

105cbc-經 1（委託研究報告）

「應用大數據提升台灣民間消費 預測」

本報告係計畫主持人的意見，不代表委託機關

及計畫主持人所服務單位之立場

委 託 單 位：中 央 銀 行

執行單位：世新大學財務金融學系

計畫主持人：郭迺鋒 副教授

中 華 民 國 106 年 1 月

目 錄

第一章 研究背景與目的	1
第二章 文獻回顧	5
第一節 搜尋大數據預測應用研究	5
第二節 網路大數據民間消費預測應用研究	9
第三章 研究架構與設計	15
第一節 網路大數據指標	16
第二節 民間消費指標資料處理架構	20
第四章 研究方法	24
第五章 研究成果	29
第一節 大數據預測指標主成分分析	29
第二節 應用大數據提升台灣民間消費預測實證分析	39
第三節 民間消費綜合指標預測效力實證研究	52
第四節 關鍵社群網站民間消費預測指標分析	65
第六章 結論	78
參考文獻	81
附件一 網路搜尋關鍵字設定討論會議會議紀錄	85
附件二 期中審查意見及回應表	89
附件三 期末審查意見及回應表	115
附件四 社群網路資料擷取平台	146
附件五 網路大數據—搜尋指標篩選過程	149
附件六 15 大電商網主成份分析成果	153
附件七 應用大數據提升台灣民間消費預測-VAR 實證分析	168
附件八 民間消費綜合指標預測效力-VAR 實證分析	172
附件九 ARIMA 模型分析(納入 CCI)	179

圖目錄

圖 1-1	消費者網路行動模式	2
圖 2-1	美國 2008 年與 2013 年家計網路關注度	6
圖 2-2	民眾網路事務關注度變化	7
圖 2-3	民間消費與調查指標線性趨勢	12
圖 2-4	詞藻集群分析範例	14
圖 3-1	研究架構圖	15
圖 3-2	研究架構圖	20
圖 3-3	民間消費預測指標模型資料處理架構	21
圖 3-4	民間消費預測指標 DFM 模型資料處理架構	23
圖 5-1	15 大電商購物網站各網站中類的分類項目總計	30
圖 5-2	網路檢索指標主成份分析結果之陡坡圖	31
圖 5-3	總體經濟領先指標主成份分析結果之陡坡圖	34
圖 5-4	民間消費預測綜合指標靜態主成份分析	35
圖 5-5	主成份分析結果之陡坡圖	36
圖 5-6	DFM 模組分析	37
圖 5-7	樣本內靜態預測趨勢	45
圖 5-8	樣本內動態預測趨勢	45
圖 5-9	樣本外靜態預測趨勢圖	47
圖 5-10	樣本外動態預測趨勢圖	47
圖 5-11	VAR 樣本外預測趨勢圖	50
圖 5-12	樣本內靜態預測趨勢圖	57
圖 5-13	樣本內動態預測趨勢圖	57
圖 5-14	樣本內靜態預測趨勢圖	58
圖 5-15	樣本內動態預測趨勢圖	58
圖 5-16	VAR 樣本外預測趨勢圖	62
圖 5-17	關鍵社群網站民間消費分析架構	67
圖 5-18	博客來-聲量圖	68
圖 5-19	博客來-情緒指標圖	68
圖 5-20	博客來-盒鬚圖	69
圖 5-21	博客來-分配圖	69
圖 5-22	博客來-相關性圖	70
圖 5-23	博客來-樹狀圖	71
圖 5-24	博客來-集群網絡圖	71

表目錄

表 1-1	2014 年各國民眾每日上網使用分鐘數	2
表 2-1	旅遊主題 Google Trends 搜尋關鍵字	9
表 2-2	BEA 與 Google 品項可對照之分類	10
表 3-1	Google Trends 關鍵字來源網.....	17
表 3-2	各電商網商品品項分類數	17
表 3-3	關鍵字設定與網路大數據指標蒐集步驟	18
表 5-1	檢索指標主成份分析結果之解釋總變異量	31
表 5-2	網路檢索指標平方差解釋變異量	32
表 5-3	總體經濟指標資料建置	33
表 5-4	領先指標主成份分析結果之解說總變異量	34
表 5-5	總體經濟領先指標平方差解釋變異量	34
表 5-6	主成份分析模組之解說總變異量	36
表 5-7	民間消費綜合指標 (PCA) 平方差解釋變異量	36
表 5-8	DFM 操作要素說明.....	38
表 5-9	變數說明	39
表 5-10	單根檢定	40
表 5-11	變數敘述統計	41
表 5-12	相關係數矩陣	41
表 5-13	ARIMA 最適落後期分析	42
表 5-14	ARIMA 分析結果	43
表 5-15	樣本內靜態及動態預測績效比較	44
表 5-16	ARIMA 樣本內預測 DM Test.....	46
表 5-17	樣本外靜態及動態預測績效比較	46
表 5-18	ARIMA 樣本外預測 DM Test.....	48
表 5-19	VAR 落後期設定篩選表	48
表 5-20	VAR-傳統模型	49
表 5-21	VAR-GT 模型	49
表 5-22	VAR 模型樣本外預測績效比較	50
表 5-23	變數說明	52
表 5-24	單根檢定	53
表 5-25	變數敘述統計	53
表 5-26	相關係數矩陣	54
表 5-27	民間消費綜合指標 ARIMA 模型分析	55
表 5-28	樣本內靜態及動態預測績效比較	56
表 5-29	民間消費綜合指標 ARIMA 樣本內預測 DM Test.....	57
表 5-30	樣本外靜態及動態預測績效比較	59
表 5-31	民間消費綜合指標 ARIMA 樣本外預測 DM Test.....	59

表 5-32	VAR 落後期設定	60
表 5-33	VAR-PCA 模型	60
表 5-34	VAR-PCATR 模型	61
表 5-35	VAR-DFM 模型	61
表 5-36	VAR-DFMTR 模型	62
表 5-37	VAR 模型樣本外預測績效比較	63
表 5-38	VAR 樣本外預測 DM Test	63
表 5-39	關鍵社群網站大數據觀測站	65
表 5-40	博客來-Google Trends 主成份分析與 FB 集群網絡圖之商品相關性分析	73
表 5-41	七家網購的粉絲專頁-聲量圖之綜合結果	74
表 5-42	七家網購的粉絲專頁-情緒指標圖之綜合結果	74
表 5-43	七家網購的粉絲專頁-盒鬚圖與分配圖之綜合結果-1	75
表 5-44	七家網購的粉絲專頁-盒鬚圖與分配圖之綜合結果-2	75
表 5-45	七家網購的粉絲專頁-相關性圖之綜合結果	76
表 5-46	七家網購的粉絲專頁-樹狀圖之綜合結果	76
表 5-47	七家網購的粉絲專頁-集群網絡圖之綜合結果	77
表 5-48	七家網購的粉絲專頁-Google Trends 主成份分析與 FB 集群網絡圖之商品 相關性分析表之綜合結果	77

摘 要

在網路世代，民眾食、衣、住、行皆離不開網路，而使用網路的行為已累積可觀的資訊數量。近年有相當多的研究，運用全球使用率最高的 Google Trends 系統，作為預測民間消費、房市、股市等各面向的基準資訊，主要在於其具有相對即時性，透過週資料的分析，比過去月或季頻率總體經濟指標更能掌握市場變動。

本研究透過主成份及動態因子模型，萃取出網路檢索指標及民間消費綜合指標，並採用時間序列預測分析模組 ARIMA 與 VAR，探討納入網路大數據資訊是否能提升民間消費預測能力。

其中，檢索指標的建構乃透過國內 15 大電商網站內商品分類項目，作為檢索指標的建構內涵，彙總 1,838 類商品檢索項目之時間序列資料，再依照各電商網個別進行成份萃取，建置各電商的商品檢索數列指標。而民間消費綜合指標則是以總體經濟領先指標、消費者信心指數及網路檢索指標萃取而成。

研究結果發現，在 ARIMA 模型中，分別納入網路檢索指標及民間消費預測綜合指標進行評估，並與簡單 ARIMA 模型及 ARIMAX 傳統模型相比，納入兩指標之樣本內預測（動態或靜態預測）均具有較佳的預測能力；惟在樣本外預測模組中，各模型間並無顯著差異。此外，本研究亦透過 VAR 進行預測能力比較，然而納入兩指標之樣本外預測亦與傳統 VAR 模型無顯著差異。

整體而言，在樣本內預測模組中，考慮網路檢索指標、民間消費預測綜合指標之民間消費預測模型在模型配適度上較佳，惟受限於檢索資料起始自 2004 年，樣本數仍有限。未來在資訊不斷累積下，希冀能透過多樣態模型持續檢視及開發檢索指標在民間消費預測的效益。

