

國立政治大學國際經營與貿易研究所

碩士學位論文

以情緒分析量化新聞資料探討新台幣匯率市場
效率性之研究

The Efficiency of USD/TW FX Market

-----Quantifying News Materials with Sentiment Analysis



指導教授：林信助 博士

研究生：張邑齊 撰

中華民國 108 年 07 月

致謝

隨著論文的完成，2年的碩士生涯也到了尾聲，在論文撰寫的過程中，最感謝的人莫過於指導教授林信助老師。感謝老師願意讓我在繁忙的論文撰寫過程中同時進行暑期實習以及出國交換，讓我有機會在即使出國交換的情況下也能準時畢業。感謝老師在指導過程中不辭辛勞一字一句地協助我修改論文，即使不在上班時間，老師也都會很願意回覆我的訊息，每當研究遇到瓶頸，老師也都很明確地給我方向與相關資料，讓我在論文的撰寫過程中逐漸增進了我研究與解決問題的能力。跟著老師的這段期間，看到老師對於學術的熱忱以及對於論文嚴謹的態度真的讓我獲益良多。

同時也感謝我的口試委員：顏汝芳教授與鄭宗記教授，感謝兩位教授願意非常仔細地閱讀我的論文，並在在論文口試時提出許多建議，讓我能夠找出論文中不夠穩健的部分並得以在論文修改時進行改善，短短40分鐘的口試讓我學到了很多。

也感謝我的父母願意提供我許多資源讓我獲取碩士學位，並願意讓我出國交換。論文的撰寫的過程中，我的父母給我許多意見與靈感，讓我在撰寫論文的過程更加地順利。

最後，感謝所有在我的論文撰寫的過程中給予幫助的同學、朋友、家人以及老師，您的支持與陪伴是我完成這篇論文的動力。

張邑齊 謹致

摘要

本研究之主要目的為透過文字探勘中的情緒分析方法以 ANTUSD 情緒詞庫量化新聞資料，利用新聞情緒分數衡量市場情緒，結合 OLS 與 GARCH(1,1)建立迴歸模型，探討新台幣匯率市場之效率性。本研究分別使用：1. 匯率相關之財經新聞全文 2. 匯率相關新聞中匯銀主管針對匯率走勢所發表之談話內容 兩種新聞資料，作為計算新聞情緒分數的樣本，並比較上述兩種新聞資料何者對於匯率之走勢有較佳的解釋能力。根據本研究之實證結果，以匯率相關之財經新聞全文計算之情緒分數作為變數之模型表現較佳。在樣本期間 2018/01/01 至 2019/03/20 內，以量化之新聞情緒分數衡量之市場情緒對於新台幣兌美元之匯率是存在解釋力的，且匯率的走勢的波動也會受到新聞情緒的變異所影響，代表新台幣匯率市場違反效率市場假說。

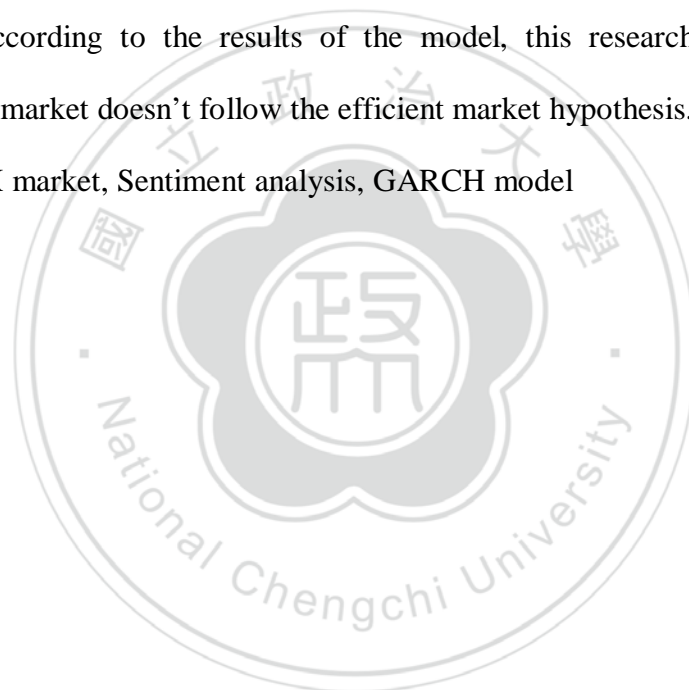
關鍵詞：外匯市場、文字探勘、GARCH 模型



Abstract

This study uses quantified news materials as a measure of market sentiment in order to test the efficiency of USD/NTD FX market. By quantifying FX rate related news materials with ANTUSD sentiment lexicon, news sentiment can be transformed into numeric variable that can be used to build up econometric model. Base on the result of the GARCH(1,1) model implemented in this thesis, within 2018/01/01~2019/03/20, news sentiment significantly affects the USD/NTW exchange rate. Moreover, the variance of the news sentiment also affects the volatility of the USD/NTW exchange rate. Hence, according to the results of the model, this research concludes that USD/NTW FX market doesn't follow the efficient market hypothesis.

Key words: FX market, Sentiment analysis, GARCH model



目錄

第一章、緒論	1
第一節、研究動機.....	1
第二節、研究方法與資料範圍.....	2
第三節、本文架構.....	3
第二章、文獻探討	4
第一節、市場微結構.....	4
第二節、效率市場假說.....	5
第三節、文字探勘在匯率研究中的應用.....	6
第三章、研究方法	8
第一節、文字探勘.....	8
第二節、資料搜集.....	9
第三節、新聞情緒分析.....	9
第四節、模型變數.....	12
第五節、變數檢驗與模型設定.....	13
第四章、實證結果	18
第一節、新聞資料的敘述性統計分析.....	18
第二節、以匯率新聞全文資料衡量市場情緒之實證分析.....	20
第三節、以匯銀主管談話衡量市場情緒之實證分析.....	23
第五章、結論與檢討	26
第一節、研究結論.....	26
第二節、研究檢討.....	27
參考文獻	29

圖目錄

圖 1:匯率走勢圖.....	16
圖 2:匯率循環變動圖.....	16
圖 3:匯率趨勢圖.....	16
圖 4:新聞分詞統計.....	19
圖 5:匯銀主管談話分詞統計.....	19



表目錄

表 1、詞庫比對率.....	11
表 2、變數 ADF 檢定之 P-value	14
表 3、新聞全文自變數之 VIF.....	16
表 4、模型變數之敘述性統計.....	19
表 5、以匯率相關新聞全文計算情緒分數之模型	22
表 6、以匯銀主管談話內容計算情緒分數之模型	25



第一章、緒論

第一節、研究動機

自美國總統川普當選後，反全球化趨勢興起，諸多如匯率操縱國、退出 WTO、中美貿易戰等政策皆對匯率有很大的影響，台灣為出口導向的經濟體，匯率的走勢與台灣出口產業的獲利有很大的影響，然匯率市場因透明度低、交易限制少以及匯率市場參與者存在異質性等因素導致預測匯率走勢的難度遠超過預測股票市場走勢的難度，且除匯率市場自身特性外，各國央行也會因為政治及經濟因素對匯率的走勢進行干預，然台灣央行並未公布央行干預的實際資料，使得新台幣匯率預測的難度大幅上升。誠如上所述諸重大經濟事件對於新台幣匯率的變動都會有極大的影響，而重大事件經濟事件皆會在經濟新聞中呈現，故引發本研究之研究者對於匯率市場的疑問，匯率市場的投資人是否會因為重大經濟新聞事件的報導或新聞報導中專業投資人對於匯率走勢的預測而改變投資決策，進而影響到匯率的走勢。

在與匯率的相關研究中，許多研究者都曾利用新聞資料來探討經濟事件對於匯率市場的影響，如張興華(2013)曾利用新聞資料作為衡量央行干預新台幣匯率市場的變數，來探討央行干預行為對新台幣每日匯率變動的影響。然在大多數以新聞資料作為分析依據的研究中多因新聞資料無法被量化的限制故只能用虛擬變數的方式納入模型，並沒有一個量化的標準可以衡量新聞資料所帶有的內容。且新聞資料在處理上費工耗時，常需要以手動閱讀的方式來過濾新聞，因此在研究樣本上也有限制。

