數間是否會因為外生變數特定區間而對於被解釋變數有負向或正向之變化之研究方法。

門檻迴歸模型介紹(threshold regression)：

$$Y_{it} = u_i + \beta_i x_{it} + \varepsilon_{it} \text{ 若 } q_{it} \leq \gamma \quad (3-4)$$

$$Y_{it} = u_i + \beta_i x_{it} + \varepsilon_{it} \text{ 若 } q_{it} \geq \gamma \quad (3-5)$$

殘差最小化做為門檻估計值，並檢驗門檻的顯著性必須先檢定模型中是否存在於門檻效果。虛無假說為門檻效果不存在；而對立假說為門檻效果存在。

第三節 機器學習方法

機器學習是透過撰寫程式使機器自動從數據中分析學習規律，並對未知數據進行預測，主要方法有兩種，監督式學習與非監督式學習，兩者差別於數據是否有被定義，本文主要是採用監督式學習中最常見的方法，其中包含決策樹、隨機森林及 XGBoost，分別介紹如下。

一、決策樹（Decision tree）

決策樹主要來源於 Quinlan(1986)提出的 Iterative Dichotomiser 3(ID3)演算法和 (1992)提出的 C4.5 演算法，及由 Breiman(1984)提出的 Classification And Regression Tree(CART)演算法，其算法採用樹形的結構，使用一層一層的分支來達到最終的分類，利用已知的類別與多個特徵來進行樹狀結構的分類，並從中歸納出高機率的規則。

而決策樹由三種元素構成，分別為：

1. 根節點：包含樣本的全集
2. 內部節點：對應特徵屬性測試
3. 葉節點：代表決策的結果
- 4.

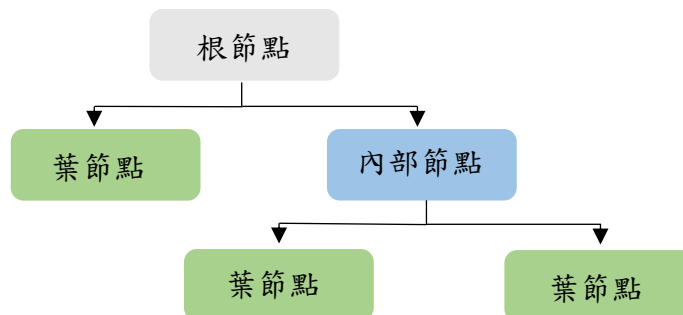


圖 3-2 決策樹元素表

決策樹的生成：

（一）挑出分類能力最好的內部節點：透過屬性選擇指標來評估資訊增益 (Information Gain)，資訊增益代表了在一個條件下，資訊複雜度（不確定性）減

少的程度。並選擇具有最大資訊增益值的特徵進行分支，本文使用基尼係數(Gini)作為屬性選擇指標來評估資訊凌亂的程度。

$$\text{Gini 公式: } G(S) = 1 - \sum_{i=1}^C p_i^2 \quad (3-6)$$

(二) 利用內部節點產生出對應的新分支，並針對每個新分支運行下一個內部節點的生成。

(三) 不斷重複上述步驟直到滿足分類的終止條件為止。

本文採用 Google Colab 作為 Python 工具，使用 Scikit-Learn 的決策樹套件，程式中的模型參數種類如下：

- (一) class_weight=None (自動給予數據集中的所有標籤相同的權重)
- (二) criterion=gini (使用基尼係數)
- (三) max_depth(限制樹的最大深度)
- (四) max_features=None(分支時最多考慮的特徵數，None=不限制)
- (五) max_leaf_nodes=None(最大葉子節點數，None=不限制)
- (六) min_samples_leaf (子節點最少樣本數)
- (七) min_samples_split (中間節點最少樣本數)
- (八) min_weight_fraction_leaf (子節點所有樣本權重和的最小值)
- (九) random_state (設置分枝中隨機模式的參數)
- (十) splitter=best(優先選擇更重要的特徵進行分枝)

決策樹透過一層一層的分支進行完整的分類，導致此模型產生過度擬合的問題，隨機森林以隨機方式建立由多顆決策樹組成的森林，每顆樹相互獨立且樣本數皆非全部的樣本，較不會有過度擬合的問題，因此本文第二個模型採隨機森林彌補決策樹的缺陷。

二、隨機森林模型(Random Forest)

隨機森林是由多個決策樹組成的監督式分類機器，本文透過 Bootstrap 方法，以隨機抽取但會放回的方式取得數個樣本而組成訓練資料，再使用多個決策樹進行投票來生成對大盤漲跌的預測。

本文採用 Google Colab 作為 Python 工具，使用 Scikit-Learn 的隨機森林套件，程式中的模型參數種類如下：

- (一) class_weight=None (自動給予數據集中的所有標籤相同的權重)
- (二) criterion=gini (使用基尼係數)
- (三) max_depth (限制樹的最大深度)
- (四) min_samples_leaf (子節點最少樣本數)
- (五) bootstrap=True (使用 Bootstrap Method)
- (六) min_weight_fraction_leaf (子節點所有樣本權重和的最小值)
- (七) max_features=auto (限制分枝時考慮的特徵個數)
- (八) random_state (設置分枝中隨機模式的參數)

效能衡量指標：

(一) 混淆矩陣 (Confusion Matrix)

混淆矩陣為分類演算法最常用來衡量效能的一種方式，由於本文主要目的是預測加權指數的漲跌是否正確，因此分為上漲(Positive)、下跌(Negative)兩類，並分別把真實狀況及預測結果的對比分成四類：

表 3-3 混淆矩陣範例表

混淆矩陣		預測結果	
		Positive	Negative
真實情況	Positive	TP (真陽性)	FP (偽陽性)
	Negative	FN (偽陰性)	TN (真陰性)

表 3-4 的說明如下：

- a. 真陽性 (True Positive, TP)：真實情況為真且預測結果亦為真
- b. 真陰性 (True Negative, TN)：真實情況為假且預測結果亦為假
- c. 偽陽性 (False Positive, FP)：真實情況為假但預測結果是真
- d. 偽陰性 (False Negative, FN)：真實情況為真但預測結果為假

註：本文 Negative 為下跌，Positive 為上漲

本文利用混淆矩陣，計算出以下指標，透過指標衡量機器學習的績效並加以調整來進行優化：

$$a. \text{準確率 (Accuracy)} = \frac{(tp+tn)}{(tp+fp+fn+tn)}$$

b. 精確率 (Precision) 為陽性的樣本中有幾個是預測正確的 $= \frac{tp}{(tp+fp)}$

c. 召回率 (Recall) 為事實為真的樣本中有幾個是預測正確的 $= \frac{tp}{(tp+fn)}$

d. F1-score 為精確率與召回率的調和平均數 $= \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$

由於精確率與召回率反映了不同面向的預測正確率，而 F1-score 為精確率與召回率的調和平均，因此本文除了考慮準確率外，主要衡量 F1-score 作為預測績效的評比標準。

(二) 特徵重要性 (Feature Importance)

使用特徵重要性判斷各個變數的重要性，再依據各個變數的重要性進行篩選，將不重要的變數刪除，而特徵重要性是透過分裂節點前後的不純度(impurity)減少量來判斷，也就是資訊增益(Information Gain)的程度，而本文使用基尼係數來計算不純度(impurity)。

隨機森林和 XGBoost 模型皆屬於集成學習，隨機森林的決策樹間是獨立的，而 XGBoost 每建構一個決策樹，新的決策樹都會修正以前訓練過的決策樹犯下的錯誤，進而優化目標函數，且於 Kaggle 競賽中許多冠軍皆使用 XGBoost 在他們的算法之中，因此本文第三個模型採 XGBoost。

三、XGBoost 模型 (eXtreme Gradient Boosting)

Boosting 就是將每個弱分離器組合成強分離器，可用於分類或迴歸，XGBoost 是通過加入新的弱學習器，來糾正前面弱學習器的殘差，不斷循環這個步驟，最終就會得到許多弱學習器運算出的最終答案，已得到較為準確的預測。之所以稱為 Gradient，是因為在添加新模型時使用了梯度下降算法來最小化損失。

本文採用 Google Colab 作為 Python 工具，使用 Scikit-Learn 的 XGBoost 套件，程式中的模型參數種類如下：

(一) max_depth (限制樹的最大深度)

(二) min_child_weight (決定最小葉子節點樣本權重和)

- (三) gamma (指定節點分裂所需的最小損失函數下降值)
- (四) subsample (控制對於每棵樹，隨機採樣的比例)
- (五) colsample_bytree (控制每棵隨機採樣的列數的佔比)
- (六) objective= binary:logistic (邏輯迴歸損失函數)
- (七) scale_pos_weight (使算法更快收斂)
- (八) random_state (設置分枝中隨機模式的參數)
- (九) seed (隨機樹的種子)

第四章 實證結果分析

本章共分為四節，主要探討第三章研究方法實證結果，第一節為樣本說明，說明時間選取和採用那些變數，第二節與第三節分別為敘述統計及統計分析結果，第四節為機器學習結果，首先使用基本變數製作基本模型，並逐一加入本文新增變數與期比較準確率與 F1-score。

第一節 樣本說明

一、統計模型期間說明

樣本期間（統計）：

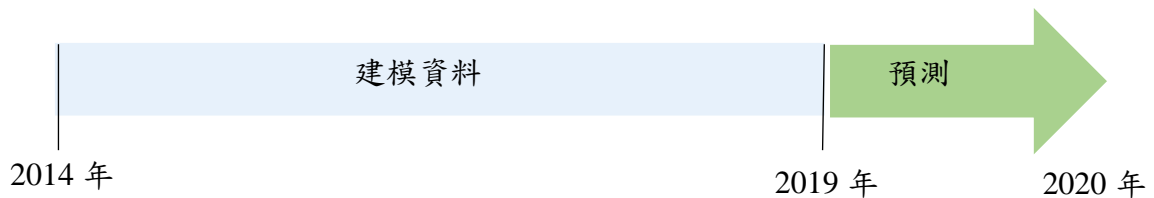


圖 4-1 統計樣本說明圖

在統計的部分，本文收集 2014/01/01-2020/12/31 資料，並將其切分為兩個時期，分別為建模期與測試期。建模期為 2014-2019，以大樣本數來降低準確率因標準誤隨機擺盪的機率，增加整體預測準確力，預測期間期為 2020/01/01-2020/12/31，用以衡量該模型之預測準確度。