第一章 研究背景與目的

民間消費為 GDP 支出面的主要構成項目，長期以來佔比多有 50% 以上，為國家經濟發展的重要引擎，維持整體經濟穩健成長，在預測研究上，具有其重要的意義。

過去在民間消費的預測研究上，多以經濟、貨幣性的指標做為預測的基準，如銀行放款利率、可支配所得、能源價格、消費者信心指數及景氣指標等，並透過各式計量模型去進行設算。但在網路發展下，具有即時性的網路型態的指標也漸漸受到關注，進而被做為預測指標的一環，並應用在消費、失業預測，並做為民間消費、國家經濟的預測分析來源之一。

Matsumoto, Matsumura, and Shiraki (2013) 指出在經濟預測的議題上，過去的問題是我們是以延遲性的相關指標，去對未來進行預測，時間的落差問題會造成預測的效度受到影響，而為了改善這項問題，臨近預測 (Nowcasting) 則是未來的發展趨勢。臨近預測是指透過延遲性低的指標來做預測基準，而網路搜尋資訊 (Internet search data) 則是目前最具潛力的預測研究指標，而 Matsumoto, Matsumura, and Shiraki (2013) 亦用此來對國家經濟進行預測。

Vosen, and Schmidt (2012) 則是以 Google 搜尋引擎的大數據來對德國民間消費進行預測，發現運用此類新型態的指標來預測，較傳統預測指標更據準確性。Askatas, and Zimmermann (2009) 認為網路資訊的即時性，在做為預測上，會較以往指標更具即時性，並以民眾搜尋資訊去預測德國失業率。

近幾年除藉由網際網路的蓬勃發展，也拜智慧型通訊設備普及率大幅提升，而使得網路使用群眾數目得以倍增，也進入了所謂 Open Data 時代。市場調查機構 Millward Brown 在 2014 年針對 37 個國家進行調查，各國民眾每日上網使用分鐘數，以我國為最多。

表 1-1 2014 年各國民眾每日上網使用分鐘數

排名	國家	分鐘數	排名	國家	分鐘數
1	台灣	197	14	香港	162
2	奈及利亞	193	17	美國	151
3	沙烏地阿拉伯	189	20	南韓	144
7	中國	170	22	日本	135

資料來源：Millward Brown（2014），網址：<https://goo.gl/ZYi3qt>。

而依據資策會 2016 年的統計，台灣人平均每天滑手機 3.28 小時，臉書活躍用戶占總人口 74%，兩者皆為全球之冠，顯示我國在網路生活的普遍度。

搭配消費者的使用習慣，我們可以發現在網路世代，注意力（attention）的產生、會帶動其搜尋力，並進而對購買力產生可能影響。此外，不可否認地，生活在資訊網路時代的現代人改變了過去的互動模式，取而代之的是透過網際網路相互傳輸的自然語言巨量訊息，所以民眾討論議題的方式，已逐漸從通訊設備、面對面交談漸轉變於網際網路上集思廣益並討論。也因此趨勢使各大討論區、部落格、社群網站數量也不斷攀升，而此現象進而連帶影響了民眾的行為模式，從以往的單方面的資訊接收者轉變為得以查詢本身想要之資訊並也可於討論區或自身的社群媒體界面上發表自身的評論及看法，影響力遍及各式食衣住行育樂。



資料來源：本研究。

圖 1-1 消費者網路行動模式

目前根據統計，谷歌（Google）公司每天得處理超過 24PB 的資料¹，是美國國會圖書館所有紙本資料量的數千倍，而臉書現在使用者每小時會上傳超過一千萬張新照片，每天所有使用者會按讚

¹ 1petabyte 大約等於 1,000 terabyte。

或留言超過三十億次。這種爆炸性的資訊量被稱作「巨量資料（大數據）」（Big data），如何利用與分析「巨量資料（大數據）」近年來被科學界或網路界大量且廣泛的討論（林俊宏譯，Mayer-Schonberger, and Cukier，2013）。

目前網路多數社群媒體中，台灣民眾使用頻率最高為臉書、及 PTT 批踢踢實業坊，在這幾個平台上，每天有無數的使用者除了分享自己本身的近況、看法、評論，也能張貼圖片或標註自己位於何處，而其他使用者也可對該則貼文做出回應，除按「讚」、「推」表示認同外，也可留下自己對於該則貼文的看法。

除了個人用戶外，公眾人物、營業廠商、上市公司、非營利組織等各式組織也紛紛架設起屬於本身的臉書頁面-粉絲頁面。希望能藉由臉書的高普度使用率，透過與支持者互動或發佈現況進而提升自身的知名度。

換句話說，臉書、PTT 等社群網站將所有使用者的想法、評論、觀點都匯集於該平台。在此類型平台產生之前當需要獲得大量的評論資料時，則須透過較繁瑣的問卷、深度訪談等方式。我們可以藉由文字探勘的技術，能夠透過電腦的運算能力，過濾及轉化大量的文字內容，讓人能夠更有效率的運用，這些資料將會是研究各種社會、經濟議題的最佳資料來源。

這也代表如果我們有辦法將這些社群網站上的評論加以分析及量化並將之轉為可運用的參考數據，或可探討其評論中的關鍵字與使用者行為，而各方評論、社會氛圍，也可做為政府發展公共政策方向的參考指標，更可做為民間消費熱度的預測指標之一。故此，本研究的主要研究方向為：

在傳統預測民間消費的相關經濟指標之外，建置運用網路檢索及社群媒體民間消費輿情時間數列資料，做為民間消費預測綜合分析指標，並透過計量模型驗證其預測效益，本研究計畫之目的如下：

1. 彙整有助預測我國民間消費之經濟及領先指標資訊，包括見諸主計總處、國發會之相關資訊，比對相關統計定義與內涵，作為總體經濟面向之重要指標。
2. 建立民間消費網路搜尋大數據指標，以消費注意力為基礎，

透過網路檢索資訊蒐集系統 Google Trends，彙整我國民間消費大數據消費指標。

3. 建立民間消費輿情大數據指標，探索社群網絡新興資訊對於我國民間消費之相關影響。
4. 透過考量網路大數據之「民間消費預測綜合指標」之實證結果，提供建置臺灣民間消費預測資料庫之基礎與建議，並可做為後續研究之架構參考。

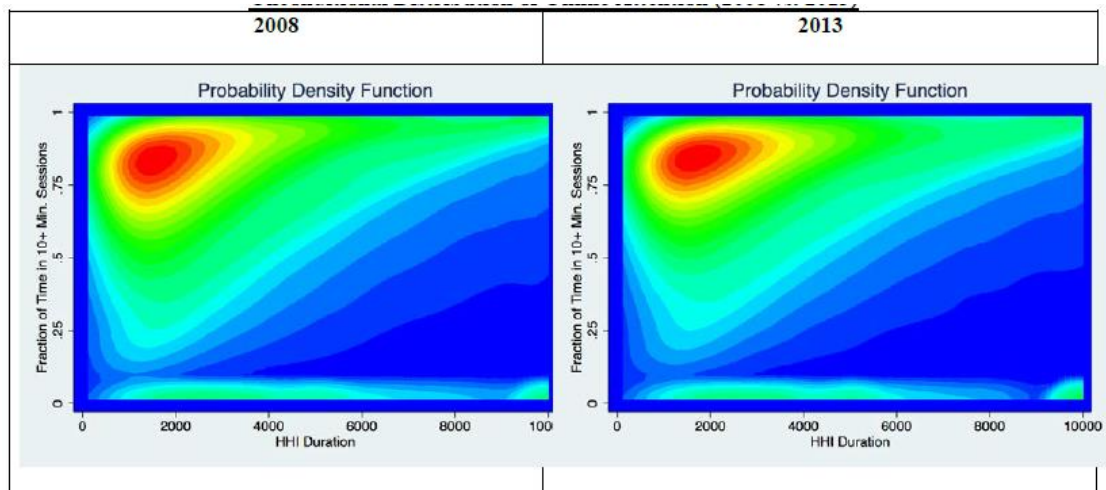
第二章 文獻回顧

諾貝爾經濟學獎得主 Herbert Simon (1971) 即提到，資訊的富有將會帶出注意力的貧窮，如何在過度豐富的資訊來源中，將注意力做有效的配置，才是重點。Davenport and Gilbert (2002) 也提及，在注意力經濟中，注意力不見得是一種「付出」，而是像貸款一樣需要借用，像投資一樣需要管理，才能學會忽略不必要的事，把注意力擺在最重要的事情上。