隨著計算機效能的提升，處理大量資料的成本降低以及機器學習技術的進步，許多利用電腦處理大量文字資料的技術都被成功的建出，輔以程式語言如 R 語言的興盛使社會科學研究者能用較簡單且有效率的方式來處理文字類型的資料，其中文字探勘中的情緒分析技術使研究者能夠輕鬆的將文字資料所傳達的情緒加以量化，並依據文字資料所傳達情緒的不同給予每篇文章相對應的情緒分數，使文字資料在計量經濟學領域的應用除了虛擬變數外多了一個新的選擇。綜合上述原因，本研究

希望透過文字探勘技術量化匯率相關新聞資料，用以探討市場情緒對於匯率市場變動的影響。

第二節、研究方法與資料範圍

本研究欲利用文字探勘方法中的詞庫分析法(Lexicon-based approach)量化新聞資料，計算新聞資料的情緒並作為衡量投資人市場情緒的變數，利用最小平方法建構迴歸模型，以探討短期匯率市場的效率性，研究議題包括：

1. 新聞資訊揭露對於匯率市場的影響，若匯率市場符合效率市場假說，則新聞資訊的揭露對匯率的收盤價不會有任何影響，若匯率市場不符合效率市場假說，則新聞資訊的揭露對於匯率的收盤價會存在解釋能力。
2. 探討新聞資料中匯銀主管所發表之談話對匯率市場的影響，本研究所採用的新聞資料來源為聯合知識庫，聯合知識庫中與匯率相關的新聞多會包含匯銀主管針對匯率走勢的預測所發表的談話，本研究認為匯率市場參與者會參考匯銀主管所發表之談話來決定投資策，故欲探討匯銀主管談話的新聞情緒對於匯率收盤價是否存在解釋力。

本研究之樣本選擇範圍為 2018/01/01 至 2019/3/20 號之匯率收盤價與新聞資料，該期間因有央行總裁彭淮南先生退休，楊金龍先生接任與中美貿易戰等對匯率有重大影響的匯率新聞，因此選擇此期間進行研究較能觀察出匯率與市場情緒之間的關聯性。

第三節、本文架構

本研究分為五章，架構如下

第一章、緒論

闡述研究動機、研究方法與資料範圍及研究架構。

第二章、文獻探討

就市場微結構及效率市場角度探討上述領域的相關研究，並闡述本研究在此領域的定位與貢獻。

第三章、研究方法

介紹文字探勘與詞庫分析法，並按文字探勘之研究步驟依序詳述量化新聞資料的過程與方法，最後闡述本研究所使用的分析方法與迴歸模型。

第四章、實證結果

探討前章所述之分析方法與迴歸模型的實證結果。

第五章、結論與建議

就前章所述之實證結果做總結，並對本研究不足之處做檢討，以供後續相關研究參考。

第二章、文獻探討

自 1973 年後 Bretton woods 制度瓦解，各國的匯率制度從固定匯率制度轉為浮動匯率制度後，各種對於匯率預測的模型也因應而生。早期的匯率預測模型多由總體經濟構面出發，認為匯率的變動主要是來自於總體經濟指標的變動，然以總體經濟指標做為匯率預測變數的模型在實證上的預測能力往往不優於隨機模型。Engel and West (2005) 認為利用總體經濟指標做匯率預測的模型忽略了市場參與者對經濟指標的預期，該研究認為就中短期而言，投資人對未來經濟指標的預期比起當期經濟指標還要更有影響性。因此，將市場參與者的預期納入匯率預測的考量是非常重要的。本節將就市場微結構、效率市場假說兩個構面探討個體及市場效率性的角度所建構的匯率模型的之理論與實證文獻的回顧，最後再針對國內外應用文字探勘技術於匯率領域的相關研究做更深入的探討。

第一節、市場微結構

多數的匯率決定理論模型皆是以總體經濟基本面來解釋匯率的變化，然而經 Meese and Rogoff (1983) 的實證研究發現，匯率決定理論在實證上對中短期匯率變動的解釋能力遠低於隨機漫步模型，更無法用於預測短期匯率的變動，因此，以微觀角度探討匯率的模型也因應而生。所謂市場微結構是研究某個交易規則下的資產交易的過程和結果，主要研究市場參與者的行為與該行為對市場價格的影響。交易量是匯率市場微觀結構的重要因子，由於匯率市場具有非常龐大的交易量，而其中與貿易相關的交易比重卻非常的低，但傳統的總經匯率模型多著重於貿易相關的總體經濟指標，並沒有把交易量納入考量，有可能就是總體經濟匯率模型在實證上表現不如預期的原因之一。French and Roll (1986) 發現交易量本身可能導致匯率的波動，在資訊量不變的條件下，匯率主要是受到交易量的影響。Evans (2002) 則提出利用委託單流量(order flow) 來衡量作為匯率市場微結構理論的重要變數。委託單流量即是帶有正負訊息的交易量，其價值在於可以反映出投資人對於市場的預期，且委託單流量帶有私人訊息以及公開訊息的資訊，原因在於擁有私人訊息的投資人會透過委託單來影響到匯率變動，而擁有公開資訊的所有投資人也會藉由買賣雙方的委託單

來影響到匯率的變動。在國內對於委託單流量的研究如李建慧、黃信棠 (2011) 以 OLS 模型探討總體經濟指標宣告、委託單流量以及匯率報酬的關係，發現委託單流量對於匯率的報酬率有顯著的影響且總體新聞宣告對於委託單流量也有顯著的影響，即公開訊息對於交易量與匯率的變動存在影響性。

有鑒於委託單流量資料取得上不易，除了委託單流量，市場情緒也是衡量投資人預期的良好變數，市場情緒的研究被廣泛的被應用在股票市場上，Barberies, Shleifer and Vishny (1998) 研究發現投資人情緒對於股價會有反應過度以及反應不足 (under react) 兩種情形，該研究發現投資人對於一年內的新聞資料會有反應不足的情況，而對於三年以上的新聞資訊則會有反應過度的情況，意即可以利用短期新聞資訊在股價上反應不足的情況來賺取超額報酬。

目前利用市場情緒分析匯率變動的研究甚少，陳裕菘、謝邦昌、李勝輝、陳郁婷 (2014) 自匯率相關新聞中利用文字探勘技巧找出出現頻率最高的關鍵字，並針對關鍵字做推論樹分析來篩選出計量模型的變數是國內少數利用新聞資料衡量市場情緒來篩選模型變數的研究。而 Komariah, Machbub, Sin and Prihatmanto (2016) 則是透過分析 Twitter 資料來分析市場情緒與匯率波動的相關性以探討匯率市場的效率性。

第二節、效率市場假說

上述的匯率決定模型與效率市場假說不符，效率市場假說由 Fama(1970) 提出，該假說認為市場是非常有效率的，利用市場資訊來預測未來價格是不可行的，因為所有市場資訊都會被快速地反映在價格上，即投資人在欲利用市場上的額外資訊來預測市場價格的走向時，該資訊早已經被充分的反映在市場價格上，因此市場上並不存在超額報酬。效率市場假說分為三大類：弱勢(weak)效率市場、半強勢(semi-strong)效率市場、強勢(strong)效率市場。所謂弱勢效率市場認為市場價格是無法被前期的市場價格所影響的，半強勢效率市場則是認為除了歷史市場價格之外，公開發佈的新資訊如經濟基本面的衝擊等，皆無法用以預測未來的市場價格，而強勢效率市場假說則是認為市場上的所有資訊不論是公開與否，皆已經充分的反映在市場

價格上，也因此這些資訊對於市場價格的預測是無益的。目前有許多研究都是使用情緒分析法來探討匯率市場效率性的問題，例如 Komariah, et al. (2016) 利用 Twitter 資料進行情緒分析(sentiment analysis) 用以衡量投資人對市場的預期，並探討該資訊與印尼盾兌美元匯率走勢的相關性。Brouwer (2017)則是透過文字探勘與量化內容分析法同時分析財經新聞資料，並使用 OLS 迴歸模型觀察當期以及前期市場情緒對於英鎊兌美元匯率走勢的影響。