二、機器學習期間說明

樣本期間（機器學習）：

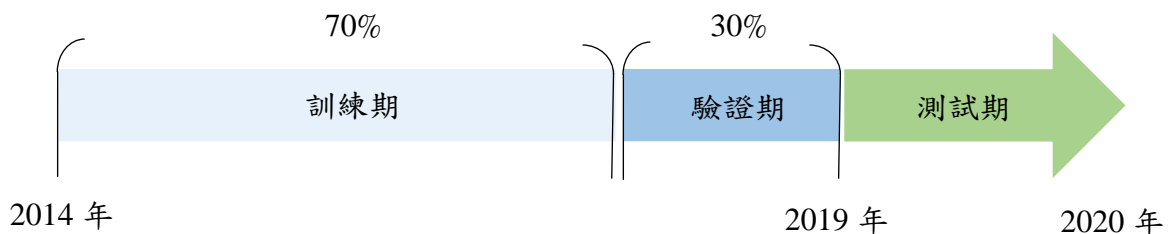


圖 4-2 機器學習樣本說明圖

在機器學習的部分，本文收集 2014/01/01-2020/12/31 資料，並將其切分為三個時期，分別為訓練期、驗證期、測試期。訓練期為 2014-2019 的前 70%，用來擬合模型，並通過設置機器學習模型的參數，訓練機器學習模型，以提高模型的

準確率及 F1-score，驗證期為 2014-2019 的後 30%，通過訓練集訓練出多個模型後，為了找出效果最佳之模型，使用各個模型對驗證集數據進行預測，並記錄模型準確率，訓練期和驗證期切分比例為參考機器學習常見使用切分比例，選取績效最佳切分比例。測試期為 2020/01/01-2020/12/31，通過訓練期和驗證期得出最優模型後，使用測試集進行模型預測，用來衡量該最優模型的性能和分類能力。

三、使用變數

本文參考前述第二節所提等前人所實證出的結果挑選出常見於運用在預測臺灣加權股價的變數...，列為經濟變數，並將經濟變數分為三類，匯率、股票指標及常見指標，再參考 Banker et al(2016)及黃裕烈(2020)等人的方法編制自創新聞性變數，列為本文新增變數。

表 4-1 使用變數分類說明表

		變數名稱
經濟變數	匯率	人民幣兌美元匯率 歐元兌美元匯率 英鎊兌美元匯率 日圓兌美元匯率 台幣兌美元匯率
	股票指標	那斯達克綜合指數 大盤股價淨值比 加權指數的成交量 蘋果公司股價 大盤本益比 標普 500 指數 上海綜合指數 道瓊工業指數 日本經濟平均指數
	常見指標	恐慌指數 臺銀一年期定存利率
本文新增變數		新聞情緒指標 Google 搜尋趨勢 新聞性經濟指標 新聞性不確定性指標 新聞性政策指標

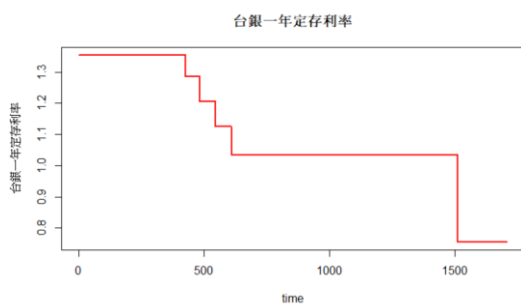
第二節 變數基本分析(傳統多變量迴歸)

一、基本敘述統計量

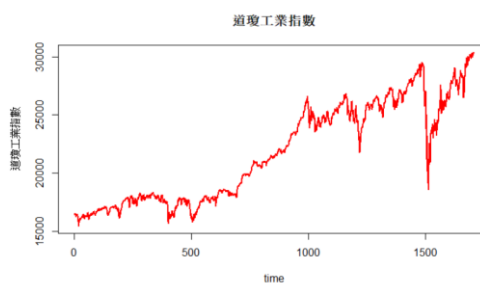
由表 4-2 與圖 4-3 基本敘述統計量可知國際指數類如道瓊工業指數、那斯達克綜合指數、日本經濟平均指數、標普 500 指數等在 2014 到 2020 年底間呈現穩定上升的趨勢，如道瓊工業指數從最小值 15440.23 點曾漲至 30409.56 點，那斯達克指數從最小值 6599.2 點也曾漲至 12899.42 點的高峰，可看出這七年間的全球股市市場愈加繁榮。蘋果公司股價、人民幣兌美元匯率、日圓兌美元匯率、英鎊兌美元匯率、日圓兌美元匯率也呈現上升的趨勢。而有關大盤的數據如大盤股價淨值比、本益比等則是呈現在一定基準上波動的現象。恐慌指數則是在 2019 年疫情及中美貿易戰爆發時攀上這七年的高峰，值得一提的是臺銀一年定存利率隨著時間逐年的降低，直到 2020 年 12 月 30 號為止近乎為 0 利率的水準。可以看得出來這幾年來因應美國聯準會的降息措施，臺灣央行也跟著執行降息政策。本文所編製的新聞性變數中，可以看出在 2020 年 3 月左右，隨著臺灣新冠肺炎的持續爆發，除了新聞情緒指標以外的新聞性變數皆有明顯大幅的上升，且大盤股價淨值比和本益比皆有顯著的下降。