網路世代的來臨，促使人們隨身的裝置有所變化，且媒體服務應用之成長，在 2005 年 2 月成立的 YouTube，2006 年 10 月時即被搜尋引擎 Google 購買。而同屬於影片觀賞供應的 Netflix 和 Hulu 兩者同時在 2008 年上線提供媒體服務，在當時，我們目前最熱門的社交媒體還未見今日榮景。如 2006 年 3 月推出 Twitter，2004 年 2 月推出臉書，在 2006 年 9 月時才被廣泛的討論。而 2013 年，社群媒體已經成為一種主流的線上工具，鑲嵌入生活之中，配合個人喜好，顯示專屬資訊，此也顯示人們的各項資料被廣泛應用在其中。

第一節 搜尋大數據預測應用研究

Boik, Greenstein and Prince (2016) 指出，現今在許多競爭市場中，廠商並非直接爭奪消費支出，而是關注消費者的注意力。其透過建模，分析家庭如何利用網路工具去對其感興趣的市場表達關注，進而分配注意力。而家庭部門的注意力配置決策可分為三個主要方向：分配數量、配置在哪、如何配置，此篇文章觀察 2008 至 2013 年的點擊流量資料，家庭使用網路搜尋的時間有明顯大幅度增加，從聊天、視頻、社群媒體中，看到廣大的數量及預期變化知道家庭部門消費狀況，當家庭所得越高、注意力彈性越大，在不同網路截點上分散的穩定度相當高，在網站中的注意力密度也相當穩定（見圖 2-1）。

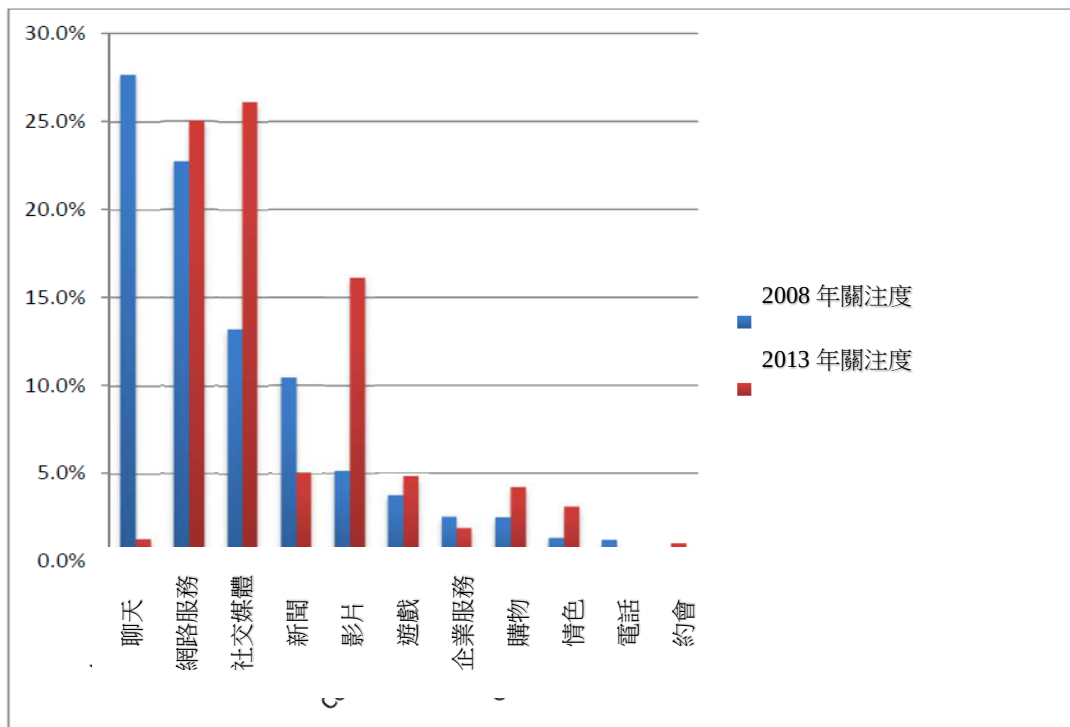


註：圖形呈現兩年注意力分佈變化，橫軸為赫芬達爾—赫希曼指數(HHI)，縱軸為聚焦時間。

資料來源: Boik, Greenstein and Prince (2016)。

圖 2-1 美國 2008 年與 2013 年家計網路關注度

Boik Greenstein, and Prince (2016) 也指出雖然有些研究指出收入和其他經濟變數會影響人們在網路的總時間。但是在人口型態的轉變，人口統計資料中能預測的廣度及深度變化趨向較少，且人們的收入不會因為家庭教育、家庭人數、年齡、是否有小孩而受影響，也相應到網路使用與關注度（注意力）的投入。而在整體趨勢上，搭配各種型態的網路裝置軟體發展，相較 2008 年，民眾配置較多的關注度在網路服務、社群網路軟體上（見圖 2-2）。



資料來源: Boik, Greenstein and Prince (2016)。

圖 2-2 民眾網路事務關注度變化

目前許多研究透過現有搜尋引擎、社群網站等各種公開場域，所累積之大數據資訊來對各項主題進行研究。Choi and Varian(2012)利用 Google 搜尋引擎的資訊去預測經濟指標的鄰近值 (near-term value)，如汽車銷售、失業救助領取、旅遊目標的規劃及消費者信心等各項議題。由於政府部門雖會定期釋出經濟指標，但通常統計資訊會有時效落後的問題，故對於預測這些經濟指標的功能是具有需求性的，相關工具如 Google Trends、UPS、Mastercard 等可產生即時日資料的工具，則如待開採的重要資料金礦，正如 Da, Engelberg, and Gao (2011) 所點出的重點，「搜尋就會帶來注意」。

以汽車銷售為例，Choi and Varian (2012) 利用 Google 搜尋關鍵字大數據資訊，如休旅車、汽車保險、卡車等搜尋次數來對汽車銷售做影響評估。其使用季節 AR (1) 模型，估計期間為 2004 年 1 月至 2011 年 7 月，結果發現各關鍵字多顯著影響汽車銷售。而將此法套用在失業救助領取、旅遊目標及消費者信心等議題，皆較傳統模型提升解釋力 (判定係數皆上升)。整體而言，考慮 Google 關鍵字進入模型後，表現都較簡單季節 AR 模型提升 5-20% 預測力。

Herzog (2014) 則以 Google 搜尋資料來分析對公司資產價格的影響，以 2004 至 2010 年，以 30 個國際銀行為標的去建立時間數列模型。有兩項主要發現，第一，就 Google 搜尋量及 Google 搜尋的點擊為影響評估要素，結果發現公司資產的價格會與 Google 搜尋量的變動率為正向關係，而 Google 搜尋的點擊會與公司股票交易量為正向。第二，有關資產的動態變化關係，Google 搜尋量及 Google 搜尋的點擊有不同的影響效果。

Da, Engelberg and Gao (2011) 探討搜尋的注意力，建立搜尋量指數 (Search Volume Index, SVI)，此指標可直接衡量民眾在搜尋引擎上的注意力，而為了判斷其效益，其以 3000 支現有股票，在 2004 至 2008 年間的搜尋情況，與現有投資關注指標 (如 IPO 指數、或恐慌指數 VIX) 進行比較，有三個主要發現。第一，SVI 與現有指標存不同的相關性；第二，使用 SVI 去捕捉投資人注意力，是更具有時效性的；第三，對於散戶的衡量，特別具有效果，如 SVI 提高的話，股價未來兩周將會較高，而未來一年有反轉的可能。而 SVI 提高的話，也會使得第一天 IPO 報酬率較高，但長期股價報酬會相對較低。Hamid and Heiden (2015) 亦以搜尋關注力去建置週指標，來去預測投資人在股市的各項行為。

Toth and Hajdu (2012) 同樣以 Google Trends 搜尋數列建立相關指標來對匈牙利的房市、車市消費進行預測。其中，Toth and Hajdu (2012) 也在驗證，是否 Google Trends 搜尋數列只適用於互聯網技術及規模較大的歐美國家，而在網路普及度相對沒這麼高的國家是否仍適用，結果發現，Google Trends 檢索指標在預測匈牙利的房市及車市上，仍有不錯表現，相對於過去僅能使用落後資訊的經濟性指標而言，有更好的臨近預測力。

Pavlicek (2014) 則點出 Google 計量學的議題，透過工作相關的 Google 關鍵字搜尋來評估對失業率的影響，其以自迴歸模型 (Autoregressive model, AR) 及向量自我迴歸 (Vector Autoregression, VAR) 兩種模型去進行分析。結果發現以捷克為例，用 Google 搜尋的關鍵字，來做解釋要素，在 VAR 的操作上，透過因果分析、衝擊反應以及預測誤差拆解，研究結果發現部份支持假說 Google Trends 資料具有影響效果，而也提出當將 Google 搜尋資料納入分析模型後會讓模型預測力較傳統變數模型更為準確。