第三節、文字探勘在匯率研究中的應用

本節將就上述兩節所提及，對利用文字探勘技術研究匯率市場的研究做更進一步地闡述。在國外文獻方面，Komariah et al. (2016) 利用情緒分析法探討印尼盾兌美元匯率市場的效率性，該研究將 Twitter 資料作為衡量匯率市場參與者的市場情緒指標，透過結合詞庫分析法與 Naive Bayes classifier 分析 3235 筆 Twitter 資料，將 Twitter 資料分為正面與負面兩種情緒，並建置混淆矩陣(confusion matrix)來探討上述的文字探勘方法的精確程度。Komariah et al. (2016) 從資料中分別人工選取正負面新聞共兩百篇，並衡量人工判讀與文字探勘的結果符合的程度已計算分類方法的準確度，發現模型的精確度為 69%，意即在 400 筆 Twitter 資料中，有 276 筆的情緒分數與人工判讀是相符的，最後，該研究使用共變異數分析法，計算出 Twitter 資料的情緒方向（正面與反面）與匯率市場的走勢的皮爾森相關係數為 65.7%。Brouwer (2017) 則是利用情緒詞庫分析量化財經新聞的情緒，探討匯率市場的效率性，Brouwer(2017)將欲研究的新聞資料分為兩類，一類為不論主題的所有新聞，另一類則是財經新聞，並利用迴歸分析分別探討兩類新聞對於英鎊兌美元匯率收盤價的影響，發現在財經新聞資料中，當期的新聞情緒對於匯率的報酬率的影響是顯著的，然在加入前一期與前兩期的新聞後，新聞情緒對於匯率的影響則皆為不顯著；而在以不分主題的全部新聞資料中作為變數的模型中，則是所有新聞情緒對匯率的走勢皆無任何解釋力，因此該研究的結論認為英鎊兌美元匯率市場符合效率市場假說。

在國內，陳裕菘等人 (2014)曾利用文字探勘與資料採礦等方法結合時間序列模型建構匯率預測模型，該研究透過文字探勘法計算不同詞彙出現的詞頻，結合過往相關研究所使用的變數，找出九種可能對匯率有影響的變數，並利用資料採礦方法

中的特徵選取法對變數進行篩選，最終篩選出『貿易餘額』、『加權股價指數』、『國際原油價格』與『貨幣市場利率』作為模型的變數，複合式時間序列演算法建構模型。最終發現匯率的走勢除與自身匯率前期的變化有關外，同時也與前三期加權股價指數有高度的關聯性，並依上述結果，選擇前兩期之匯率與前三期之加權股價指數作為預測匯率的指標。

目前國內鮮少有人利用情緒分析法研究新聞內容對匯率之影響的實證分析，因此本研究欲透過文字探勘將新聞資料的情緒量化，作為市場投資人對於匯率未來走勢看法的衡量，用以檢測新台幣兌美元的匯率市場是否符合效率市場假說，以及是否市場上的所有資訊都直接且迅速的反應在匯率的價格中，並判斷匯率相關新聞的情緒對於預測匯率市場的走向是否有影響。



第三章、研究方法

在前述的章節中提及了許多衡量投資人預期的方法，而本研究所欲採取的方法是利用文字探勘技術量化新聞資料的情緒，即利用匯率相關新聞所傳達的情緒來衡量投資人對匯率的預期。舉例來說，2018年10月3號經濟日報一篇探討中美貿易戰的新聞提及美國欲利用加拿大、墨西哥等國達成的貿易協定作為與中國談判的籌碼，因而導致市場擔憂此舉會影響到台灣的出口市場，使得市場多偏向美元市場，導致新台幣重貶 1.12 角。在上述新聞中，可以明顯地感受到這是一篇對於新台幣匯率帶有負面情緒的新聞。透過最小平方法建構迴歸模型分析新聞情緒與匯率變動的關係後，若匯率新聞的情緒並無法解釋匯率的變動（匯率投資者在得知這個消息後更新了他們對匯率的預期，並反映在他們的交易行為上），則效率市場假說成立。反之，若新聞情緒的正面與負面對匯率有所影響，則效率市場假說不成立。

第一節、文字探勘

文字探勘即是處理文字資料的方法，透過處理非結構化的資料，找出文章中隱含的資訊。近年來網路發達，許多資訊都是以文字的形式來傳達，也因而產生了大量的文字資料，而這些文字資料往往隱含許多有用的資訊。舉例來說，行為財務學與行為經濟是分析消費者或投資人因為心理或其他因素而使其在消費或投資決策上會有不理性的行為，而在實證研究上，消費者或投資人的心裡因素往往是以文字資料如通訊軟體中的對話、對新聞的評論與社群軟體的留言等方式來呈現，而在研究上文字資料的應用相對數值資料有較多的限制，因此在進行上述領域的研究會遇到阻礙，如今透過文字探勘的方式，消費者或投資人的情緒得以被量化，因此文字探勘技術對於諸多社會科學領域的研究有很大的幫助。進行文字探勘的方法有很多種，例如機器學習，統計學、詞庫分析...等，而本研究欲採取的研究方法為情緒詞庫分析法，該方法是將文章斷詞後與情緒詞庫相比對，找出文章中與情緒詞庫資料庫相符的詞彙，並將配對成功的詞彙的情緒分數加總，除以文章的詞彙總數，計算出每一篇新聞的情緒分數，本研究將辭典分析法分為資料搜集、資料處理、情緒分數計算三個步驟，以下將依序詳述詞庫分析法各步驟。

第二節、資料搜集

本論文資料搜集的方法是參考張興華(2013)與柯秀欣(2018)中蒐集央行干預新聞的方法，其中匯率資料來源為台北外匯市場發展基金會的每日匯率資料，而新聞資料的來源則為聯合知識庫的新聞資料，其中匯率相關新聞的資料是透過關鍵字“匯銀主管”、“新台幣/美元”來擷取 2018/01/01 至 2019/05/30 間共 230 篇與新台幣兌美元匯率相關的新聞資料。將“匯銀主管”納入關鍵字的原因在於本研究主要的目的為分析新台幣/美元相關中文新聞對於匯率投資者的行為的影響，因此在進行新聞情緒分析時也必須將欲分析的新聞資料鎖定在對新台幣或美元會有影響的新聞。若僅將關鍵字設定為“新台幣/美元匯率”時，時常會找到包含他國匯率變動相關的內容，若依此類新聞進行新聞的情緒分析對新台幣/美元的分析時，會因為分析的新聞中包含非新台幣美元匯率相關的資訊而使得新聞情緒分數受到影響。經研究者手動篩選的方式發現，透過“匯銀主管”、“新台幣/美元”的關鍵字搜尋方式所找出的新聞較符合研究所欲達成的目標。此外，由於透過上述關鍵字所篩選出來的匯率新聞中通常包含了各家報社對匯銀主管得訪談內容。本研究除了分析匯率相關新聞的情緒外，也欲研究匯銀主管所發表的談話是否會對匯率變動有所影響。

第三節、新聞情緒分析

詞庫分析法簡介

本研究欲採用的文字探勘方式為詞庫分析法(Lexicon-based approach)，所謂詞庫分析法係指將文章切割為數個詞彙所組成的詞袋(bag of words)，再將詞袋中的字與已經被建立好的情緒詞庫中的詞彙做比對，透過加權平均的方式計算出該篇文章的情緒分數，簡而言之，即是透過計算一篇文章中正負面情緒辭彙出現的頻率去衡量一篇文章的情緒分數。在情緒詞庫中通常會包含許多詞彙以及各個詞彙所對應的情緒分數（大部分的情緒詞庫的情緒分數僅分為 3 種，即負面詞彙的-1 分，中性詞彙的 0 分以及正面詞彙的 1 分）。情緒詞庫分析法的優點在於操作簡單，許多文字探勘方法是由機器學習或統計學的方式來操作，意即需要大量資料作為訓練集，才可以建立出有效的情緒分類模型，例如著名的支援向量機(support vector

machine)即是透過輸入大量的訓練資料，透過建立一個超平面(hyper plane)將資料切為兩個部分（在文字探勘的例子中即是將詞彙集切分為正面詞彙與負面詞彙）。然而，由於詞庫分析法所使用的情緒詞庫是已經被建立好的，也因此操作時並不需要特別龐大的資料做為訓練集，分析時也僅需比對文章中的詞彙與詞典中相符的詞彙並處理該詞彙的情緒分數即可，在實務操作上較為簡單便利。