圖 4-3 各變數時序表

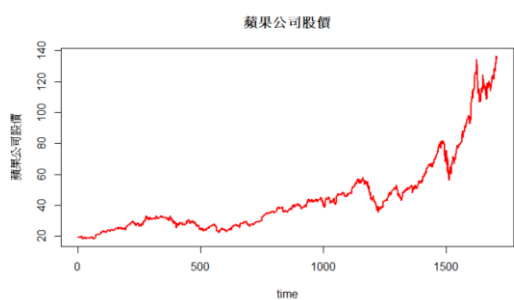
臺銀一年定存利率



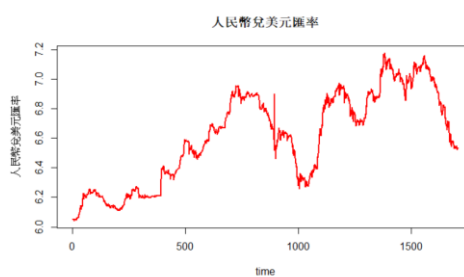
道瓊工業指數



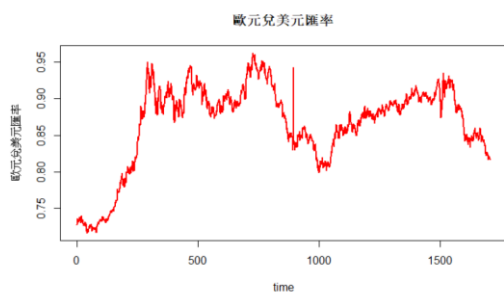
蘋果公司股價



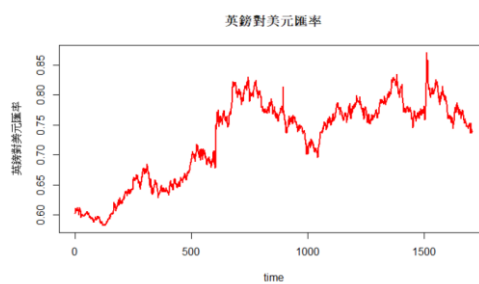
人民幣兌美元匯率



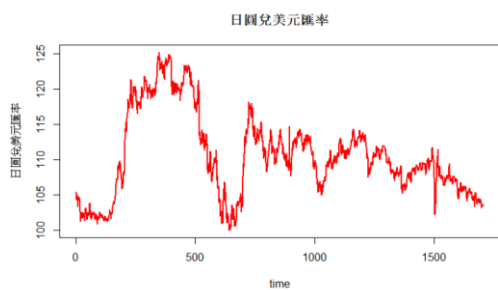
歐元兌美元匯率



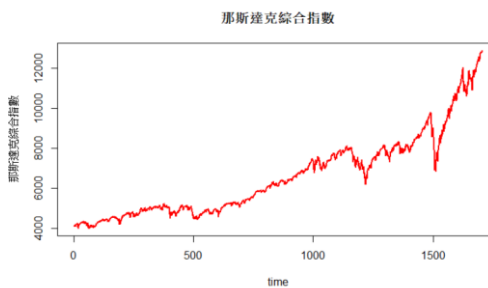
英鎊對美元匯率



日圓兌美元匯率

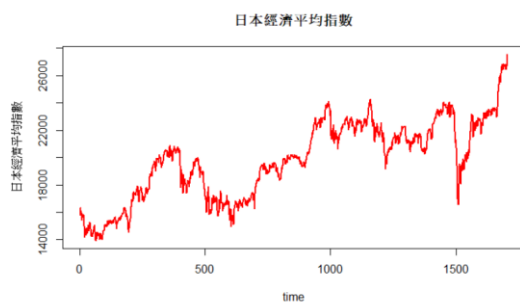


那斯達克綜合指數

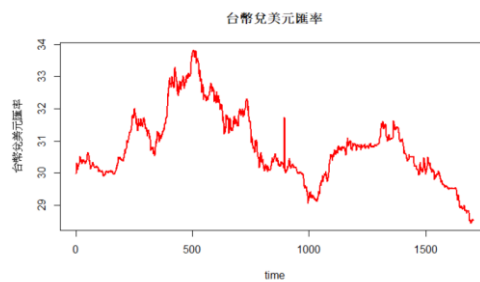


新聞性，你敢信？運用新聞不確定性與新聞情緒指標預測臺灣加權指數漲跌

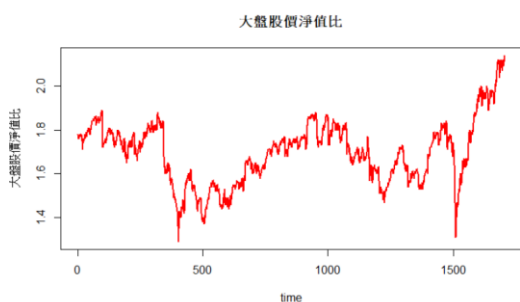
日本經濟平均指數



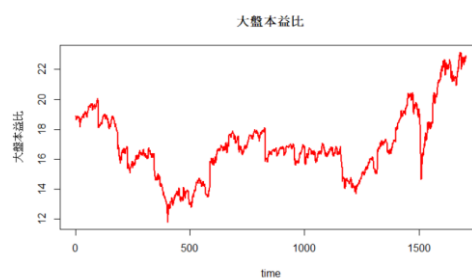
台幣兌美元匯率



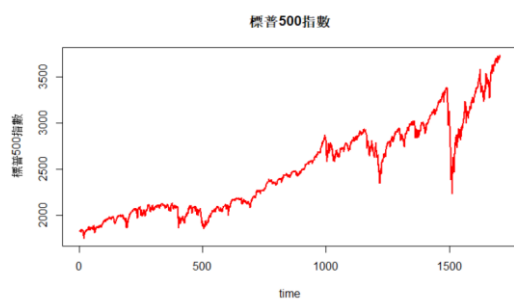
大盤股價淨值比



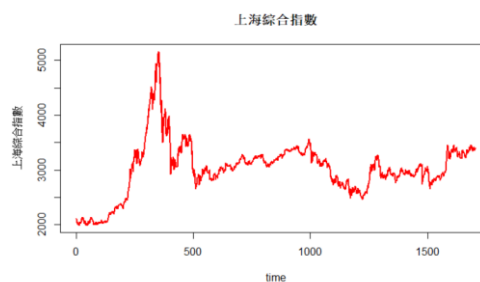
大盤本益比



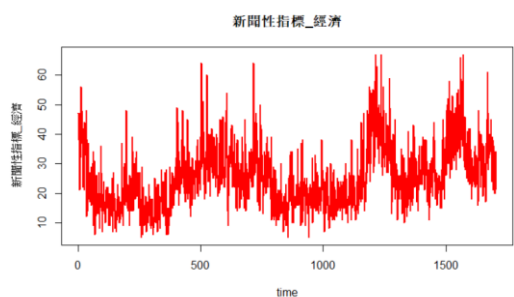
標普 500 指數



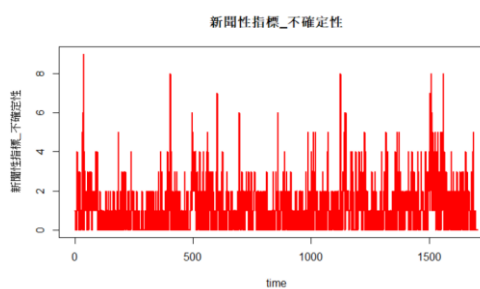
上海綜合指數



新聞性經濟指標

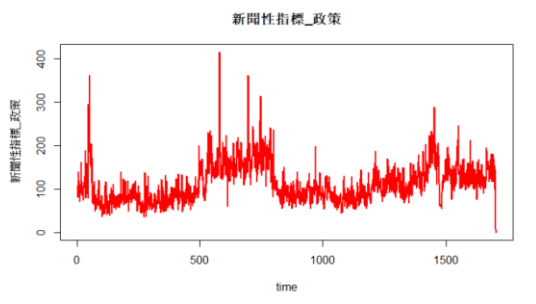


新聞性不確定性指標

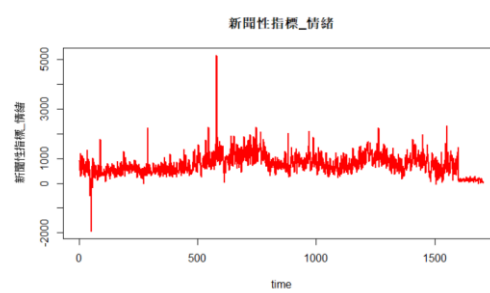


新聞性，你敢信？運用新聞不確定性與新聞情緒指標預測臺灣加權指數漲跌

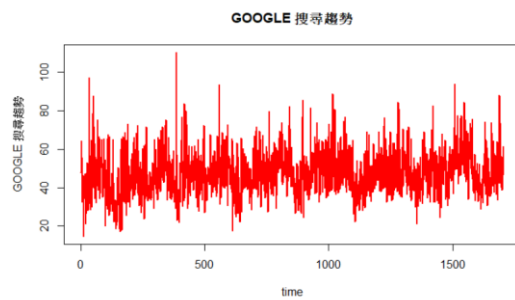
新聞性政策指標



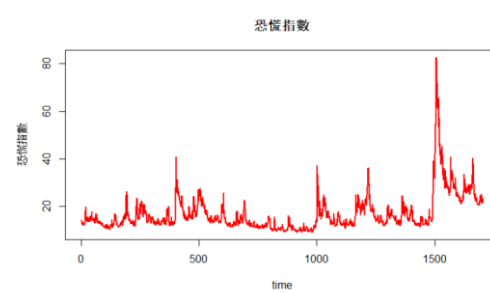
新聞情緒指標



Google 搜尋趨勢



恐慌指數



加權指數成交量

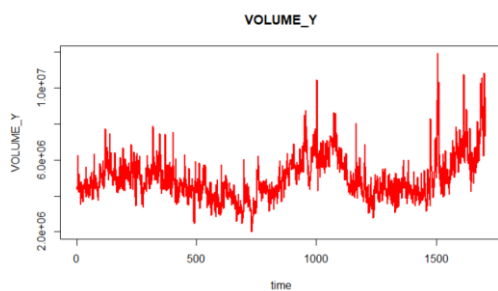


表 4-2 基本敘述統計量表

	平均數	最大值	最小值	標準差
臺銀一年定存利率(%)	1.10	1.36	0.76	-0.08
道瓊工業指數	21870.15	30409.56	15440.23	4255.12
蘋果公司股價	44.66	136.69	18.30	25.01
歐元兌美元匯率	6.61	7.72	6.04	0.31
人民幣兌美元匯率	0.87	0.96	0.72	0.06
英鎊對美元匯率	0.73	0.87	0.58	0.07
日圓兌美元匯率	110.59	125.22	100.03	5.90
那斯達克綜合指數	6599.20	12899.42	3999.73	2036.52
日本經濟平均指數	19807.94	27568.15	13910.16	2868.44
台幣兌美元匯率	30.81	33.84	28.44	1.07
大盤股價淨值比	1.61	2.14	1.29	0.14
大盤本益比	16.95	23.28	11.79	2.24
標普 500 指數	2490.25	3735.36	1751.64	478.24
上海綜合指數	3027.07	5166.35	1991.25	499.58
新聞性經濟指標	25.55	67.00	5.00	10.51
新聞性不確定性指標	1.14	9.00	0.00	1.26
新聞性政策指標	112.62	417.00	0.00	43.45
新聞情緒指標	763.31	5171.00	-1937.00	401.67
Google 搜尋趨勢	47.57	110.31	14.65	11.49
恐慌指數	16.94	82.69	9.14	7.81
加權指數成交量	4987998.00	11942087.0	2015119.00	1207452.00

第三節 多元迴歸結果

表 4-3 基本變數篩檢過程表⁸

基本比較變數	結果	第一次	第二次	結果
人民幣兌美元匯率	○ -48.57**	-146.27**	-208.89**		
歐元兌美元匯率	○ -925.14**	-740.09**	370.17*		2140.69**
英鎊對美元匯率	○ 2664.42**	1646.59**	1152.18**		
日圓兌美元匯率	○ -11.1**	-4.41**	-5.18*		-21.68**
台幣兌美元匯率	○ -78.46**	-65.84**	-61.2**		-113.35**
日本經濟平均指數	○ 0.02**	0.02**	-0.01*		0.07**
那斯達克綜合指數	○ 0.06*	-0.04*	0.12**		0.39**
大盤股價淨值比	○ 2875.75**	2750.35**	2386.63**		2391.26**
蘋果公司股價	○ -11.68**	-11.11**	-2.85		
大盤本益比	○ -28.06**	-8.01**	24.06**		-4.56**
標普 500 指數	○ 4.32**	4.4**	1.71**		
上海綜合指數	○ 0.1**	0.09**	0.15**		
恐慌指數	○ 0.44**	1.06**	-3.63**		-12.39**
臺銀一年定存利率	○ 743.74				
道瓊工業指數	○ -0.23**	-0.21			
加權指數成交量	○ 2.96E-05**	2.99E-05**	3.86E-05**		3.48E-05**

一、經濟變數分析

此部份為了確保多變量線性迴歸能完整有效率的對變數充分解釋，故先將參考論文中之經濟變數先做初步顯著性測試(以顯著水準 10%作為標準)，藉由逐步篩選將不顯著之變數刪除並挑選出對於臺灣加權指數具有顯著性之變數。

⁸ **與*分別代表掩住水準 5%及 10%顯著者

篩選變數過程與結果:根據文獻中經濟變數篩選對於臺灣加權指數是否有顯著影響，本文將共 16 種總體經濟變數刪減為 9 種，刪除依序為臺銀一年期利率、道瓊工業指數、蘋果公司股價、上海綜合指數、標普 500 指數、人民幣兌美元匯率、英鎊對美元匯率。

篩選結果為保留變數:歐元兌美元匯率、日圓兌美元匯率、台幣兌美元匯率、日本經濟平均指數、那斯達克綜合指數、大盤股價淨值比、大盤本益比、恐慌指數、加權指數成交量。

二、本文主要變數分析

將上述刪減過後參考變數加入本文自選變數重新做顯著測試(以顯著水準 10% 作為標準)，結果發現新聞指標經濟、新聞指標不確定、新聞指標政策、Google 搜尋趨勢、新聞性指標情緒皆為不顯著，為節省篇幅故省去圖示說明。

三、交乘項變數分析

變數之間關係探討：

此節為了研究及解決本文自編新聞性變數與總體經濟指標彼此在迴歸中因為彼此之間線性糾結而導致最終迴歸不顯著之問題，故探討一些變數之間潛在的關係以及影響相對大小與正負。在臺灣總體經濟中，確實不確定性指標亦或是政策經濟面指標往往可能會因為外在因素而產生非線性或互相加乘互相抵減之現象，如恐慌指數大幅變動可能間接影響臺灣大眾媒體對某方面討論熱度之影響，進而影響投資人判斷市場準則而影響臺灣加權指數走勢，下面會針對上述相關經濟洞見而進行實證結果與闡述。