第二節 網路大數據民間消費預測應用研究

Matsumoto, Matsumura and Shiraki (2013)以日本大地震為例，就搜尋資料的即時性來看，日本官方資料的新車註冊資訊落後 3 天出現、經濟觀測調查延後 20 天後，食品服務銷售產業、商業及工業生產調查延後 30 天，服務業的調查、旅遊公司的營收調查延後 60 天，季 GDP 估計調查延後 70 天，各式官方資訊皆是明顯具有落後性的。為驗證搜尋關注力指標的代表性，其針對旅遊進行研究。並將旅遊分為 11 大類，包含特色旅遊、旅遊目的地、旅行紀錄、旅遊辦事處等等，並在 11 大類下又細分為 21 個搜尋數列(詳見表 2-1)，經 Google Trends 搜尋數列的建立後，資料期間為 2004 年 1 月至 2011 年 2 月，並利用主成份分析分為 3 因子，並經季節調整後進行預測。

表 2-1 旅遊主題 Google Trends 搜尋關鍵字

主題	大項	細項
旅行	特色旅遊	農業旅遊
		生態旅遊
		葡萄園 & 酒 旅遊
		觀光旅遊
		冒險活動
	公車 & 鐵路	
	酒店 & 住宿	
	租車 & 計程車服務	
	旅遊目的地	主題公園
		海灘 & 小島
		湖 & 河
		區域公園 & 花園
		山 & 天空 渡假村
		歷史遺跡 & 建築物
		動物園-水族館
		博物館
	空中旅行	康樂航空
		飛機場 & 運輸
	遊輪 & 特許	
	拼車 & 座位共用	
	旅行紀錄 & 遊記	

主題	大項	細項
	行李 & 旅行配件	
	旅遊辦事處 & 服務	假期優惠
		旅遊局 & 遊客中心

資料來源：Matsumoto, Matsumura and Shiraki (2013)

實證結果發現，第一主成份因子對旅遊支出整體解釋力 0.08，個別係數非常顯著的正向影響，係數值 0.61，相當顯著，而加入調查資料後，整體預測力也較單獨使用調查經濟預測問卷解釋力為高，顯示 Google Trends 指標的代表性。

而 Vosen and Schmidt (2011) 更是直接透過 Google Trends 來對民間消費建立搜尋聲量的時間序列研究。以 Google Trends 做為新的指標，相對於密西根大學的消費者情緒指數 (Consumer Sentiment Index) 以及 The Conference Board 的消費者信心指數 (Consumer Confidence Index)，在樣本內及樣本外的預測更具優勢，故其建議在對民間消費進行預測時應該將 Google Trends 的資訊納入解釋變數之一。而文內也整理美國經濟分析局 (Bureau of Economic Analysis, BEA) 與 Google 分類中可對應到的項目如下表 2-2：

表 2-2 BEA 與 Google 品項可對照之分類

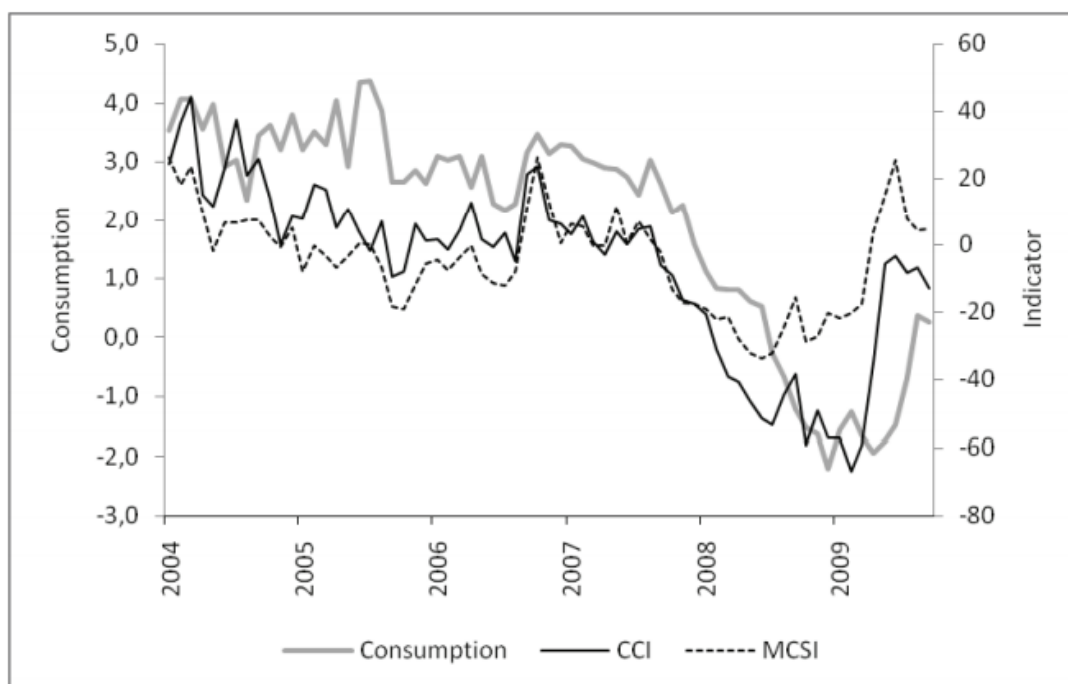
類別	大項	細目
耐久財	汽車與汽車零件	汽車、汽車零件、汽車融資、汽車保險、汽車品牌、購物車
	傢俱與耐久設備	電腦與電子、消費類電子產品、家電、房貸、家居擺設、家庭與花園、家裝、家庭保險、家政與室內裝飾、室內設計
	娛樂商品與車輛	圖書零售商、娛樂、娛樂產業、電影、電玩遊戲
	其他耐久商品	圖書零售商、移動與無線、電信
非耐久財	買食品和飲料的前提消費	含酒精飲料、食物與飲料、食品零售商、非酒精性飲料
	服裝與鞋類	服裝、服裝標籤與設計、服裝零售商、鞋類、內衣、T 恤
	汽油與能源產品	電力、能源與公共事業、石油與天然氣
	其他非耐久財商品	美容與個人護理、化學品、藥物、臉與身體的照顧、頭髮護理與產品、健康、報紙、菸草產品

類別	大項	細目
服務	家庭消費支出	
	住房與公共設施	房貸、家庭改善、家庭保險、家政與家裝、室內設計、房地產、房地產代理商
	衛生保健	藥物、健康、健康保險、醫療設施與服務、移動與無線
	運輸服務	汽車融資、汽車保險
	娛樂服務	娛樂、娛樂產業、電影、門票銷售、電玩遊戲
	食品服務與住宿	食物與飲料、食品零售商、酒店與住宿、餐廳供應、餐廳
	金融服務與保險	金融與保險、房貸、家庭保險、保險
	其他服務	娛樂與退休金、社會服務、電信、廢棄物管理

資料來源：整理自 Vosen and Schmidt （2011）。

Vosen and Schmidt （2011）指出民間消費在美國總支出中約占 7 成，所以有關於即時資訊、民間消費支出都是重要的依據，但是官方資料公布的時間為月資料，為落後一個月的資訊。其中，領先指標、高頻資料有助於預測未來，還可以呈現當下的情況，因此高頻資料的公布，是特別有用的，當總體經濟遭受很大的不利、不確定性、重大事件時，過去的資料沒辦法有效的去解決。

傳統都是以問卷為基礎的情緒指標，這些指標考量經濟、心理面向，實證的文獻大致認為，美國的消費者情緒指標跟消費行為，有很強之關係性，透過 The Conference Board 之消費者信心指數、及密西根消費者情緒指數，發現兩者之間有高度正相關(詳見圖 2-3)，會隨時間有共同的變化方向，在預測民間消費時，總體經濟指標代表他們的消費情形，問卷為基礎的指標為消費者的購買意願，Google Trends 指標則提供衡量消費者預備下一步購買的前階準備。



資料來源： Vosen and Schmidt （2011）。

圖 2-3 民間消費與調查指標線性趨勢

Vosen and Schmidt （2011）其在預測上的設計，利用簡單自迴歸模型，並以民間消費成長率作為標竿模型，後續則將 MCSI、CCI、Google Trends 加入標竿模型中，並加入三總體指標：（1）實質個人所得（2）三個月公債殖利率（3）股價，以 S&P500 為代表。實證結果呈現出 Google Trends 為有利的資訊來源，不管是在樣本內、樣本外的預測，都比傳統的問卷基礎指標好，建議可以使用 2 年多的資料，Google Trends 也應該作季節性調整，有助於後面做實證分析。

Kholodilin, Podstawski, and Siliverstovs （2010）以 Google Trends 之搜尋指標來評估其對美國民間消費臨近預測之效益，在模型評估上，其納入一般性總體經濟變數，金融相關及網路檢索引指標。在檢索引指標建立上，其利用消費品項大項目搜尋，進一步讓 Google 帶出之前 10 大搜尋品項來建立之。研究結果發現，納入檢索引指標的民間消費預測模型，在臨近預測上，相較於過去更具有預測效度。