（二）情緒詞庫選擇

目前中文情緒詞庫仍非常稀少，本研究所欲使用的情緒詞庫為台灣大學自然語言處理實驗室陳信息教授所建立的擴增台灣大學情緒分析詞庫 ANTUSD (Augmented National Taiwan University Sentiment Dictionary)情緒詞庫，該詞庫對於每一個詞有六個欄位，分別是 Cope0pi 語意分數、正面標記數、中立標記數、負面標記數、非意見詞標記數。¹其中 Cope0pi 語意分數即是本研究欲用以計算文章情緒分數之工具。Cope0pi 是一個語意分析的工具，概念來自於中文構詞，是一個非監督式的工具，其特點在於相較於一般的情緒詞庫，Cope0pi 所計算出來的情緒分數不僅僅只有正反情緒分類，還能包含不同情緒的強弱，由 Cope0pi 所計算出的情緒分數範圍介於-1 至 1 之間，其中 1 代表最正面，-1 則是代表最負面，此外，由於 ANTUSD 就是以新聞資料來建立的，因此在新聞情緒的衡量上也會比其他的中文情緒詞庫還要來得好。

（三）資料處理—斷詞

利用情緒詞庫分析法進行中文文章情緒分析時，最重要的資料處理步驟就是對文章進行斷詞，由於中文的書寫方式並不會像英文一樣字與字之間並不會有空格，因此需要透過斷詞的方式將句子切割成數個有意義的詞彙，此步驟被稱之為斷詞。斷詞方法的選擇對於文章語意的判讀會有非常大的影響，不佳的斷詞套件甚至有可能會計算出與文章語意完全相反的情緒分數。本研究所採用的斷詞套件為 JiebaR，在斷詞的過程中儘管會有一些不精確，但整體來說由於新聞文章的架構以及用字都較其他文體來得精確，因此大多數情緒相關的詞匯皆有精準的被此套件成功斷詞。

¹ 感謝中央研究院資訊科學研究所提供 ANTUSD 情緒詞庫供本研究做使用。

表 1、詞庫比對率

	極小值	中位數	平均數	極大值
新聞全文	12.28%	26.34%	26.75%	48.91%
匯銀主管談話	5%	28.57%	29.17%	62.50%

註：詞庫比對率之計算方式為：文章斷詞後與情緒詞庫相符之詞彙數/斷詞後詞彙總數。

舉例來說，透過 JiebaR 套件對“國際經濟走弱，美元轉強”這句話進行斷詞後的結果為“國際”“經濟”“走弱”“美元”“轉強”。

(四) 詞庫比對率之計算

在計算新聞情緒分數之前，我們必須要先探討新聞斷詞後的詞彙成比對情緒詞庫詞彙的比率有多少。原因在於成功比對的詞彙量之多寡會間接影響到情緒分析的精準度，若成功配對的詞彙太少，可能會導致情緒分數在計算上與實際的新聞情緒有所落差，因此在計算情緒分數前，必須要先驗證一篇文章有多少比率的詞彙是有相對應的新聞分數的。經統計後發現，每一篇新聞全文大平均有 27% 的詞彙是被建立在 ANTUSD 詞典裡的，而新聞內文中匯銀主管所發言的詞彙則是有 29% 的比對成功率，在考量到一篇新聞裡通常會有許多專有名詞、數字、人名及公司名是不會被建立在情緒詞典中，且上述類型的詞彙屬於中性詞匯，對於情緒分析上也不會有任何的影響，因此認為這樣的比對成功率是可以穩健地進行情緒分析而不會有太大的偏誤的，表 1 為比對率之敘述性統計資料。

(五) 新聞資料情緒分數計算

利用情緒詞庫分析法的研究中，大多數的研究者是以文章中正面文字出現的比例減去負面文字出現的比率，即衡量一篇文章的正面程度。然而，因為 ANTUSD 與大多數的情緒詞庫不同，由 Cope0pi 系統所計算出之詞彙情緒分數有強弱之分，所以在情緒分數的計算上與其他情緒詞典的計算方式不同。本研究透過將所有詞彙的情緒分數依照每個辭彙出現的平率做加權平均計算出一篇新聞的情緒分數。計算公式如下：

$$\frac{\sum_{i=1}^n (score_i)}{n}$$

其中 n 代表一篇文章配對成功的詞彙總數， $score_i$ 則代表新聞斷詞後第 i 個詞彙的情緒分數。

第四節、模型變數

在成功量化新聞情緒後，即可透過迴歸分析來觀察新聞情緒對匯率變動的解釋能力。首先，為了討論新聞情緒對匯率的變動是否存在顯著影響，本文分別將 1.匯率新聞全文 2.新聞資料中匯銀主管針對匯率走勢所發表之談話 作為衡量新聞情緒的指標，以比較上述兩種文體何者對於匯率之走勢有較佳的解釋能力。此外，除了情緒分數外，本研究也將美元指數與聯邦基金率納入自變數，用以鎖定新台幣兌美元匯率因為美國經濟景氣變化對美元本身所產生的影響，下面就分別就納入模型的變數做解釋。²

1. 當期匯率新聞/匯銀主管談話情緒分數

即透過前述方法所計算出來之匯率新聞/匯銀主管談話的情緒分數，若經模型結果檢定係數顯著，即代表匯率收盤價是可以被新聞情緒所解釋，意即市場資訊對匯率的變動是有解釋能力的，效率市場假說不成立，即匯率市場是沒有效率的，過去的市場資訊對於匯率的預測是有效的。若迴歸係數不顯著，則表示新聞情緒分數對於匯率收盤價是沒有解釋能力的，即匯率市場是有效率的，所有的市場資訊都已經被反映在當前匯率的價格之中，市場上不存在超額報酬，效率市場假說成立。

2. 前期匯率新聞/匯銀主管談話情緒分數

本研究除了探討匯率是否能被當期的新聞情緒分數解釋外，也欲研究前期的匯率新聞情緒是否能用以解釋本期的匯率收盤價，意即，前期的新聞匯率新聞/匯銀主管談話的情緒分數是否可以用來解釋今日匯率的變動，若經模型檢定係數顯著，代表前期的新聞/匯銀主管談話的情緒分數對匯率收盤價具有一定的預測能力。

² 本文也嘗試要放入台灣的總經變數做為控制變數，但結果顯示放入之變數於模型中都不顯著，可能是資料頻率所致，感謝口試委員顏汝芳教授的指正與提醒。

3. 美元指數

美元指數是衡量美元在外匯市場匯率變化的指標，美元指數的組成包含了與美國貿易頻繁知國家的貨幣，包含日元、加幣、歐元、英鎊、瑞典克朗、瑞士法郎，其中歐元佔指數的權重最大，為 58%，14%得日圓次之。本研究欲透過加入美元指數作為自變數來鎖定美元本身因為貿易條件或美國國內經濟等因素而產生的升值或貶值效果，藉以提升新聞情緒分數的係數值之顯著性，原因在於本研究所篩選的新聞資料為新台幣匯率相關的國內中文新聞，該類新聞對於新台幣兌美元匯率收盤價的影響有很大一部份僅包含了台幣自身的升貶值，較無法包含美元自身的變動，因此為了鎖定美元自身的變動，本研究將美元指數加入自變數來增加模型的顯著性，然因美元指數的數值較其他變數之數值大，故以取對數的方式納入模型。美元指數資料取自 FRED economic data。

4. 聯邦基金率

聯邦基金率為美國一家金融機構之間借貸的利率，即是各家銀行間的同業拆款利率，由於聯邦基金率的高低會影響到美元的供給，因此會對美元的匯率產生影響。本研究將聯邦基金率納入自變數乃是考量到美元匯率的變動也會受到聯邦基金率得影響，為避免迴歸模型忽略重要變數而影響到新聞情緒變數之係數的顯著性，故將 FFR 放入模型中。聯邦基金率資料取自 FRED economic data。