(一) 恐慌指數與新聞性經濟指標的交乘項

由表 4-4 知當加入恐慌指數與新聞性經濟指標為交乘項時，原不顯著之新聞性經濟指標(未加入交乘項前)對模型反而有顯著貢獻，可由表中係數項可得知，當令其他變數不變以單純兩者關係來探討，可推得

$$\text{臺灣加權指數} = \text{常數} - 1.08 * \text{新聞性經濟指標} + 0.06 * \text{新聞性經濟指標} * \text{VIX} + \dots$$

平均而言，當恐慌指數大於 18⁹ 時，新聞性經濟指標會因為恐慌指數而對大盤股價呈現正向影響；反之當恐慌指數小於 18 時，新聞性經濟指標會因為恐慌

⁹ 當臺灣加權指數被新聞性不確定性指標與交乘項單位變動所改變量為 0 時，即可推得

$1.08/0.06 = 18 \left(\frac{\Delta y}{\Delta x_1} = a + bx_2, \text{當 } \frac{\Delta y}{\Delta x_1} = 0 \text{ 時, 可解得 } x_2 = \frac{-a}{b} \right)$

指數而對大盤股價呈現負向影響。與實際市場比對，恐慌指數高於一定程度時、當媒體面之經濟新聞指標係數上升也就是國內討論氣氛高漲時，確實欲可能推動股市流動性及增加避險需求進而正向影響股價推升。

（二）新聞性不確定性指標與恐慌指數的交乘項

由表 4-4 知當加入恐慌指數與新聞性不確定性指標為交乘項時，新聞性不確定性指標對臺灣加權指數為顯著，具有解釋力，由附表係數項可得知當令其他變數不變時單純以兩者關係來看時，可推得

$$\text{臺灣加權指數} = \text{常數項} + 6.39 * \text{新聞性不確定性指標} - 0.44 * \text{新聞性不確定性指標} * \text{VIX} + \dots$$

平均而言，當恐慌指數大於 14.52¹⁰時，不確定性指標因為恐慌指數交乘影響而對被解釋變數(大盤股價)由正向影響轉為負向影響，比對在 2019 中美貿易戰時，恐慌指數一度突破 20 大關，確實在當時影響民眾對未來股價報酬率預期，標普 500 指數也因此受民眾不確定性因素增加而下跌，由附表得知投資人受到不確定性的影響於恐慌指數約略 14.52 左右便會轉為負向影響，提前對於股票市場未來市場衝擊做理性預期，確實敏感之跡象。

（三）Google 搜尋趨勢與加權指數成交量的交乘項

由表 4-4 知當加入 Google 搜尋趨勢指標與加權指數成交量變數為交乘項時，Google 搜尋趨勢指標對被解釋變數為顯著，具有解釋力，由附表係數項可得知當令其他變數不變時單純以兩者關係來看時，可推得

$$\text{臺灣加權指數} = \text{常數項} + 1.99 * \text{Google 搜尋趨勢} - (3.9 * 10^{-7}) * \text{Google 搜尋趨勢} * \text{加權指數成交量} + \dots$$

平均而言，當加權指數成交量大於 5102654¹¹時，Google 搜尋趨勢因為加權指數成交量交乘影響而對臺灣加權指數由正向影響轉為負向影響，反之為正，實際上 Google 搜尋趨勢確實可能因為加權指數成交量突發性過高而產生極端預期的發生，而就附表而言當加權指數成交量飆高時，可能在一定程度以上民眾會加乘對未來股市可能發生動盪的情緒，進而負向影響大盤股價。

¹⁰ 當臺灣加權指數被新聞性不確定性指標與交乘項單位變動所改變量為 0 時，即可推得 $6.39/0.44=14.52$

¹¹ 當臺灣加權指數被 Google 搜尋趨勢與交乘項單位變動所改變量為 0 時，即可推得 $1.99/3.9*10=5102654$

（四）Google 搜尋趨勢與恐慌指數的交乘項

由表 4-4 知當加入 Google 搜尋趨勢指標與恐慌指數為交乘項時，Google 搜尋趨勢對被解釋變數為顯著，具有解釋力，由附表係數項可得知當令其他變數不變時單純以兩者關係來看時，可推得

$$\text{臺灣加權指數} = \text{常數項} + 1.28 * \text{Google 搜尋趨勢} - 0.07 * \text{Google 搜尋趨勢} * \text{VIX} + \dots$$

平均而言，當恐慌指數在 18.29¹² 以上時，Google 搜尋趨勢愈高，愈加負向影響大盤股價，反之為正向影響，網路民眾對於資訊的散播，從經濟角度來分析確實討論熱度往往為整體市場未來變動的濫觴，當市場越恐慌，網路討論度越高，市場上投資者往往只以名目變數之變動來做投資決策，導致可能未來對股價自然容易因為恐慌指數而有偏離理性之判斷。

（五）那斯達克綜合指數與新聞性政策指標的交乘項

由表 4-4 知當加入那斯達克綜合指數與新聞性政策指標為交乘項時，政策對臺灣加權指數為顯著，具有解釋力，由附表係數項可得知當令其他變數不變時單純以兩者關係來看時，可推得

$$\text{臺灣加權指數} = \text{常數項} - 0.35 * \text{新聞性政策指標} + (5 * 10^{-5}) * \text{那斯達克綜合指數} * \text{新聞性政策指標} + \dots$$

平均而言，當那斯達克綜合指數在 7000¹³ 以上時，那斯達克綜合指數越高，新聞性政策指標對臺灣加權指數有愈正向的影響，反之則為負面影響。實際市場來說，確實往往美國相關股票指數相對於臺灣經濟市場變動為領先指標，而我國之政策方也常為之影響，新聞性政策指標越高，換言之大眾媒體對政府對短期產出改善之政策有高度關注，確實從行為經濟學來說，短期確實能卓著改善產出，但大眾媒體的報導也可能影響民眾預期，而導致民眾提前反映預期進而導致產出無增加之結果。

¹² 當臺灣加權指數被 Google 搜尋趨勢與交乘項單位變動所改變量為 0 時，即可推得

$1.28/0.07 = 18.29$ ($\frac{\Delta y}{\Delta x_1} = a + bx_2$, 當 $\frac{\Delta y}{\Delta x_1} = 0$ 時，可解得 $x_2 = \frac{-a}{b}$)

¹³ 當臺灣加權指數被新聞性指標政策與交乘項單位變動所改變量為 0 時，即可推得 $0.35/5 * 10^{-5} = 7000$

表 4-4 交乘項迴歸結果¹⁴

變數	係數				
截距項	963.29**	1030.03**	889.83**	1016.68**	1079.02**
歐元兌美元匯率	188.11**	220.71**	214.12**	211.11**	220.81**
日圓兌美元匯率	1.17	1.30	1.11	1.46*	0.89
那斯達克綜合指數	0.07**	0.06**	0.06**	0.06**	0.06**
日本經濟平均指數	-0.02**	-0.02**	-0.02**	-0.02**	-0.02**
台幣兌美元匯率	-19.40**	-23.97**	-21.80**	-24.97**	-22.04**
大盤股價淨值比	116.10*	69.75	98.80*	64.67	103.78*
大盤本益比	-8.07**	-7.21*	-7.71**	-7.20*	-8.50**
加權指數成交量	0.00**	0.00*	0.00**	0.00**	0.00**
大盤前期價格	0.93**	0.94*	0.93**	0.94**	0.93**
恐慌指數	-5.91**	-2.95*	-3.87**	0.07	-4.10**
新聞性經濟指標	-1.08*	-0.05	-0.03	0.00	-0.08
新聞性不確定性指標	-2.19	6.39*	-1.94	-1.92	-2.05
新聞性政策指標	-0.03	-0.03	-0.04	-0.03	-0.35*
新聞情緒指標	0.01	0.00	0.00	0.00	0.00
Google 搜尋趨勢	0.00	0.05	1.99*	1.28**	0.04
交乘項 1 ¹⁵	0.06*	-	-	-	-
交乘項 2	-	-0.44**	-	-	-
交乘項 3	-	-	0.00*	-	-
交乘項 4	-	-	-	-0.07**	-
交乘項 5	-	-	-	-	0.00*
AIC	11.79005	11.78797	11.789	11.78579	11.79084
SC	11.8443	11.84222	11.84326	11.84004	11.84509

¹⁴ **與*分別代表掩住水準 5%及 10%顯著者

AIC(Akaike information criterion)與 SC(Schwarz criterion)，其主要功能為眾多模型選擇的準則愈低為優

¹⁵ 交乘項 1:新聞性經濟指標*恐慌指數
 交乘項 2: 恐慌指數*新聞性不確定性指標
 交乘項 3:Google 搜尋趨勢* 加權指數成交量
 交乘項 4:Google 搜尋趨勢*恐慌指數
 交乘項 5:那斯達克綜合指數*新聞性政策指標

四、門檻迴歸

選定變數：

本文試圖從本文選取之總體經濟變數中找尋在不同值域中可能會出現對於大盤股價或是本文所編制之新聞性變數間有無不同影響的變數，新聞性變數為本文自行編制的新聞性經濟指標，新聞性不確定性指標、新聞性政策指標，Google 搜尋趨勢以及新聞情緒指標。

（一）新聞性經濟指標

由表 4-5 發現，新聞性經濟指標與加權指數存在門檻值，為了解新聞性經濟指標對加權指數的影響，故運用門檻迴歸(Threshold regression)，以新聞性經濟指標作為門檻變數，進一步了解新聞性經濟指標與加權指數為何種關係，以及各解釋變數對加權指數的影響。

以門檻迴歸估計新聞性經濟指標與加權指數結果彙整如下表，透過門檻迴歸可知新聞性經濟指標門檻值為 25 與 33，當新聞性經濟指標小於 25 時為低度經濟相關媒體報導次數，於 25 與 33 之間為中度經濟相關媒體報導次數，大於 33 為高度經濟相關媒體報導次數，總共有 879 筆落入低度、438 筆於中度、388 筆落入高度。

不論一般迴歸模型亦或門檻迴歸，都須具備個別變數間缺乏自我相關的條件，若各變數間存在殘差自我相關，則會直接影響到整體模型之不偏性，為了避免估計錯誤，本文亦運用 Newey-west 於 1987 年提出的變異樹與自我相關的修正方法，以穩健原則穩定迴歸模型的充分一致有效性，以增加模型之可信度。