Schmidt and Vosen （2012）以 Google 搜尋引擎的大數據來對德國民間消費進行預測，發現運用此類新型態的指標來預測，較傳統預測指標更具準確性。其建立月頻率的消費指標，並使用自迴歸模

型去進行預測。Askatas and Zimmermann (2009) 認為網路資訊的即時性，在做為預測上，會較以往指標更具即時性，並以民眾搜尋資訊去預測德國失業率，其中則以時間序列因果分析去進行驗證。Mclaren and Shanbhogue (2011) 則同樣以搜尋資料來對英國的失業率及房價資料進行預測。

劉欣姿 (2012) 也利用 Google Trends 搜尋指標，對我國消費模型進行預測推估，其以商業營業額為消費變數，並以家庭消費類別為搜尋依據，篩選與消費變數相關性較高的因子，最後則透過主成份分析萃取其因子特性。而在預測模型中，也納入總體經濟指標，如消費者信心指標、勞動市場、金融面因子等變項，研究結果發現，以搜尋指標作為估計消費的高頻率指標對推估具有助益，惟搜尋指標有期間較短的問題存在。

第三節 社群網路大數據相關研究

現今民眾的網路使用習慣，除了搜尋引擎之外，在社群媒體上所花費的時間比例更是相當明顯。社群媒體的功能型態也逐漸擴張，現今也有許多電子商務是透過臉書等社群媒體去經營，各家廠商為了更加抓緊消費者，紛紛也透過臉書經營搭配其電子商務系統。Kantrowitz (2015) 指出，電子商務直接導入臉書則是未來的趨勢。有了網路、社群之後，消費者的生活型態、需求，以及購物旅程有很大的轉變，人們開始受到口碑的影響，包括親友口耳相傳、網路上的評價，作為消費決策判斷（動腦新聞，2015）。

在社群網路的分析上，由於其資訊是依據消費者在社群網站將所有使用者的想法、評論、觀點各項資訊的匯聚。在此類型平台產生之前當需要獲得大量的評論資料時，則須透過較繁瑣的問卷、深度訪談等方式。我們可以藉由文字探勘的技術，能夠透過電腦的運算能力，過濾及轉化大量的文字內容，讓人能夠更有效率的運用

Godfrey et al. (2014) 以 Twitter 為標的，來分析世界盃粉絲專業，其中用集群分析可以來看到資料的潛在型態。在操作上其蒐集在世界盃舉辦前的 3 萬筆推文，但實際上資料中會有詞藻噪音，其採用解層 (algorithm) 來加以處理，方法為 NMF (Non-Negative Matrix Factorization) 法，結果發現用 NMF 在找到文字結構上相當

有效，後續並對文字做集群分析，結果呈現如圖 2-4。



資料來源：Godfrey et al. (2014:10)

圖 2-4 詞藻集群分析範例

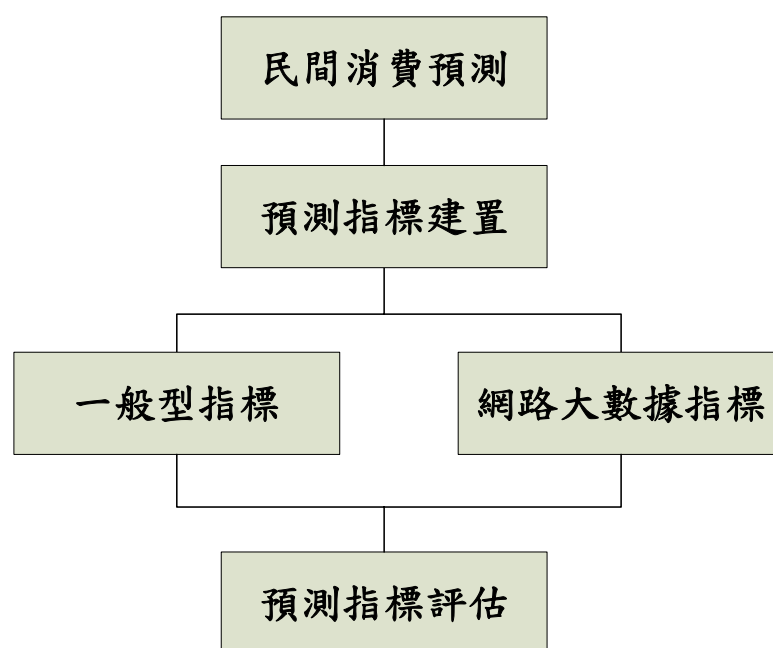
其中，可發現即便是民眾言論的詞藻，也會具有集群與相關性，故研究者可從中發掘各議題間的網絡關係。而社群網站將所有使用者想法、評論、觀點都匯集於該平台，較傳統的調查分析，更能貼近使用者實際的各項行為。

而在民間消費面向，配合各家電商網在社群網絡上的新興經營模式，粉絲專頁已成為其銷售平台之一，並可掌握民眾需求的第一手訊息，在社群網絡中的發言、點讚、轉 PO 各種行為模式，掌握口耳相傳的傳遞模式是相當值得開發研究，作為後續民間消費預測的重要指標。

第三章 研究架構與設計

本研究之目標，為在傳統預測民間消費的相關指標之外，建置運用網路檢索時間數列資料，做為民間消費預測綜合分析指標。在民間消費預測指標建置的部份，會就一般型、網路大數據指標各別建置，並透過計量模型分析方式（主成份分析）以及時間序列預測模型（VAR、ARIMA），去驗證新型態預測指標及各項指標組成的效果。

其中，一般型指標將參考過往民間消費預測研究去建構相關預測資訊，而在網路大數據指標的部份，會就搜尋引擎大數據進行分析，除此之外，本研究亦透過多變量分析方法觀察關鍵社群網站的網路大數據資訊。以下則就此二部份進行說明。



資料來源：本研究。

圖 3-1 研究架構圖

第一節 網路大數據指標

(一) 搜尋引擎民間消費指標

Google 引擎是目前廣泛受到國人使用的平台，民眾會利用此平台搜尋趨勢各種日常生活所需事務，食、衣、住、行、育、樂皆涵蓋其中，故在建立搜尋引擎指標的部份，本研究將以 Google 引擎做為主要觀測平台。

其中，將以攸關民間消費的搜尋議題做為指標建置的內涵，Vosen and Schmidt (2011) 即指出搜尋注意力可作為消費者購買前準備動作，故本研究在指標建構上本研究則以實務上品項購買類別作為檢索指標篩選關鍵字基準，並透過後續相關性分析等，確定指標的影響性，最後則進行主成份分析，將其做為搜尋引擎指標分析資訊來源。

在關鍵字的選擇上，考量貼近民眾消費搜尋字串，故依據電商數據平台 EagleEye 所公佈之 2014 年第 4 季「台灣購物網站排行榜」中前 50 大電商網，²本研究篩選出前幾名的指標性購物網站（詳見表 3-1），並納入好市多、Famipart 全家等電商平台，依據網站中的商品分類，作為關鍵字的來源依據。³在實務上，本研究整理出各網站中的分類項目，分為大類、中類、小類、品牌列表、注目商品，比較各電商網商品類的重複性作為後續關鍵字的選擇基礎。表 3-2 則為各電商網在各商品分類數上的分佈。

² EagleEye 在此次調查總計統計樣本為 15 萬熱愛網購族群，統計區間：2014 年 10 月至 12 月，統整出綜合購物網站的人次佔比排行榜。資料來源：<http://goo.gl/jLUOqt>。

³ Gomaji 團購網因品項資訊呈現並非依據商品層級列表呈現，而是依照廠商以單頁面方式呈現，須結合廠商與商品的判斷，故在此未將其納入參考電商網頁之一。

表 3-1 Google Trends 關鍵字來源網

名次	電商名稱	人次佔比%	名次	電商名稱	人次佔比%
1	Yahoo!奇摩購物中心	53.64%	10	樂天市場購物網	29.50%
2	PChome 線上購物	51.35%	11	GoHappy	23.61%
3	Yahoo!奇摩拍賣	50.42%	15	7-net 統一超商	14.30%
4	Yahoo!超級商城	48.63%	16	EHS 東森購物網	13.95%
5	露天拍賣	44.52%	23	myfone 購物	9.83%
6	momo 富邦購物網	38.89%	-	好市多	-
7	PChome 商店街	37.02%	-	Famiport 全家	-
8	博客來	33.68%	-	Yahoo 服務+	-
9	淘寶網	29.57%			