第五節、變數檢驗與模型設定

(一) 單根檢定

上節所討論的變數皆屬於時間序列變數，然在使用時間序列變數作為迴歸模型的變數時，必須要先檢驗變數是否為定態時間序列，因為迴歸模型中的兩變數可能會因存在隨機趨勢而導致虛假迴歸的問題，意即原本不應該相關的兩個變數因為隨機趨勢的存在而導致迴歸係數為顯著。因此，在進行迴歸分析之前，必須要先對重要的變數：匯率以及新聞情緒分數做單根檢定本研究選擇使用 Augmented Dickey-Fuller test 來檢驗上述變數中的單根存在與否。下表 2 即為 ADF 檢定之結果。自表中

表 2、變數 ADF 檢定之 P-value

	匯率	新聞情緒分數	匯銀主管談話分數	美元指數	聯邦基金率
<i>P-value</i>	0.7571	<0.01	<0.01	0.2778	0.0874

註：假設檢定：H0：存在單根 H1：不存在單根

可以看出除了新聞情緒分數與匯銀主管談話分數以外，其餘變數皆非定態變數，故在進行分析之前，必須要針對非定態的變數做處理。

(二) Hodrick-Prescott Filter

首先，我們先繪製出資料期間匯率的走勢圖以探討為何匯率變數並非定態時間序列。如圖 1 所示，發現匯率收盤價在本研究鎖定的樣本期間內存在明顯的時間趨勢，為了解決時間趨勢的問題使匯率變數回歸到定態的時間序列變數。本研究欲使用 Hodrick-Prescott Filter (HP-Filter) 將變數中定態的部分分離出來，HP-Filter 可以將一存在時間趨勢的時間序列資料拆解為長期趨勢(Trend)與循環變動(Cycle)兩個部分，其計算之公式如下：

$$TR_t^{HP} = \arg \min \sum_{t=1}^T (y_t - TR_t) + \lambda \sum_{t=2}^{T-1} [(TR_{t+1} - TR_t) - (TR_t - TR_{t-1})]^2 \quad (1)$$

HP-Filter 的原理為自變數本身 y_t 中找出一隨機趨勢 TR_t ，該隨機趨勢可以極小化 y_t 與 TR_t 之間差異的變異，然若僅僅透過尋找能夠極小化 y_t 與 TR_t 之間差異的 TR_t 的解會是 y_t 本身，因此 HP-Filter 在公式中加入懲罰項，即是公式(3)中加號右邊的部分，目的是限制趨勢項 TR_t 的波動。利用 HP-Filter 處理匯率收盤價資料後，即可得到匯率收盤價中屬於循環變動的部分。在上方公式(3)中之 λ 為控制隨機趨勢平滑程度之參數，越大的 λ 表示隨機趨勢越平滑，根據 Hodrick and Prescott (1997) 參數 λ 之設定會根據資料之頻率而有所不同，在嘗試許多不同 λ 後，本研究決定設定 λ 為 14400。

透過 HP-Filter 所分解出來的趨勢代表的是匯率收盤價受到恆常性衝擊所導致永久性改變的部分，而循環變動則是代表匯率收盤價受短暫衝擊造成的短期變動。本研究所欲探討的是當一篇匯率新聞出現時，市場參與者是否會根據新聞情緒來調整

投資策略，進而影響匯率未來的走勢，並依此探討匯率市場的效率性，故著重於新聞情緒對匯率短期變動的部分，因此選擇循環變動作為迴歸模型的應變數。

為了更直觀的了解 HP-Filter 的功能，本研究以圖像的方式呈現在使用 HP-Filter 後的產出。下圖 1 為匯率收盤價在樣本期間內的走勢，而圖 2 為經 HP-Filter 拆解後匯率循環變動的部分，而圖 3 則是透過 HP-Filter 所拆解出匯率長期時間趨勢的部分，從圖 1 與圖 3 可以發現兩者線條非常相似，代表 HP-Filter 成功的分離出匯率收盤價資料中屬於長期趨勢的部分。此外，對剩下的匯率收盤價循環變動資料做單根檢定後發現 P-Value 小於 0.01，表示匯率循環變動資料為定態資料，作為時間序列迴歸模型的變數是穩健的。

(三)、共線性診斷

在變數的選擇上，美元指數與聯邦基金率皆是作為美金自身波動的變數，然聯邦基金率可能與美元指數具有高度的相關性，因為美元指數衡量的是美國與其貿易頻繁之國家的匯率的相對強弱，一定程度上可以反映出美國經濟的狀況，進而影響到聯邦基金率的設定。在此情況下，美元指數與聯邦基金率可能存在共線性問題，若將存在高度共線性的變數皆放入模型中，會導致迴歸係數變異數過大，造成迴歸係數不顯著。

本研究採用變異數膨脹因子(VIF)作為衡量高度共線性問題存在與否的指標，變異數膨脹因子的計算方式為將欲研究的迴歸模型中的自變數分別對模型中其他自變數進行迴歸分析，即輔助迴歸(auxiliary regression)，並計算出每個輔助迴歸的判定係數，將 1 減去判定係數後做倒數即可得出變異數膨脹因子，如下所示：

$$VIF_i = \frac{1}{(1 - R_i^2)} \quad (2)$$

由上式可以看出，若一自變數可以被模型中其他自變數以線性迴歸模型解釋的話，代表該變數與其他自變數之間存在高度的共線性，就實務上來說變異數膨脹因子大於 10 的自變數即被視為存在共線性問題。下表 3 為本研究兩個迴歸模型自變數之變異數膨脹因子。

圖1 匯率走勢圖



圖2 匯率循環變動圖

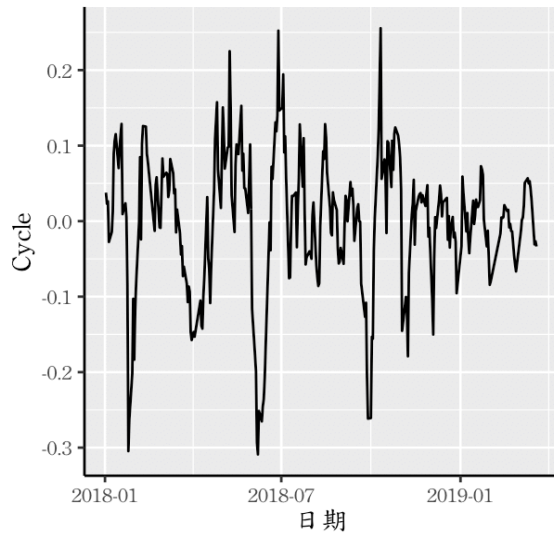
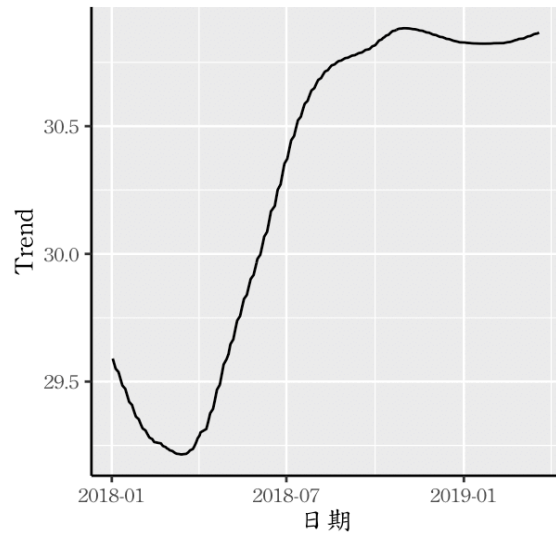


圖3 匯率趨勢圖



註：

1. 圖 1 為 HP-filter 前匯率之走勢，圖 2 圖 3 分別為經 HP-filter 分解後之循環變動與時間趨勢。
2. 分解後屬於匯率循環變動的 Cycle 變數(圖 2)，經檢定後為定態時間序列資料。

表 3、新聞全文自變數之 VIF

註：根據實務法則，變異數膨脹因子超過 10 之變數即存在高度共線性問題。

	$textscore_t$	$textscore_{t-1}$	$textscore_{t-2}$	USDX	F.F.R
VIF	1.03	1.05	1.02	6.36	6.31
	$speechscore_t$	$speechscore_{t-1}$	$speechscore_{t-2}$	USDX	F.F.R
VIF	1.12	1.07	1.06	5.17	5.58