本文發現大盤本益比只於低度新聞性經濟指標時，才顯著影響臺灣加權指數，其餘皆不顯著，可能投資人於新聞指標低度時，並不會考慮大盤指標作為投資策略的指標的。而加權指數成交量為高度新聞性指標經濟時，才顯著影響臺灣加權指數。而值得一提的是新聞性政策指標於新聞性經濟指標於高度時對加權指數呈顯著影響，且為正相關。其餘情況皆對加權指數無顯著影響，而恐慌指數於經濟指標高、中、低度時皆為顯著影響。由實證結果可猜測臺灣新聞經濟指標其對臺灣加權指數可能有負向的加乘效果。其劇烈影響程度遠高於新聞性經濟指標低度與高度，亦即媒體討論熱度在一區間範圍內，可能恐慌指數對於加權指數影響具卓越貢獻，大部分解釋變數在新聞性經濟指標中度時對大盤呈現不顯著影響。可能原因為當國內經濟面向屬於穩定階層時，變數變動可能對民眾所期盼未來加權

指數影響較為溫和，而經濟面向為極端階層時，變數微小變動可能大幅度地影響民眾與政府在干預臺灣經濟決策時所作影響的範圍。

表 4-5 門檻迴歸結果(新聞性經濟指標)¹⁶

變數	新聞性經濟指標 <25	25≤新聞性經濟指標 <33	33≤新聞性經濟指標
截距項	771.18**	752.72	2021.13**
歐元兌美元匯率	173.34*	233.50	-169.98
日圓兌美元匯率	1.20	-0.94	4.76
那斯達克綜合指數	0.08**	0.08**	0.08**
日本經濟平均指數	-0.02**	-0.01	-0.01
台幣兌美元匯率	-13.77*	-8.20	-41.65*
大盤股價淨值比	234.63**	211.64	-358.77
大盤本益比	-19.51**	-1.28	11.10
加權指數成交量	0.00	0.00	0.00**
恐慌指數	-5.91**	-5.57**	-5.34**
新聞性經濟指標	-1.25*	-1.90	0.31
新聞性不確定性指標	-3.98	-1.85	-6.01
新聞性政策指標	-0.07	-0.21	0.44*
新聞情緒指標	-0.00	0.02	-0.02
Google 搜尋趨勢	-0.25	0.75*	-0.27
大盤前一期價格	0.95**	0.89**	0.89**
樣本數	879	438	388
AIC	11.75847		
SC	11.91166		

¹⁶ **與*分別代表掩住水準 5%及 10%顯著者

C(Akaike information criterion)與 SC(Schwarz criterion)，其主要功能為眾多模型選擇的準則愈為優

（二）新聞性不確定性指標

本文試圖以不同層面來觀察來衡量加權指數受影響範疇的大小，由門檻迴歸實證研究中發現，以新聞性不確定性指標作為加權指數估計並不存在結構性轉折點。

（三）新聞性政策指標

由表 4-6 門檻迴歸發現，新聞性經濟指標與加權指數存在門檻值。以門檻迴歸估計新聞性政策指標與加權指數結果彙整如下表，透過門檻迴歸可知新聞性政策指標門檻值為 74、92 和 154，當新聞性政策指標小於 74 時為極低度政策相關媒體報導指標，於 74 與 92 之間為低度政策相關媒體報導指標，92 與 154 之間為中度政策相關媒體報導指標、大於 154 為高度政策相關媒體報導指標，總共有 315 筆落入極低度、329 筆於低度、770 筆落入中度、291 筆落入高度。

由表 4-6 發現，恐慌指數於四個轉折點皆為顯著影響臺灣加權指數，可看出臺灣投資人可能對於國際間恐慌較為敏感，在投資之間傾向以市場風險作為投資的主要考量點。大盤本益比除了中度以外皆為顯著影響臺灣加權指數，故於大部分政策討論熱度時，皆有可能影響投資人投資情緒之高低，而討論熱度為穩定(中度)時，影響較為微小。而那斯達克綜合指數於四個階段皆為顯著影響臺灣加權指數。而本文發現，當新聞性政策指標落入極低度時，除了新聞性經濟指標、新聞性政策指標外，其餘本文編制的變數(新聞性不確定性指標、Google 搜尋趨勢、新聞情緒指標)皆對加權指數無顯著影響，而新聞性政策指標於極低度以外，皆都為不顯著，故新聞性政策指標高低長期下來(資料為 2014-2020 年)對於加權指數整體性影響有較弱的可能性，換言之政策改變帶來的影響高低似乎在中長期(2014 至 2020)對於整體金融市場產出影響較為溫和，Muth(1961)所提出理性預期理論中也曾提過，若整體市場皆為理性預期下，的確有可能大幅度地減少系統性偏誤，在政府介入市場以財政政策或貨幣政策時，民眾往往能事先調整預期進而減少政策帶來的產出改變。而新聞性政策指標，為刻畫一段時間內政府相關政策於媒體討論之熱度，是故，當新聞性政策指標為低度時，可理解為媒體報導相關政府施政政策關鍵字頻率較低，而高度則為頻率較高。

表 4-6 門檻迴歸結果(新聞性政策指標)¹⁷

變數	新聞性指 標_政策 <74	74<=新聞 性指標_政 策<92	92<= 新聞 性指標_政 策<154	154<= 新 聞性指標_ 政策
截距項	1188.42*	2053.38**	1183.82**	1971.06
歐元兌美元匯率	362.61*	252.45	187.80	-352.48*
日圓兌美元匯率	0.70	-1.36	1.38	4.39
那斯達克綜合指數	0.11**	0.09**	0.06**	0.10**
日本經濟平均指數	-0.04**	-0.04**	-0.02**	0.01
台幣兌美元匯率	-22.73	-29.05*	-27.97**	-30.10
大盤股價淨值比	478.14**	211.83	41.47	-396.17
大盤本益比	-33.87**	-42.16**	-3.95	36.90**
加權指數成交量	0.00	0.00*	0.00	0.00
恐慌指數	-6.62**	-6.41*	-4.58**	-4.77**
新聞性經濟指標	-1.33*	-0.44	0.09	-0.59
新聞性不確定性指標	-2.79	2.10	-1.71	-13.21*
新聞性政策指標	-1.41**	0.73	0.29	-0.02
新聞情緒指標	0.01	-0.02	0.00	0.01
Google 搜尋趨勢	-0.58*	-0.39	0.51*	-0.92
大盤前期價格	0.91**	0.93**	0.93**	0.80**
樣本數	315	329	770	291
AIC	11.77			
SC	11.98			

¹⁷ **與*分別代表掩住水準 5%及 10%顯著者IC(Akaike information criterion)與 SC(Schwarz criterion)，其主要功能為眾多模型選擇的準則
低為優

（四）Google 搜尋趨勢

由表 4-7 門檻迴歸發現，於模型中 Google 搜尋趨勢與加權指數存在門檻值。以門檻迴歸估計 Google 搜尋趨勢與加權指數結果彙整如下表，透過門檻迴歸可知 Google 搜尋趨勢門檻值為 2 段，門檻值為 54.55789。當 Google 搜尋趨勢小於 54.55789 時為低度網路討論熱度，大於 54.55789 為高度網路討論熱度，總共有 1288 筆落入低度、417 筆於高度。

本文發現那斯達克綜合指數、台幣對美元匯率、恐慌指數於兩轉折點皆為顯著影響臺灣加權指數，故可能不管網路討論熱度為何，往往投資人皆會納入這些變數作為進場買股票的時機點。且當 Google 搜尋趨勢由低度邁入高度時，對加權指數顯著性有明顯改善，當網路搜尋熱度為高度區間時，為最顯著影響臺灣加權指數，由上述實證結果可猜測，當 Google 搜尋趨勢飆升時，可能影響加權指數的力量有上升的趨勢，而本次文所編列變數(新聞性經濟指標，新聞性不確定性指標、新聞性政策指標)也於網路討論度高度影響加權指數時對加權指數為顯著影響。故猜測 Google 搜尋趨勢有可能是間接影響新聞性政策指標，新聞性經濟指標，新聞性不確定性指標強弱的催化劑。

（五）新聞情緒指標

本文試圖以不同層面來觀察來衡量加權指數受影響範疇的大小，由線性迴歸實證研究中發現，以新聞情緒指標作為加權指數估計並不存在任何顯著影響及交互作用影響，因此無法藉由觀察新聞情緒指標變動是否會影響加權指數。然而，雖實證結果發現新聞情緒指標對臺灣加權指數漲跌無任何顯著性影響，但本文仍不排除新聞情緒指標在非線性影響中可能有顯著影響。確實在近代經濟發展中，行為財務經濟學漸為經濟學中之顯學，試圖在經濟學理中加諸心理學以刻畫不同投資人情緒影響投資之規律性，是故，本文認為，於中長期時間序中，新聞情緒指標對於臺灣加權指數無顯著影響，但於短期有可能因為股票市場無效率之影響以及影響變數相對較少的條件下擴大對於投資人情緒反應之力度。

表 4-7 門檻迴歸結果(Google 搜尋趨勢)¹⁸

變數	Google 搜尋趨勢<54.56	54.56<=Google 搜尋趨勢
截距項	878.30**	1624.62**
歐元兌美元匯率	102.14*	177.58
日圓兌美元匯率	0.98	2.47
那斯達克綜合指數	0.05**	0.09**
日本經濟平均指數	-0.02**	-0.01
台幣兌美元匯率	-19.19**	-39.82**
大盤股價淨值比	72.90	114.00
大盤本益比	-11.17**	0.65
加權指數成交量	0.00**	0.00
恐慌指數	-2.80**	-5.45**
新聞性經濟指標	-0.62*	0.86*
新聞性不確定性指標	0.90	-10.43*
新聞性政策指標	0.12	-0.25*
新聞情緒指標	0.00	0.00
Google 搜尋趨勢	0.09	0.97*
大盤前一期價格	0.96**	0.87**
樣本數	1288	417
AIC	11.77	
SC	11.87	

五、預測結果

本文以匯率、國際指數、利率、大盤股價淨值比、大盤本益比、大盤成交量、利率、VIX 波動率指數等預測股價趨勢的變數做為參考變數，刪減不顯著之變數後再逐一加入本文新聞性變數與臺灣加權指數前一期合稱為基本變數，並以前述研究交乘項為測試變數，找尋最適預測力模型。