註：人次佔比%=(瀏覽該商品頁的不重覆人次/全部不重覆人次) x 100%

資料來源：EagleEye (2015)、本研究。

表 3-2 各電商網商品品項分類數

電商名稱	商品品項數		
	大類	中類	小類
Yahoo!奇摩購物中心	-	42	171
PChome 購物中心	-	8	141
Yahoo 拍賣	22	393	-
Yahoo 品牌旗艦分類		28	172
露天拍賣	24	475	-
momo 富邦購物網	-	19	98
Pchome 24h 購物	-	14	355
博客來	-	29	290
淘寶網	-	26	378
樂天市場	12	76	436
GoHappy	15	135	493
7-11 集團 ibon	8	62	85
EHS 東森購物網	14	198	137
Myfone 購物	-	38	246
好市多 Costco	15	118	279
FamiPort 全家	9	57	65
Yahoo 服務+	4	120	-

資料來源：本研究。

而在關鍵字設定與網路大數據指標蒐集上，主要可分為三步驟，結合前述關鍵字蒐集說明。考量在進行後續研究分析，關鍵字及資料蒐集過多將有影響分析效能，故以中類品項為關鍵字設定基礎。

在第二步驟，則將中類分類項目，作為搜尋引擎大數據指標的建立基礎，為週的資料頻率。第三步驟則配合社群網站大數據指標的建立。指標搜尋步驟請見表 3-3。

表 3-3 關鍵字設定與網路大數據指標蒐集步驟

關鍵字設定與後續操作	作法
第一步驟： 關鍵字設定	整理出各網站中的分類項目，分為大類、中類、小類、品牌列表、注目商品
第二步驟： 搜尋引擎大數據指標	將各電商網中類的分類項目，放入 Google Trends 搜尋趨勢。其中，數據資料為：2004 年至 2016 年，每週被人們搜尋的頻率次數，建置本研究網路大數據指標。
第三步驟： 社群網站大數據指標	將 Google Trends 搜尋趨勢中的關鍵字，放入本研究設計之爬文系統「fbCrawler」，在各購物網站中的粉絲專頁做搜尋，數據資料為：2014 年 12 月至 2016 年 7 月，每日人們在臉書版面中留下的紀錄，數據資料包含了文章的發表時間與更新時間、使用者 ID、訊息、按讚次數、分享次數、幾則留言、文章的連結。（搭配如下分析架構說明）

資料來源：本研究。

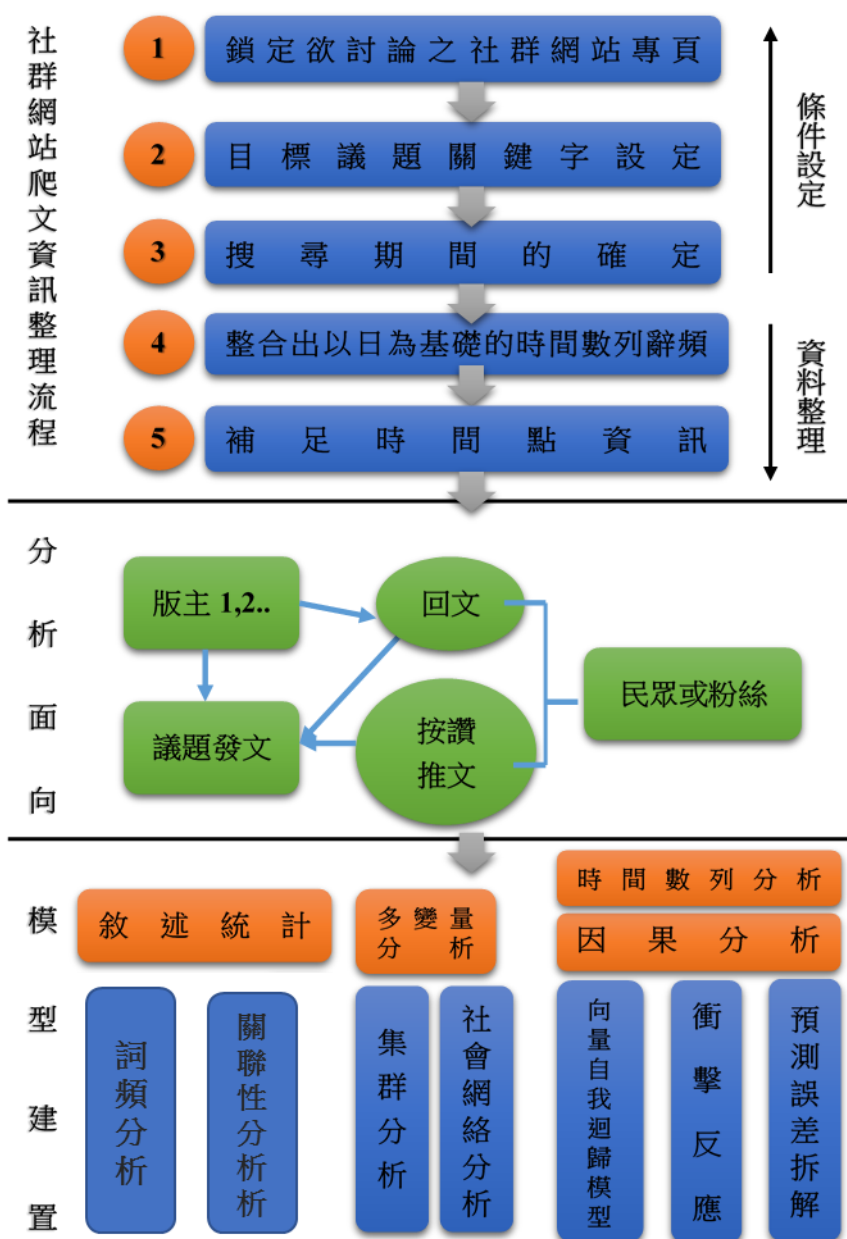
（二）社群網頁民間消費指標

在社群網頁指標的部份，在資料處理流程上，首先鎖定欲討論社群網站專頁（如各大購物或討論版面）；第二步驟，透過關鍵字的設定，如商品、消費等，作為搜尋專頁發文及回文的依據；第三步驟，則須設定所欲討論的期間。而為了後續的模型分析，在第四、五步驟，則需整合以日為基礎的時間數列詞頻，並補足未有發言之時間點。本研究為蒐集社群網頁民間消費資訊，特建置社群網路資料擷取平台，詳見附件四之平台簡介及操作說明。

而依據臉書社群網站的性質，分析面向可分為版主的議題發文，民眾的回文、按讚等面向，進行各種敘述統計及模型分析。在模型分析上，則可應用敘述統計、多變量分析、時間序列及因果分析，除可了解期間內主要民意討論重點、關聯性、集群性，並可了解關

鍵字間相互影響的情況，是否一議題容易帶動其他議題的討論，以及與民間消費的相關性，並可做後續的預測分析。而在操作細節上，針對議題觀測頁面，對其做正負向詞彙之聲量觀測，另會對不同社群網域平台進行相關性分析。最後，則會進行主成份分析，建置社群網頁民間消費指標。

在社群網路資料擷取上，本研究設計臉書資料擷取套件，其可針對社群網站內核心人物或關注版頁（各大消費臉書版頁等）於各時間點的各項發言、民眾的回文、按讚或推文數等各項內容資料進行系統性的蒐集，並轉成文字資料，以關鍵字進入資料庫，以利後續的研究分析。