由表 3 可以發現，雖然美元指數與聯邦基金率的确如前所述，具有較高的變異數膨脹因子，但仍未超過 10，因此本研究認為將美元指數與聯邦基金率放入迴歸模型中並不會有嚴重的共線性問題。

(四)ARCH-LM 檢定與模型設定

觀察圖 2 匯率循環變動之圖像，可以發現匯率循環變動之波動似乎存在波動群聚(volatility clustering)的現象，即匯率收盤價循環變動之波動因時點變動而改變，而此一現象會導致迴歸殘差存在變異數異質性的問題。因此本研究採用經 ARCH-LM 檢定來檢定匯率收盤價變數是否存在 ARCH 效果。根據檢定結果，匯率收盤價在顯著水準 0.01 以下拒絕無 ARCH 效果之虛無假設，代表匯率收盤價之循環變動存在 ARCH 效果，必須要在模型的設計上對變異數進行處理。

因此，為解決波動性群聚現象所帶來之變異數異質性問題，本研究決定以 GARCH(1,1)模型估計匯率收盤價循環變動之條件變異數，用以捕捉波動性群聚現象的特性，使模型再度回歸同質變異數假設。此外，在建立 GARCH 模型時，本研究在模型中也加入當期與前期新聞情緒分數的平方作為衡量新聞情緒分數變異之變數，以探討匯率收盤價之循環變動的變異是否也會受到新聞情緒分數的變異所影響。總結本節針對變數處理與篩選之結果，設定本研究之模型如下所示：

$$Cycle_t = \beta_0 + \beta_1 sentiment_t + \beta_2 sentiment_{t-1} + \beta_3 sentiment_{t-2} + \beta_4 USDX_t + \beta_5 F.F.R_t + \varepsilon_t. \quad (3)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 (sentiment_t)^2 + \beta_3 (sentiment_{t-1})^2. \quad (4)$$

第四章、實證結果

第一節、新聞資料的敘述性統計分析

(一)詞頻分析

在探討模型結果之前，本研究欲先以詞頻分析找出新聞資料中出現頻率較高的詞彙，並以較簡易的角度探討新聞資料中所帶有的資訊與匯率收盤價走勢之間的關係。

本研究在搜集新聞資料時發現某些報社會將許多國際財經新聞、股市新聞與匯率新聞融合為一篇新聞報導，因此一篇新聞有可能會包含了許多對新台幣兌美元匯率的變動影響較小的資訊，造成後續利用情緒分數衡量匯率變動時會有失準的情形，因此在進行新聞情緒分數的計算之前，必須先觀察新聞內文中出現頻率較高的詞彙，並探討這些高頻率出現的詞彙與匯率之走勢是否存在相關性。此外，分析辭彙出現的頻率也可以使研究者篩選出停用詞(stop words)，所謂停用詞係指對情緒分析沒有任何幫助的詞彙例如“的”、“之”等不帶有意義的连接詞。不同主題的文本根據研究者欲研究的方向會有不同的停用詞。如本研究將“匯率”、“匯銀”、“主管”等詞彙設為停用詞，原因在於上述詞彙即是本研究篩選新聞的關鍵字，因此該詞彙出現頻率自然就會比較高，在分析詞頻時納入討論就會失去意義。圖 4 與圖 5 分別為兩種文本中出現頻率最高的 25 個詞彙。

自上述的詞頻圖表中可以觀察到，“外資”詞彙不管是在匯銀主管談話或是新聞全文中出現頻率都遠高於其他詞彙，故可推論外資對於新台幣兌美元匯率有顯著的影響，此外我們也可以觀察到在兩種文本當中“貶值”出現的頻率都比“升值”出現的頻率還要高，因此可以推斷在這段期間內市場對於新台幣兌美元市場的看法較偏向貶值。與圖 1:新台幣兌美元 2018 年 1 月 1 號至 2019 年 3 月 20 號之間的匯率收盤價走勢比較，可觀察出匯率的走勢與詞頻分析所觀測到的市場情緒方向一致。

圖4 新聞全文分詞統計

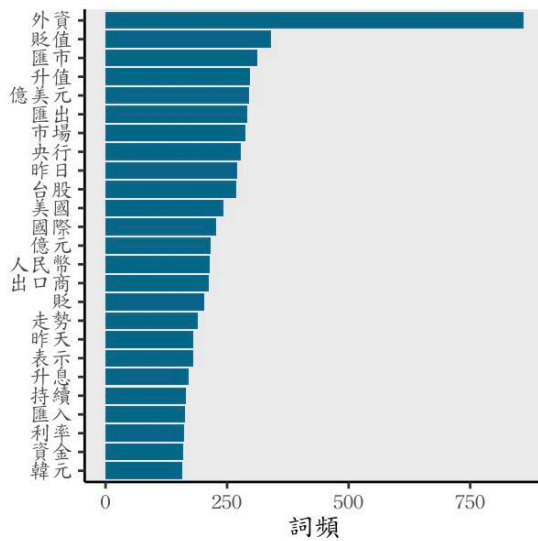
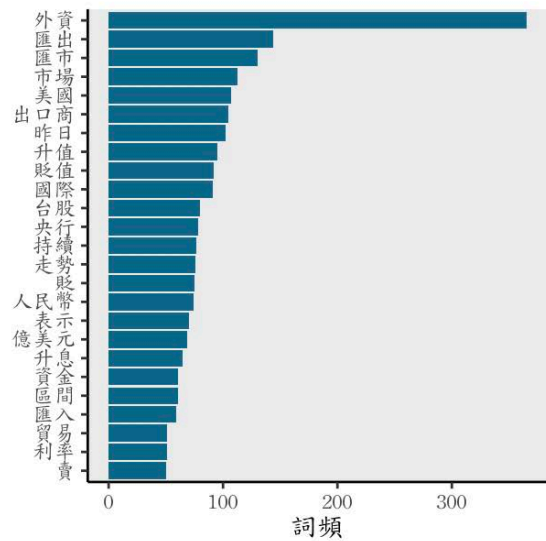


圖5 匯銀主管談話分詞統計



註:

- 圖 4、5 為斷詞後之新聞全文與匯銀主管談話中，出現頻率最高的前 25 個詞彙。

(二)模型變數的敘述性統計分析

表 4 為模型所使用之變數資料的敘述性統計分析，其中新聞全文資料與匯銀主管談話之資料係將當日所有新聞之新聞情緒分數加總後所計算出之結果進行統計。由表中可以發現，匯率循環變動、新聞全文資料情緒分數以及匯銀主管談話的情緒分數，平均數都接近 0。

表 4、模型變數之敘述性統計

	標準差	極小值	中位數	平均數	極大值
匯率循環變動	0.1763	-0.3349	-0.0406	-0.0100	0.3560
新聞全文	0.0591	-0.0918	0.0314	0.0380	0.2494
匯銀主管談話	0.0842	-0.1825	0.0420	0.0456	0.3022
美元指數	3.9844	115.2	123.9	122.8	129.1
聯邦基金率	0.3040	1.42	1.91	1.88	2.40

第二節、以匯率新聞全文資料衡量市場情緒之實證分析

(一)模型結果

表 4 為迴歸模型之結果，本研究首先僅將當期之新聞情緒分數放入模型中建立簡單迴歸模型(模型(1))，之後再將美元指數與聯邦基金率兩個控制美元自身變動的重要變數加入模型中(模型(2))，緊接著將落後一期的新聞情緒也加入模型中(模型(3))，最後將落後兩期之新聞情緒分數放入模型(模型(4))，並比較上述 4 種模型的結果並依據 AIC 選出最佳模型以做出結論。

在僅有當期新聞情緒分數的模型(模型(1))中，當期新聞情緒分數對於匯率收盤價循環變動是存在顯著的解釋力的，且係數為負。在加入美元指數與聯邦基金率(模型(2))後，當期之新聞情緒分數之係數變為不顯著，不過美元指數與聯邦基金率之係數皆為顯著。而在納入落後期新聞情緒的模型中，在僅包含當期與前期新聞情緒分數的模型(3)中，當期的新聞情緒分數與前期新聞分數的係數都是顯著的，而加入落後兩期之新聞情緒變數的模型(4)中則是所有變數都是顯著的。而在 GARCH 的部分中，可以發現當期的新聞情緒分數之變異的係數在除了模型(1)以外之其他模型中都是顯著的，而落後一期之新聞情緒分數則是在所有模型中都是不顯著的。