¹⁸ **與*分別代表掩住水準 5%及 10%顯著者

IC(Akaike information criterion)與 SC(Schwarz criterion)，其主要功能為眾多模型選擇的準則，低為優。

表 4- 8 模型使用變數表

模型名稱	使用變數
模型 1	基本變數(人民幣兌美元匯率、歐元兌美元匯率、英鎊兌美元匯率、日圓兌美元匯率、台幣兌美元匯率、台股兌美元匯率、那斯達克綜合指數、大盤股價淨值比、加權指數的成交量、蘋果公司股價、大盤本益比、標普 500 指數、上海綜合指數、道瓊工業指數、日本經濟平均指數、恐慌指數、臺銀一年期定存利率)
模型 2	基本變數、新聞情緒指標交乘項(經濟*恐慌指數)
模型 3	基本變數、Google 搜尋趨勢*恐慌指數
模型 4	基本變數、新聞情緒指標交乘項(不確定*恐慌指數)
模型 5	基本變數、模型 1~4 之變數

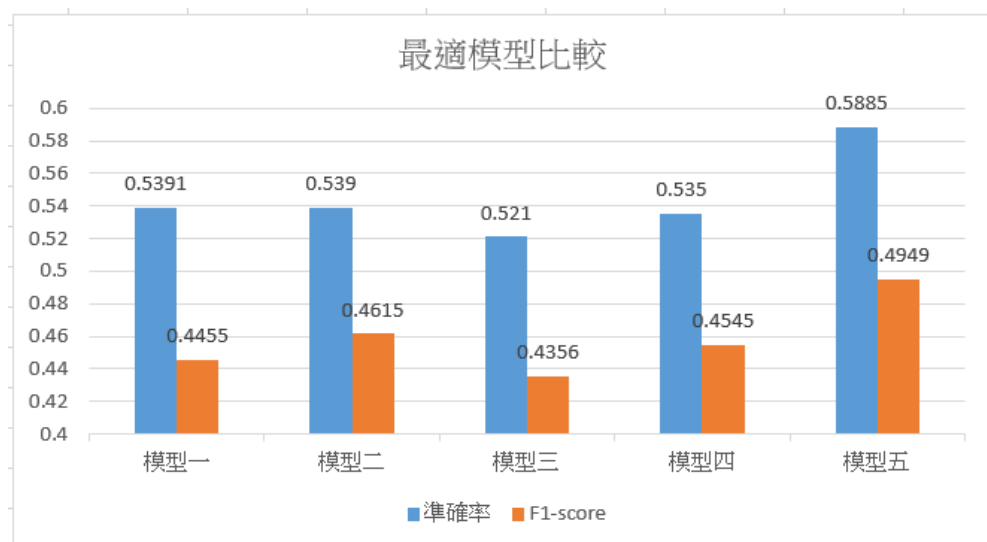


圖 4- 4 各模型績效比較圖

如圖 4-4 可知使用多變量迴歸預測準確率為由模型一至五為 53.91%、55.9%、52.1%、53.5%、58.84%。著實能由上述實證結果觀察得知模型五(準確率 58.85%、F1-score 49.49%)為加入模型一至四交乘項的結果，確實對預測績效有卓著提升且皆較其他四個單獨模型優越。故本文推測結論為新聞指標關鍵字會影響臺灣加權指數漲跌，但新聞性政策指標在上述模型中，因顯著性不足而並未採用，並不代表短期政策媒體討論熱度與臺灣加權指數為獨立關係，由實證結果僅能確定兩者在樣本期間內，不具有線性關係。

第四節 機器學習結果

一、機器學習變數說明

本文以匯率、國際指數、利率、大盤股價淨值比、大盤本益比、大盤成交量利率、VIX 波動率指數等預測股價趨勢的變數做為參考變數，再逐一加入本文新聞性變數，透過決策樹、隨機森林和 XG-Boost 預測股價漲跌方向，並以準確率 (accuracy) 和 F1-score 為評判標準，判斷本文加入之新聞性變數可否提升預測股價漲跌方向能力。

二、模型預測結果

透過本文第二章第一節總體經濟指標對股價關聯之相關文獻，本文將較常被拿來運用的經濟變數，作為使用變數做成經濟變數模型，以此模型作為後面比較的基礎，去探討本文的新增變數是否能夠提升模型的預測能力，另外三個模型為新聞情緒模型、Google 搜尋趨勢模型、新聞性模型，以基本模型為基礎，分別加入本文新增變數新聞情緒指標、Google 搜尋趨勢、新聞性指標，去分別比較各個本文新增變數對於基本模型的影響。

表 4-9 模型使用變數表

模型名稱	使用變數
經濟變數模型	經濟變數(人民幣兌美元匯率、歐元兌美元匯率、英鎊兌美元匯率、日圓兌美元匯率、台幣兌美元匯率、台幣兌美元匯率、那斯達克綜合指數、大盤股價淨值比、加權指數的成交量、蘋果公司股價、大盤本益比、標普 500 指數、上海綜合指數、道瓊工業指數、日本經濟平均指數、恐慌指數、臺銀一年期定存利率)
新聞情緒模型	經濟變數、新聞情緒指標
Google 搜尋趨勢模型	經濟變數、Google 搜尋趨勢
新聞性模型	經濟變數、新聞性指標

(一) 經濟變數模型

表 4-10 在驗證期，以隨機森林的模型表現最佳 F1-score 和準確率皆為最高，在測試期隨機森林的準確率最高，F1-score 略小於決策樹，以整體而言隨機森林為最佳的模型。

表 4-10 經濟變數模型績效表

		決策樹	隨機森林	XG-Boost
驗證期	準確率	0.5239	0.5695	0.5460
	F1-score	0.6220	0.6845	0.6207
	混淆矩陣	[[58 144]	[[45 157]	[[[79 123]
		[65 172]]	[32 205]]	[75 162]]
測試期	準確率	0.5164	0.5328	0.5328
	F1-score	0.6266	0.62	0.5476
	混淆矩陣	[[27 77]	[[37 67]	[[61 43]
		[41 99]]	[47 93]]	[71 69]]

王鼎宏(2016)使用機器學習方法預測加權指數之研究，也使用此經濟變數預測加權股價漲跌方向，標普 500 指數、那斯達克指數、道瓊工業指數、臺幣兌美元匯率、VIX 指數為隨機森林模型中權重前五高的變數。在表 4-11 中決策樹、隨機森林、XG-boost 三個模型中加權指數成交量、道瓊工業指數、人民幣兌美元匯率、臺幣兌美元匯率皆占相當高的權重。隨機森林模型中加權指數的成交量、道瓊工業指數、臺幣兌美元匯率、人民幣兌美元匯率、上海綜合指數權重前五高的變數。

表 4- 11 經濟變數模型特徵比重表

	決策樹	隨機森林	XG-Boost
加權指數成交量	0.1046 [1]	0.1632 [1]	0.0638
道瓊工業指數	0.08416 [2]	0.1550 [2]	0.0713[1]
恐慌指數	0.08240 [3]	0.03608	0.0624
人民幣兌美元匯率	0.08108 [4]	0.1393 [4]	0.0685[2]
大盤本益比	0.07828 [5]	0.03393	0.0631
歐元兌美元匯率	0.07294	0.10898	0.0605
上海綜合指數	0.07110	0.11069 [5]	0.0616
標普 500 指數	0.06136	0	0.0656[3]
英鎊對美元匯率	0.05770	0	0.05896
大盤股價淨值比	0.05765	0	0.06152
日圓兌美元匯率	0.05512	0	0.05997
臺幣兌美元匯率	0.05331	0.1532 [3]	0.06390[5]
那斯達克綜合指數	0.04971	0.02967	0.06178
蘋果公司股價	0.04393	0.06988	0.06441[4]
日本經濟平均指數	0.04298	0	0.05853
臺銀一年定存利率	0.00371	0	0.05417

(此表為使用機器學習建立預測模型時，各特徵值於該模型所占比之權重)

（二）新聞情緒模型

表 4-12 使用參考變數加上新聞情緒指標當作自變數，在驗證期，以隨機森林的模型表現最佳，F1-score 和準確率皆為最高，在測試期，決策樹的 F1-score 和準確率皆為最高，因此決策樹為此組變數的最佳模型。

表 4-12 新聞情緒模型績效表

		決策樹	隨機森林	XG-Boost
驗證期	準確率	0.5467	0.5831	0.5421
	F1-score	0.6059	0.6945	0.6257
	混淆矩陣	[[87 115] [84 153]]	[[48 154] [29 208]]	[[70 132] [69 168]]
測試期	準確率	0.5738	0.5574	0.5
	F1-score	0.6994	0.6897	0.4831
	混淆矩陣	[[19 85] [19 121]]	[[16 88] [20 120]]	[[65 39] [83 57]]

在表 4-13 中，決策樹、隨機森林、XG-boost 三個模型中加權指數成交量、道瓊工業指數、歐元兌美元匯率、人民幣兌美元匯率皆占相當高的權重，決策樹模型中加權指數的成交量、歐元兌美元匯率、上海綜合指數、道瓊工業指數、臺幣兌美元匯率權重前五高的變數，新聞情緒指標只佔了 3.83%，遠低於其他變數。因此可以發現新聞情緒指標雖然佔比權重偏低，但能夠影響模型的演算，改變其他變數的權重，間接提升整體準確率和 F1-score，但是會有大部分預測結果為上漲的現象，將股價下跌誤預測為股價上漲，造成混淆矩陣中偽陽性(False Postive)遠大於偽陰性(False Negative)，此組變數在 XG-Boost 中不會有此現象，但是準確率和 F1-score 偏低。

表 4- 13 新聞情緒模型特徵比重表

	決策樹	隨機森林	XG-Boost
新聞情緒指標	0.0383	0.0629	0.0578
加權指數成交量	0.1358 [1]	0.0814 [2]	0.0606
道瓊工業指數	0.1025 [4]	0.0853 [1]	0.0832 [1]
恐慌指數	0.0508	0.0682	0.0524
人民幣兌美元匯率	0.0921	0.0719	0.0610 [5]
大盤本益比	0.0934	0.0733 [4]	0.0595
歐元兌美元匯率	0.1202 [2]	0.0758 [3]	0.0562
上海綜合指數	0.1159 [3]	0.0629	0.0583
標普 500 指數	0	0.0726 [5]	0.0614 [4]
英鎊對美元匯率	0	0.0711	0.0591
大盤股價淨值比	0	0.0383	0.0634 [2]
日圓兌美元匯率	0.0372	0.0475	0.0546
臺幣兌美元匯率	0.1013 [5]	0.0522	0.0542
那斯達克綜合指數	0.0665	0.0428	0.0511
蘋果公司股價	0.0462	0.0420	0.0623 [3]
日本經濟平均指數	0	0.0407	0.0554
臺銀一年定存利率	0	0.0025	0.0492