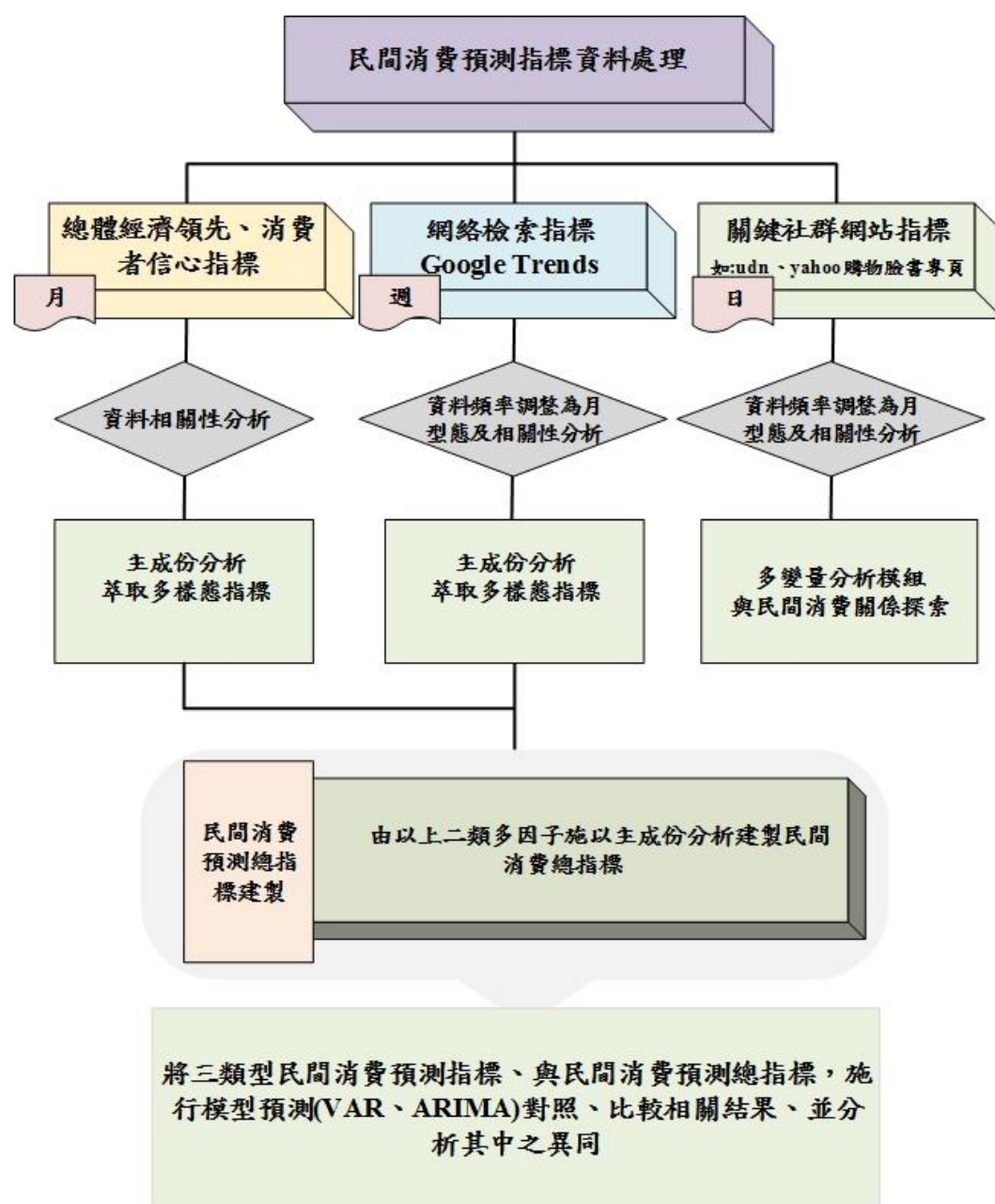
資料來源：本研究。

圖 3-2 研究架構圖

第二節 民間消費指標資料處理架構

於資料處理流程上（詳見圖 3-3），在三類型指標蒐集後，應各資料形態差異，首先施以資料頻率處理，將週資料的網路檢索指標與日週期的關鍵社群網站指標處理為月週期⁴。

⁴資料由週聲量轉月聲量或季聲量，是透過計量軟體 Eviews 進行，以總量法（量數）及平均法（指數）處理。



資料來源：本研究。

圖 3-3 民間消費預測指標模型資料處理架構

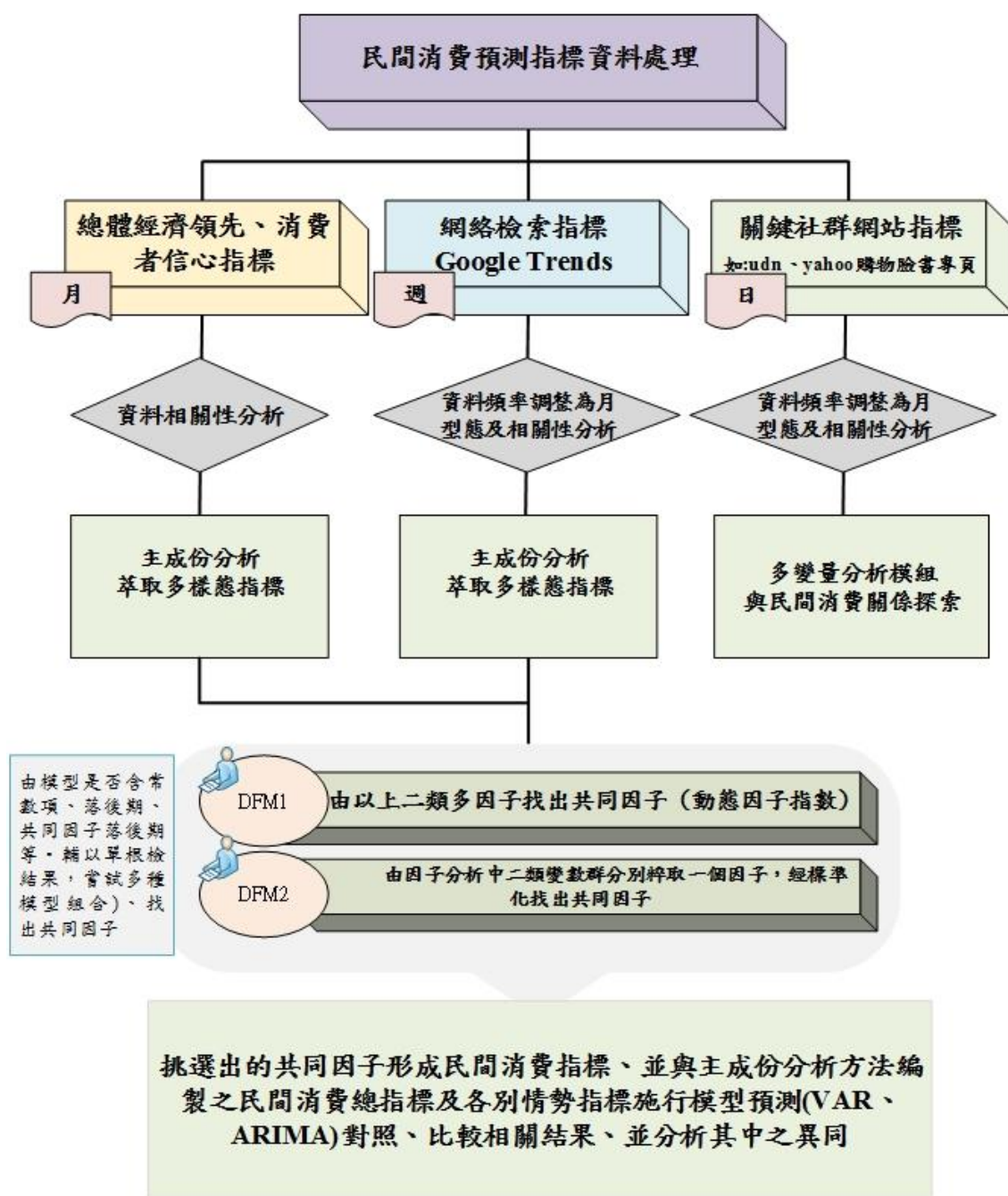
第二步，則就三類型指標各以主成份分析萃取多樣態指標，再進行民間消費預測指標的檢測下，本研究除了就三類指標各別驗證之外，並考慮民間消費總指標建製。然而，關鍵社群網站指標是透過關鍵臉書粉絲團網站資訊進行蒐集，但許多粉絲團臉書之經營為近三年左右始有發展。考量資料期間過短，將會造成模型衡量問題，故在關鍵社群網站指標則不納入預測模型之中，後續則會透過時間

序列相關分析以及網絡分析，探索社群網站指標與我國民間消費間之關係，期望作為後續民間消費預測之重要參考資訊。

最後則就總體經濟領先指標及網絡檢索指標，施以預測(VAR、ARIMA)模型對照，比較相關結果，找出具預測效度之指標。此外，由於主成份分析能將變數由繁至簡，濃縮精煉變數訊息，透過萃取臺灣重要民間消費走勢相關之重要因子，藉以觀察與掌握相關因子走勢以及據此編製之民間消費情勢指數。

惟考量臺灣自 21 世紀以來之經濟結構轉型，以及美國 Fed 編製之 LMCi 實證方法之運用特色，故而本研究擬於因素分析之外，加入動態因素考量，比較並對照主成份分析與加入動態因子後的動態因子模型 (DFM) 的異同，以更清楚了解臺灣民間消費市場情勢如何隨時間與外在衝擊而變化，操作模式請見圖 3-4。

動態因子模型 (DFM) 同樣使用前二類型各種頻率變數，施以頻率調整，並透過主成份分析實證操作同樣萃取出多樣因子後，並運用兩種方式找出共同因子，命名為動態民間消費指數。方法 A 為直接運用此多樣因子並經基準化 (反向基準化) 後找出共同因子，即 DFM1；方法 B 為先針對此二類因子所含之變數群，個別進行因素分析萃取一因子，再將此二類別因子基準化及反向基準化後，找出共同因子，即 DFM2。經過多種模型組合嘗試，本研究依據統計原則篩選最適模型，包括如 Log likelihood、模型估計結果、與靜態指數相關性及動態因子走勢等，選取最能反映民間消費市場情勢變化的模型。