(二)模型選擇與小結

本研究以 AIC 作為模型選擇之準則，根據表 4 之結果，可以發現模型(4)具有最小的 AIC，故選擇此模型作為本節之最佳模型，並依此模型之結果做出小結。在模型(4)中，當期、前期與前兩期的新聞情緒分數都是顯著的，代表新聞情緒分數對於匯率收盤價循環變動的走勢存在解釋能力。就係數之正負做討論，可以發現不論是哪一期的新聞情緒分數的係數都是負的，代表其他變數不變的情況下，當新聞情緒下降一單位時，新台幣兌美元匯率會上升一單位，即新台幣貶值。在第三章中，我們曾提及，雖然匯率貶值並不能代表負面的情緒，但與貶值相關的詞彙在情緒詞庫中的語意分數多為負分，反之亦然，因此一篇文章帶有較多與貶值有關的資訊時，該文章的情緒分數多會是負的。因此，當一篇帶有負分的貶值新聞資訊出現時，匯率市場參與者會根據此資訊來調整投資策略，進而導致匯率貶值。因此，此一結果是符合經濟直覺的。

接著，就係數絕對值之大小做分析，可以發現，落後兩期之新聞情緒分數之係數絕對值 0.6710 比落後一期之新聞分數的係數絕對值 0.6231 大，而該值又比當期新聞情緒分數之係數的絕對值 0.2575 還大。本研究認為，造成此一結果的主要原因為投資人對於接收到的新聞資訊無法及時反應在當下的投資行為中，因此會將今日所得到的資訊反應在未來的投資決策中，才會導致落後期的新聞情緒比當期的新聞情緒還要更有解釋能力。因此，根據此結果，匯率收盤價的循環變動是可以被新聞情緒給解釋的。

最後，本研究也欲探討匯率循環變動的變異是否會受到新聞情緒分數變異的影響，根據模型(4)中 GARCH 部分的結果，新聞情緒分數的變異對於匯率循環變動走勢的變異的影響是顯著的，且為正相關，表示新聞情緒的變異越大時，匯率循環變動的變異程度也會越大。因此，根據上述結果，新聞情緒對於匯率的循環變動存在解釋能力，且匯率循環變動的條件變異數也會受新聞情緒影響，表示匯率並無法及時反應所有市場資訊，投資人可以透過收集新聞資訊以獲得超額報酬，效率市場假說不成立，新台幣兌美元匯率市場是不效率的。

表 5、以匯率相關新聞全文計算情緒分數之模型

	(1)	(2)	(3)	(4)
截距	-0.0627*** (0.0066)	-1.0024*** (0.0102)	-1.0024*** (0.0023)	-1.0024*** (0.0004)
當期情緒分數	-0.3268** (0.1578)	0.0090 (0.1718)	-0.6599*** (0.1365)	-0.2575** (0.1018)
前期情緒分數			-0.9013*** (0.0223)	-0.6231*** (0.0741)
前兩期情緒分數				-0.6710*** (0.0281)
美元指數		0.0027*** (0.0001)	0.0038*** (0.0000)	0.0047*** (0.0000)
聯邦基金率		0.4000*** (0.0052)	0.3274*** (0.0010)	0.2801*** (0.0050)
GARCH				
截距	0.0013 (0.0008)	0.0012 (0.0000)	0.0001 (0.0001)	0.0001 (0.0001)
ARCH	0.6154*** (0.0243)	0.8854*** (0.1586)	1*** (0.1882)	0.9237*** (0.0149)
GARCH	0.3384*** (0.0742)	0.0853 (0.0507)	<0.0001 (0.0021)	<0.0001 (0.1151)
當期情緒分數變異	<0.0001 (0.4685)	1.3876*** (0.4662)	2.7838*** (0.9225)	1.5291*** (0.4085)
前期情緒分數變異	<0.0001 (0.2214)	<0.0001 (0.1611)	<0.0001 (0.0379)	<0.0001 (0.0268)
AIC	-1.0376	-1.2353	-1.2550	-1.3932

顯著水準: 0.01 '****' 0.05 '***' 0.1 '**'

括弧內為標準差

註:

1. 本研究將同一日發布之多篇新聞之情緒分數加總，作為該日之情緒分數，故模型之樣本數為 137。
2. 模型(4)因具有最小 AIC，故選擇模型(4)作為本小節之最佳模型

第三節、以匯銀主管談話衡量市場情緒之實證分析

(一)模型結果

表 4 為迴歸模型之結果，與上節相同，本研究首先僅將當期之匯銀主管談話之情緒分數放入模型中建立簡單迴歸模型(模型(1))，之後再將美元指數與聯邦基金率兩個控制美元自身變動的重要變數加入模型中(模型(2))，緊接著將落後一期的匯銀主管談話情緒也加入模型中(模型(3))，最後將落後兩期之匯銀主管談話情緒分數放入模型(模型(4))，並比較上述 4 種模型的結果並依據 AIC 選出最佳模型以做出結論。

在僅有當期匯銀主管情緒分數的模型(模型(1))中，當期情緒分數對於匯率收盤價循環變動的係數是不顯著的。在加入美元指數與聯邦基金率(模型(2))後，當期之新聞情緒分數之係數仍為不顯著，不過美元指數與聯邦基金率之係數仍為顯著。而在納入落後期新聞情緒的模型中，在僅包含當期與前期新聞情緒分數的模型(3)中，情緒分數的係數都是不顯著的，而加入落後兩期之情緒變數的模型(4)中落後兩期的情緒分數是顯著的，而其他期的新聞情緒分數都是不顯著的。而在 GARCH 的部分中，可以發現當期的情緒分數之變異的係數在除了模型(1)以外之其他模型中都是顯著的，而落後一期之新聞情緒分數則是在所有模型中都是不顯著的。

(三)模型選擇與小結

與上節相同，本研究以 AIC 作為選擇模型的準則，根據表 5 找出模型(4)為本節之最佳模型，並依此模型之結果做出小結。在模型(4)中僅有落後兩期之情緒分數是顯著的，且係數為正，然在上節我們已經討論過，在合理的情況下，情緒分數與匯率走勢之間的關係應該是負相關，正的係數並不符合經濟直覺。為探討發生此結果之原因，本研究手動比對匯銀主管談話資料以及期相對應的情緒分數後發現常有情緒分數與文意不符的情形發生，即情緒詞庫無法準確的衡量匯銀主管談話實際的情緒分數，原因在於：

1. 匯銀主管談話內容過少：在部分新聞資料中，匯銀主管的看法有時只會有一兩句文字，在利用情緒詞庫計算少量文字時，會因為字數過少導致某些特定詞彙的情緒分數會主導匯銀主管談話的情緒分數，造成情緒分數的計算上出現偏誤。
2. 口語類型的資料斷詞效果較差：口語類型的資料相對新聞文體資料在斷詞上較容易出現錯誤，原因在於新聞資料通常有固定的格式，且在用詞上會比較精準，

較不會出現縮寫、簡稱類型的詞彙。而匯銀主管談話在新聞資料中多以引述的方式呈現，因此匯銀主管談話資料相對於新聞全文資料的情緒分數衡量較容易產生偏誤。

綜上所述，由於匯銀主管資料之特性，因此情緒詞庫分析法並無法準確的衡量出匯銀主管談話的實際情緒，而依據此錯誤的情緒分數作為自變數所估計出來的模型結論也會是錯誤的，故認為匯銀主管談話相較於新聞情緒分數來說，較不適合使用情緒詞庫分析法所計算出的情緒分數來分析匯率循環變動。