（三）Google 搜尋趨勢模型

表 4-14 使用參考變數加上 Google 搜尋趨勢當作自變數，在驗證期期間，以 XG-Boost 的模型表現最佳，F1-score 和準確率皆為最高，來到測試期期間，隨機森林的 F1-score 和準確率皆為最高，因此隨機森林為此組變數的最佳模型。

表 4- 14 Google 搜尋趨勢模型績效表

		決策樹	隨機森林	XG-Boost
驗證期	準確率	0.4943	0.5330	0.5467
	F1-score	0.5432	0.5988	0.6224
	混淆矩陣	[[85 117] [105 132]]	[[81 121] [84 153]]	[[76 128] [71 164]]
測試期	準確率	0.5040	0.5615	0.5246
	F1-score	0.5568	0.6581	0.5797
	混淆矩陣	[[47 57] [64 76]]	[[34 70] [37 103]]	[[48 56] [60 80]]

在表 4-15 中，決策樹、隨機森林、XG-boost 三個模型中，Google 搜尋趨勢、加權指數成交量、人民幣兌美元匯率皆占相當高的權重，決策樹模型中 Google 搜尋趨勢、加權指數成交量、人民幣兌美元匯率、臺幣兌美元匯率、道瓊工業指數權重前五高的變數，Google 搜尋趨勢為權重最高，在加入後影響演算法改變其他權重的重要性，提升整體準確率和 F1-score，可推斷 Google 搜尋趨勢對此模型具有參考價值。

表 4- 15 Google 搜尋趨勢模型特徵比重表

	決策樹	隨機森林	XG-Boost
Google 搜尋趨勢	0.1767 [1]	0.0905 [1]	0.0543
加權指數成交量	0.1137 [2]	0.0777 [2]	0.0607
道瓊工業指數	0.0699	0.0675 [5]	0.0545
恐慌指數	0	0.0666	0.0637[2]
人民幣兌美元匯率	0.0924 [5]	0.0702 [3]	0.0596
大盤本益比	0.1050 [4]	0.0637	0.0586
歐元兌美元匯率	0.0816	0.0605	0.0526
上海綜合指數	0.0346	0.0611	0.0698[1]
標普 500 指數	0.0236	0.0571	0.0610[5]
英鎊對美元匯率	0.1054 [3]	0.0642	0.0545
大盤股價淨值比	0.0155	0.0388	0.0621[3]
日圓兌美元匯率	0	0.0517	0.0582
臺幣兌美元匯率	0.0512	0.0681 [4]	0.0591
那斯達克綜合指數	0.0393	0.0501	0.0498
蘋果公司股價	0.0674	0.0528	0.0591[4]
日本經濟平均指數	0.0235	0.0571	0.0593
臺銀一年定存利率	0	0.0020	0.0607

(四) 新聞性模型

表 4-16 中，使用參考變數加上新聞性指標當作自變數，驗證期，以隨機森林的模型表現最佳，F1-score 和準確率皆為最高，測試期，隨機森林的 F1-score 和準確率也皆為最高，但發現預測結果嚴重偏向上漲，所以認為決策樹是最佳的模型。

表 4- 16 新聞性模型表

		決策樹	隨機森林	XG-Boost
驗證期	準確率	0.5239	0.5808	0.546
	F1-score	0.6122	0.6964	0.6440
	混淆矩陣	[[65 137] [72 165]]	[[44 158] [26 211]]	[[60 142] [57 180]]
測試期	準確率	0.5820	0.5902	0.5451
	F1-score	0.6710	0.7340	0.6159
	混淆矩陣	[[38 66] [36 104]]	[[6 98] [2 138]]	[[44 60] [51 89]]

在表 4-17 中，決策樹、隨機森林、XG-boost 三個模型中加權指數成交量、道瓊工業指數，皆占相當高的權重，決策樹模型中加權指數成交量、新聞性政策指標、人民幣兌美元匯率、新聞性經濟指標、蘋果公司股價權重前五高的變數，可以發現新聞性政策指標和新聞性經濟指標是具有一定的參考價值，在加入後影響演算法改變其他權重的重要性，提升整體準確率和 F1-score。而新聞性不確定性指標在此模型權重為 0，對此模型毫無參考價值。

表 4- 17 新聞性模型特徵比重表

	決策樹	隨機森林	XG-Boost
新聞性經濟指標	0.0870 [4]	0.0367	0.0491
新聞性不確定性指標	0.0000	0.0152	0.0465
新聞性政策指標	0.1185 [2]	0.0598	0.0528
加權指數成交量	0.1579 [1]	0.0798 [2]	0.0525
道瓊工業指數	0.0758	0.0686 [5]	0.0658[1]
恐慌指數	0.0678	0.0866 [1]	0.0527
人民幣兌美元匯率	0.0948 [3]	0.0559	0.0506
大盤本益比	0.0299	0.0675	0.0566[4]
歐元兌美元匯率	0.0461	0.0346	0.0525
上海綜合指數	0.0616	0.0765 [4]	0.0580[3]
標普 500 指數	0.0331	0.0795 [3]	0.0549[5]
英鎊對美元匯率	0.000000	0.0357	0.0483
大盤股價淨值比	0.0273	0.0481	0.0604[2]
日圓兌美元匯率	0.0104	0.0532	0.0473
臺幣兌美元匯率	0.0668	0.0553	0.0520
那斯達克綜合指數	0.0010	0.0515	0.0506
蘋果公司股價	0.0848 [5]	0.0496	0.0542
日本經濟平均指數	0.0374	0.0401	0.0504
臺銀一年定存利率	0.0000	0.0058	0.0445

（五）最適模型比較

由圖 4-5 可知，加入本文新增變數的模型績效皆優於經濟變數模型，證明本文新增變數對於預測股價漲跌方向是有貢獻的，從 F1_score 比較，新聞情緒模型預測績效最佳。情緒指標在新聞情緒模型中權重占比不高，但是會影響其他變數權重占比，Google 搜尋趨勢在 Google 搜尋趨勢模型和全體變數模型中權重占比都極為重要，但對於模型績效的提升不如新聞情緒模型，由此結果可知即使占比權重高對於模型績效亦無較大幫助。由於將所有變數加入模型形成的全體變數模型並無明顯績效提升，因此本文省略其內容，將結果放置在附錄。

表 4- 18 各模型績效表

	經濟變數模型	新聞情緒模型	Google 搜尋趨勢模型	新聞性模型
機器學習模型	隨機森林	決策樹	隨機森林	決策樹
準確率	0.5368	0.5738	0.5615	0.582
F1-score	0.62	0.6994	0.6581	0.671

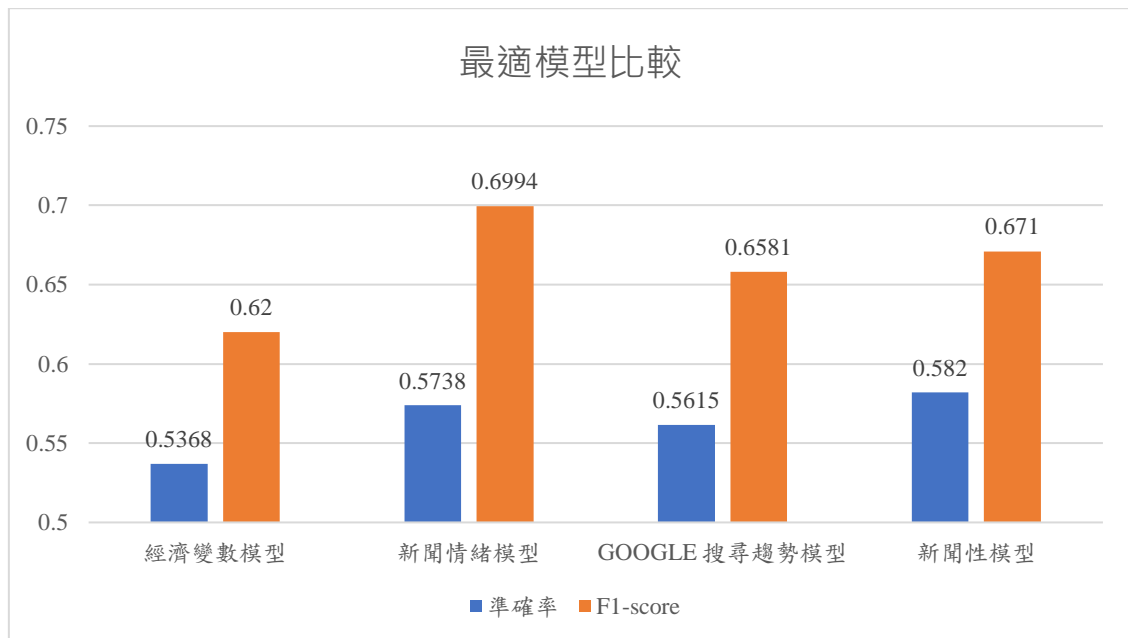


圖 4- 5 各模型績效比較圖

圖 4-6 在股價漲跌預測正確的日數將臺灣加權指數日報酬率以藍線畫出，預測錯誤則以橘色線畫出。兩圖比較可明顯看出，在臺灣加權指數波動大時，新聞情緒模型，預測漲跌方向正確力遠大於經濟變數模型，因此可看出此時新聞情緒變數扮演重要的角色，但是在臺灣加權指數波動小時卻較無明顯作用。

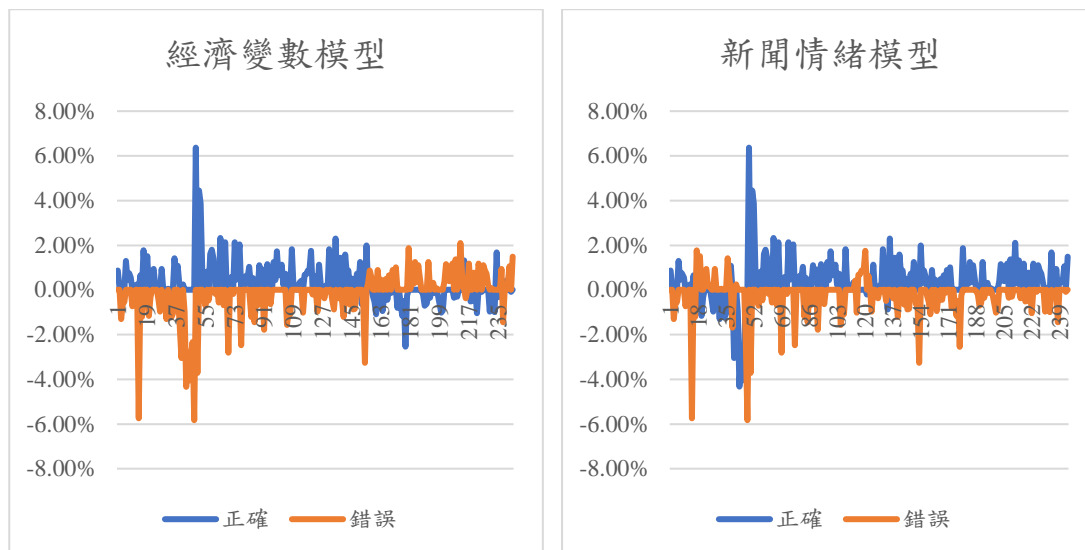


圖 4-6 模型預測比較圖

第五章 結論與建議

第一節 結論

實證結果發現，本文新增的新聞情緒指標、Google 搜尋趨勢、新聞性經濟指標、新聞性不確定性指標、新聞性政策指標跟臺灣加權指數都不具有線性關係，可能原因為現今金融市場越來越複雜，或投資人在新聞熱度不在一定高點以上時，對新聞無感，甚於不同時點、不同經濟環境下，經濟變數之解釋能力也會隨之改變，是故單以線性迴歸模型作為建模基礎著實手不應心。因此，本文認為本文新增變數和臺灣加權指數可能為非線性關係，因此採取兩類型的方式去處理。第一個類型是非線性迴歸模型包含：兩個解釋變數的交乘項模型，可以掌握特定經濟變數對於本文新增變數是否有加乘或抵減之現象。實證結果發現恐慌指數作為交乘項時能顯著改善本文新增變數顯著影響臺灣加權指數之良窳，而透過門檻迴歸模型，發現新聞性經濟指標、新聞性政策指標以及 Google 搜尋趨勢皆存在門檻值，代表這三項變數在不同期間對於加權指數的影響有所不同。第二個類型是機器學習的模型包含決策樹、隨機森林、XGBoost 三種模型，XGBoost 有內至交叉驗證的功能和每建構一個決策樹就會修正以前訓練過的決策樹犯下的錯誤的特色，因此本文原先認為此模型預測績效為最佳，但是研究結果發現，XGBoost 皆非最佳模型，反而是決策樹和隨機森林有比較好的表現，推測因為變數間相互關係較小，XGBoost 模型修正錯誤的能力有限導致預測績效較差。

另外，本文實證結果亦發現在經濟變數加入任一個本文新增變數皆可提高模型預測能力，其中又以使用決策樹方法的新聞情緒模型有最佳的 F1-score，比起經濟變數模型提高 7.94%。若將每日預測結果做對比，可以發現當臺灣加權指數當天有波動 3% 以上時，經濟變數模型準確率為 30% 但是新聞情緒模型高達 60%，準確率提升兩倍。因此，發現當臺灣加權指數出現劇烈波動時，加入新聞性情緒變數，可以有效提高模型預測臺灣加權指數漲跌方向預測能力。

第二節 後續研究建議

本文在收集資料的過程中，遇到許多限制，如有些關鍵字在運用 Google 搜尋趨勢時由於資料時間過短所以不符合我們預設樣本期間，或者有些關鍵字無法爬取，因此將部分關鍵字刪除，導致本文新增變數沒有當初設計時來的完善。在爬取新聞的過程中，部分新聞網站，在大量爬取資料過程中，會阻擋使用者爬取，造成無法爬取該新聞網站的資料。若後續 Google 搜尋趨勢有更完整的資料

或是可以突破新聞網站的爬取限制，或許可以得到更多新聞資料，提升新聞性指標的有效性，進而提高模型績效。

參考文獻

一、中文部分

- 王鼎宏(2016)，「使用機器學習方法預測加權指數之研究」。國立成功大學經營管理系碩士論文。
- 江秉倫(2020)，「納入情緒指標及資金管理系統之交易策略」。國立中正大學財務金融系碩士論文。
- 吳怡慧(2011)，「臺灣加權股價指數與總體經濟變數之關聯性研究」。國立中正大學經濟系碩士論文。
- 李嘉洲(2016)，「應用深度學習於財經新聞來源對股價趨勢預測之研究」。淡江大學資訊管理學系碩士論文。
- 翁小蘅(2009)，「新臺幣匯率、利率與股價報酬率關聯性之研究」。國立臺北大學國際財務金融系碩士論文。
- 陳識安(2020)，「氣候變遷的新聞情緒指標之投資組合績效-以美股為例」。天主教輔仁大學金融與國際企業學系碩士論文。
- 張映文(2018)，「臺灣的通貨膨脹與經濟成長—以門檻迴歸模型分析」。國立政治大學行政管理碩士學程論文。
- 陳振霖(2016)，「資產定價和政策不確定性」。國立東華大學財務金融學系碩士論文。
- 莊家睿(2017)，「三大法人資訊對於大盤指數的解釋力及期貨市場的報酬率-以臺灣加權指數為例」。國立交通大學資訊管理與財務金融學系碩士論文。
- 陳建州(2005)，「利用總體指標進行台股投資的擇時策略」。天主教輔仁大學金融研究所在職專班碩士論文。
- 黃裕烈，葉錦徽，陳重吉(2020)，「臺灣經濟政策不確定性指標之建構與分析」。
- 詹佑維(2019)，「投資人情緒再剖析以肯定及懷疑情緒預測臺灣加權指數及加權指數期貨」，國立雲林科技大學財務金融系碩士論文。
- 劉靜玟(2018)，「經濟政策不確定性對中國股市之影響」。大葉大學財務金融系碩士論文。
- 魏楚蕎(2015)，「應用資料探勘技術及總體經濟變數以預測臺灣加權指數」。朝陽科技大學財務金融系碩士論文。
- 魏宏泰(2003)，「臺灣股價與總體經濟變數關係之實證研究」。朝陽科技大學財務金融系碩士論文。

羅欣茹(2020)，「貨幣超額報酬與經濟政策不確定性和流動性收益之關聯性」。
國立中央大學財務金融學系碩士論文。

二、英文部分

- Bernanke Ben S.(1980)，"Irreversibility, Uncertainty, and Cyclical investment "，
Working paper No.502.
- Bloom Nicholas(2007)，"The impact of uncertainty shocks"，Nber working paper
series.
- Baker Malcolm, Wurgler Jeffery(2009)，"Investor Sentiment and the Cross-Section of
Stock Returns "，The journal of finance.
- Baker R.Scott, Bloom Nicholas, Davis J. Steven(2016)，"Measuring economic policy
uncertainty "，The quarterly journal of economics.
- Choi Hyunyoung, Varian Hal (2011)，"Predicting the Present with Google Trends ".
- Christiano J. Lawrence, Motto Roberto, Rostagno Massimo(2014)，"Risk Shocks "，
American Economic Review.
- Christou Christina, Gupta Rangan, Hassapis Christis, Suleman Tahir (2017)，"The role
of economic uncertainty in forecasting exchange rate returns and realized
volatility: Evidence from quantile predictive regressions ".
- Gu Ming, Sun Minxing, Wu Yangru, Xu Weike(2020)，"Economic Policy Uncertainty
and Momentum ".
- Huang Yun, Luk Paul(2018)，"Measuring Economic Policy Uncertainty in China ".
- Shiller J. Robert(2002)，"Bubbles, Human Judgment, and Expert Opinion "，Financial
Analysts Journal.
- Schaal Edouard(2015)，"Uncertainty and Unemployment "，New York University.
- Xavier C. Gustavo, Vasconcelos N. C. Lucas(2018)，"Economic Policy Uncertainty
and Momentum: International Evidence "，SSRN Electronic Journal.

附錄

全體變數模型績效				
		決策樹	隨機森林	XG-Boost
驗證期	準確率	0.5193	0.5672	0.5262
	F1-score	0.5720	0.6885	0.5840
	混淆矩陣	[[87 115] [96 141]]	[[39 163] [27 210]]	[85 117] [91 146]]
測試期	準確率	0.4795	0.5451	0.5451
	F1-score	0.4774	0.6687	0.6687
	混淆矩陣	[[59 45] [82 58]]	[[21 83] [28 112]]	[[21 83] [28 112]]

全體變數模型特徵比重			
	決策樹	隨機森林	XG-Boost
新聞性經濟指標	0	0.024	0.0476
新聞性不確定性指標	0	0.0121	0.0506 [4]
新聞性政策指標	0.0771	0.0475	0.0484
新聞情緒指標	0.3388 [1]	0.0478	0.0519
Google 搜尋趨勢	0.1066 [5]	0.1042 [1]	0.0561 [1]
加權指數成交量	0.1500 [2]	0.0749 [2]	0.0553 [2]
道瓊工業指數	0.1113 [4]	0.0602 [4]	0.0549 [3]
恐慌指數	0	0.0638 [3]	0.0477
人民幣兌美元匯率	0.1340 [3]	0.0540	0.0502
大盤本益比	0	0.0544	0.0504 [5]
歐元兌美元匯率	0	0.0542	0.0443
上海綜合指數	0.0288	0.0518	0.0442
標普 500 指數	0	0.0581 [5]	0.0459
英鎊對美元匯率	0.0535	0.0431	0.0577
大盤股價淨值比	0	0.0388	0.0468
日圓兌美元匯率	0	0.0435	0.0342
臺幣兌美元匯率	0	0.0428	0.0498
那斯達克綜合指數	0	0.0415	0.0436
蘋果公司股價	0	0.0402	0.0445
日本經濟平均指數	0	0.0392	0.0432
臺銀一年定存利率	0	0.0040	0.0327