表 6、以匯銀主管談話內容計算情緒分數之模型

	(1)	(2)	(3)	(4)
截距	-0.0675*** (0.0056)	-1.0024*** (0.0102)	-1.0024*** (0.0020)	-1.0024*** (0.0239)
當期情緒分數	0.0381 (0.1318)	-0.0801 (0.1343)	-0.0148 (0.0971)	0.0339 (0.1177)
前期情緒分數			-0.1325 (0.0948)	-0.0314 (0.1483)
前兩期情緒分數				0.2419* (0.1253)
美元指數		0.0098*** (0.0001)	0.0101*** (0.0000)	0.0098*** (0.0002)
聯邦基金率		-0.1435*** (0.0010)	-0.1580*** (0.0025)	-0.1556*** (0.0156)
GARCH				
截距	0.0009 (0.0007)	0.0001 (0.0001)	<0.0001 (0.0002)	0.0006 (0.0007)
ARCH	0.9293*** (0.1558)	0.9299*** (0.1734)	1*** (0.1564)	0.9519*** (0.1943)
GARCH	0.0174 (0.0325)	0.0398 (0.0455)	<0.0001 (0.0100)	<0.0001 (0.0960)
當期情緒分數變異	1.6593*** (0.5092)	1.3740*** (0.3854)	1.2440*** (0.3579)	0.9519*** (0.2704)
前期情緒分數變異	<0.0001 (0.0023)	<0.0001 (0.0049)	<0.0001 (0.0281)	<0.0001 (0.0602)
AIC	-1.0419	-1.1803	-1.1777	-1.1814

顯著水準: 0.01 '****' 0.05 '***' 0.1 '**'
括弧內為標準差

註:

1. 本研究將同一日發布之多篇新聞之情緒分數加總，作為該日之情緒分數，故模型之樣本數為 137。
2. 模型(4)因具有最小 AIC，故選擇模型(4)作為本小節之最佳模型

第五章、結論與檢討

第一節、研究結論

(一) 新台幣兌美元匯率市場違反效率市場假說

本研究試圖用量化的新聞資料衡量市場情緒，藉此探討新台幣兌美元匯率市場的效率性，在比較以匯率新聞全文情緒分數作為自變數的模型與以匯銀主管談話情緒分數作為自變數的模型後，發現匯率新聞全文的情緒分數較適合作為衡量市場情緒的變數。根據實證結果，在以新聞全文作為自變數的迴歸模型中，市場情緒對於匯率循環變動的影響是顯著的，且匯率市場參與者對於新聞情緒分數存在反應時間落差，即因投資人在接收到新資訊時並無法及時的反應在當期的投資行為當中，因此會將當期所接收之資訊反應在下一期的匯率投資策略，所以當期新聞資訊之揭露對於當期匯率循環變動的影響較前一期新聞資訊來的小。另一方面，就 GARCH 部分之結果，匯率循環變動的變異也會受到當期新聞情緒分數的變異所影響。因此，就市場效率性的角度來看，新聞全文資訊是用來解釋匯率循環變動的，投資人可以透過參考新聞資訊來獲取超額報酬，因此就本研究之結論而言，新台幣兌美元市場是不符合效率市場假說的。

(二) 與 Brouwer(2017)之比較

綜上所述，本研究不僅發現新台幣兌美元市場違反效率市場假說，也發現在使用不同的新聞文體作為衡量市場情緒的變數時，會因為不同文體所包含的資訊不同而使預測的結論不同。此結果可以用來解釋為何本研究之結論與 Brouwer(2017)的結論不同，在 Brouwer(2017)的研究中認為英鎊兌美元市場是符合效率市場假說的，因為在 Brouwer(2017)的研究中，所有落後期的新聞情緒分數皆是不顯著的，本研究認為此結果上之差異是來自於新聞資料選擇上的差異，在 Brouwer(2017)研究中分別以不分類之所有新聞與財經新聞作為變數衡量市場情緒，結果為以在所有新聞作為自變數的模型中，係數都是不顯著的，而在以財經新聞作為變數的模型中，僅當期的財經新聞情緒分數之係數是顯著的。結合本研究之發現可以推斷 Brouwer(2017)的模型係數之所以會不顯著的原因除了研究的市場不同外，也有可能因為 Brouwer(2017)

選擇的新聞資料範圍過大，包含了許多與匯率走勢無關的資訊，導致情緒分數並無法正確的反應匯率市場的市場情緒。

第二節、研究檢討

本研究所採用的文字探勘方法為詞庫分析法，優點在於操作容易且使用方便，不如其他機器學習方法需要大量的資料做為訓練集，但詞庫分析法具有下列缺點：

1. 情緒分析的精確度高度仰賴情緒詞典的準確性：由於新聞情緒分數來自於詞庫中情緒詞彙的分數加總，文字探勘的精確度會使用的詞庫不一樣而有所不同，且每種詞彙在不同類型的文本當中的情緒分數也會有所不同，故詞庫分析法存在精確度不足的風險。
2. 忽略詞彙之間的關聯性：由於詞庫分析法必須將文章進行斷詞後分別計算情緒分數，因此詞彙語詞彙之間的關聯性並沒有被納入考量，導致反諷語氣的詞彙在情緒分數的計算上會嚴重偏誤，許多詞彙會因為前後文的不同而有完全相反的情緒意義，因此在忽略詞彙關聯性的情況下進行分析會有情緒分數失準的風險，然在本研究中因為新聞文體是屬於較正式的文章，因此較不會存在諷刺語氣類的詞彙，因此此缺點對本研究的影響較小。
3. 主詞判別困難：詞庫分析法只計算了一篇文章中情緒詞出現的頻率，然而同樣一個形容詞詞彙有可能會因為主詞的不同而使句子想要傳達的情緒不同，然透過情緒詞庫法所計算出的分數卻會完全相同，有可能會導致情緒分數的偏誤。

故後續研究可採用其他文字探勘方法衡量新聞情緒，或待更多中文詞庫建出後分別用不同的詞庫做情緒分析，並與本研究結果進行比較以加強情緒分析的精確度。

除情緒分析的準確性之外，本研究因中文新聞全文資料搜集不易，且本研究之研究者對於資料爬蟲技術的了解不夠深入，無法透過爬蟲的方式自資料庫中搜集大

量的新聞資料，僅能透過手動方式搜集，故若後續有研究者對爬蟲技術較有鑽研，可以搜集更大量的樣本做分析，可以有更穩健的結論。



參考文獻

中文文獻

1. 張興華 (2013)。從央行干預新聞分析台灣央行外匯市場干預與台幣匯率之關係。證券市場發展季刊，25，95-112。
2. 李建慧、黃信棠 (2011)。市場交易訊息對匯率之影響。亞太經濟管理評論，15，71-88。
3. 陳裕崧、謝邦昌、李勝輝、陳郁婷 (2014)。運用文字探勘與資料採礦技術建立匯率預測模型 - 以人民幣兌新台幣為例。數據分析，9，133-146
4. 柯秀欣 (2018)。台灣央行外匯市場干預對台美匯率之影響 - 媒體資料之應用。經濟論文叢刊，46，297-322。

英文文獻

5. Barberis, N., Shleifer, A., and Vishny, R. (1998), "A model of investor sentiment," *Journal of Financial Economics*, 49, 307-343.
6. Brouwer, Y. (2017), "News Sentiment and Foreign Exchange Markets." (Master's thesis, Erasmus University Rotterdam)
7. Engel, C. and West, K. D. (2005), "Exchange Rates and Fundamentals," *Journal of Political Economy*, 113, 485-517.
8. Evans, M. D. D., and Lyons, R. K. (2002), "Order Flow and Exchange Rate Dynamics.," *Journal of Political Economy*, 110
9. Fama, E. (1970), "American Finance Association Efficient Capital Markets: A Review of Theory and Empirical," *Journal of Finance*, 25, 383-417.
10. French, K. R. and Roll, R. (1986), "Stock Return Variances: The Arrival of Information and Reaction of Traders.," *Journal of Financial Economics*, 17, 5-26.
11. Hodrick, R. J. and Prescott, E. C. (1997), "Postwar U.S. Business Cycles : An Empirical Investigation.," 29, 1-16.
12. Komariah, K. S., Machbub, C., and Prihatmanto, A. S. (2016), "A Study of Efficient Market Hypothesis to Predict Exchange Rate Trends Using Sentiment Analysis of Twitter Data.," *Journal of Korea Multimedia Society*, 19, 1107-1115.
13. Ku, L. W., Ho, H. W., & Chen, H. H. (2009). "Opinion mining and relationship discovery using CopeOpi opinion analysis system.," *Journal of the American Society for Information Science and Technology*, 60(7), 1486-1503.

14. Meese, R. A., and Rogoff, k. (1983), "Empirical Exchange Rate Models of The Seventies.," *Journal of International Economics*, 14, 3-24.