資料來源：本研究。

圖 3-4 民間消費預測指標 DFM 模型資料處理架構

第四章 研究方法

由於本研究透過總體經濟指標、網路檢索指標及關鍵社群網站三種範疇建立民間消費預測指標，其中涵蓋時間數列資訊及多變量資料型態，故在分析上，則會透過主成份、時間數列分析、多變量等模組進行分析研究，以下為各項研究方法內涵說明。

壹、主成份分析

為建立綜合性民間消費預測指標，在一般型（傳統）指標、搜尋引擎民間消費指標等指標確立後，本研究將透過靜態及動態主成份分析，去建立民間消費預測綜合性指標。

有關動態主成份資料萃取與合成之方法即使用「動態因素模型」（Dynamic Factor Model，DFM）。此一計量方法近年被廣泛運用於經濟議題的分析上，例如 Bernanke, Boivin, and Eliasziw (2002)，以及 Bernanke and Boivin (2003)。透過 DFM 可以整合並提煉民間消費相關變數的訊息成為數個「主成份」因素。並選取兩個可以解釋 80% 民間消費變數變異的因素。然後用最大方差法（variance-max method）旋轉這兩個因素所得到的結果，組成民間消費情況指標。以下將針對動態因素模型加以介紹。

所謂動態因素模型，是指透過因素分析法（factor analysis），對大維度構面（large-scales）的樣本資料，進行變數構面刪減，並藉由主成份分析法，使研究者得在眾多原始樣本變數中，找出數個共同因素（factors），用以替代原始的龐大資料，並同時保有原始資料的樣本特徵，以下敘述引自 Stock and Watson (2005) 對動態因素模型的詳細模型設定與相關定義。

令 X_{it} 為一定態（stationary）且為可觀察（observed）的時間序列變數（ $t=1, \dots, T$ ）。動態因素模型指一藉由若干共同因素（common factor）與特徵干擾項捕捉的動態因素分析方法（詳見（1）式、（2）式）。

$$X_{it} = \tilde{\lambda}_i(L)f_t + u_{it}, \quad i=1, \dots, n, \quad (1)$$

$$u_{it} = \delta_i(L)u_{it-1} + v_{it}, \dots \dots \dots (2)$$

在（1）式當中， x_t 為可觀察的（observed），它能夠被 q 個向量維度在 $q \times 1$ 的「動態共同因素」（dynamic common factor） f_t 捕捉；而 $\tilde{\lambda}_i$ 表示

矩陣維度 $1 \times q$ 的落後期多項式，稱為「動態因素負荷」(dynamic factor loading)； u_{it} 為特徵干擾項 (idiosyncratic disturbance)， u_{it} 本身可能有序列相關性存在。

在動態因素模型中，共同因素 f_t 和為特徵干擾項 u_{it} 兩者被假設為互不相關，即 $E(f_t u_{is}) = 0$ ，而特徵干擾項 u_{it} 彼此間也被假設為互不相關，如 (3) 式所示：

$$E(u_{it} u_{is}) = 0 \quad \text{for all } i, j, t, s, i \neq j \dots\dots\dots (3)$$

符合上述相關定義稱為「完整動態因素模型」(exact dynamic factor model)，由 Geweke (1977)、Sargent and Sim (1987) 所發展之，其中 (3) 式表示特徵干擾項被假設為強烈無關性 (strong uncorrelatedness)。Chamberlain and Rothschild (1983) 發展了一個可以有效區別「完整動態因素模型」與「近似動態因素模型」(approximate dynamic factor model) 的方法。所謂「近似動態因素模型」，它允許特徵干擾項得在不同期間為有限相關性 (Limited amount of correlation across the idiosyncratic terms for different i and j)。

為解決動態因素模型特徵誤差項 (idiosyncratic errors) 序列相關性問題，將 (1) 式等式兩邊同乘上 $1 - \delta_i(L)L$ ，可推導出 (4) 式：

$$X_{it} = \lambda_i(L) f_t + \delta_i(L) X_{it-1} + v_t \dots\dots\dots (4)$$

其中 $\lambda_i(L) = (1 - \delta_i(L)L) \tilde{\lambda}_i(L)$ ；進一步推導後，完整的動態因素模型可以下列形 (5) 式、(6) 式表示之：

$$X_{it} = \lambda_i(L) f_t + D(L) X_{it-1} + v_t \dots\dots\dots (5)$$

$$f_t = \Gamma(L) f_{t-1} + \eta_t \dots\dots\dots (6)$$

$$\lambda(L) = \begin{bmatrix} \lambda_1(L) \\ \vdots \\ \lambda_n(L) \end{bmatrix}, D(L) = \begin{bmatrix} \delta_1(L) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_n(L) \end{bmatrix}, v_t = \begin{bmatrix} v_{1t} \\ \vdots \\ v_{nt} \end{bmatrix} \dots\dots\dots (7)$$

$\Gamma(L)$ 為一包含落後期多項式矩陣， η_t 為 $q \times 1$ 的干擾項向量，且在不同時點下 $E(\eta_t v_{is})$ 皆為 0。

根據 Stock and Watson (2002) 指出，在動態因素模型的靜態形式中有 r 個靜態因素 (F_t)，它包含當期共同因素 F_t (亦可能包含 q 個動態因素的落後期)。假設 (5) 式 $\lambda(L)$ 為有限落後期 $p-1$ ，且令 $F_t = \begin{bmatrix} f_t' & f_{t-1}' & \cdots & f_{t-p+1}' \end{bmatrix}$

(其中 p 為動態因素的落後期， F_t 維度是 r ，且 $q \leq r \leq qp$)，則 (5)、(6) 可改寫為：

$$X_t = \Lambda F_t + D(L)X_{t-1} + v_t \dots\dots\dots (8)$$

$$F_t = \Phi(L)F_{t-1} + G\eta_t \dots\dots\dots (9)$$

其中 Λ 為 $n \times r$ 的矩陣， Λ 矩陣中的第 i 列包含 (7) 式 $\lambda_i(L)$ 的係數， $\Phi(L)$ 包含 (6) 式 $\Gamma(L)$ 的係數， G 為 $r \times q$ 的矩陣，如果 $\Gamma(L)$ 的階次 (order) 為 p 且 VAR 模型中的 F_t 不含落後期，即 $\Phi(L) = \Phi$ ，則 (8) 式、(9) 式被稱為動態因素模型的靜態形式 (即 F_t 在 (9) 式中不包含落後期)。

在靜態因素變異拆解部份，令 $\tilde{X}_{it} = (1 - \delta_i(L)L)X_{it}$ ，且 (靜態) 因素負荷 (factor loading) 矩陣為 $\Lambda = [\Lambda'_1 \dots \Lambda'_n]'$ ，而 $\tilde{X}_t = [\tilde{X}_{1t} \dots \tilde{X}_{nt}]$ ， $v_t = [v_{1t} \dots v_{nt}]'$ ， $\Sigma_{\tilde{X}}$ 、 Σ_F 、 Σ_v 為 $\tilde{X}_t$ 、 F_t 、 v_t 的共變異數矩陣，則一般靜態因素分析的變異。

將上述動態因素模型中的 (9) 式代入 (8) 式可推導出式 (10) 式：

$$X_{it} = \Lambda_i \Phi(L)F_{t-1} + \delta_i(L)X_{it-1} + \varepsilon_{X_{it}} \dots\dots\dots (10)$$

靜態因素 F_t 可利用主成份分析法對經標準化 (standardized) 後的 X_t 進行估計，在取得共同因素 F_t 與因素負荷 Λ 之後，以因素標準特徵值大於 1，或累積解釋變異達 70% 以上，來判斷最適因素個數。

貳、時間序列預測模型

1. 自我迴歸整合移動平均 (Autoregressive Integrated Moving Average,

ARIMA) 模型

自迴歸模型 AR (p)，為 p 階自迴歸模型，其方程式為：

$$u_t = c + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + \varepsilon_t, \quad t = 1, 2, \dots, T \dots (11)$$

其中 c 為常數； ψ 為自迴歸模型係數； P 為自迴歸模型階數； ε_t 為白噪音項。

移動平均模型 MA (q)，為 q 階移動平均模型，其方程式為：

$$u_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad t = 1, 2, \dots, T \dots\dots\dots (12)$$

其中 μ 為常數； $\theta_1, \theta_2, \dots, \theta_q$ 是 q 階移動平均係數。

移動平均模型 ARMA (p,q) 模型包括自迴歸模型 AR (p) 及移動平均模型 MA (q)。

$$u_t = c + \phi_1 u_{t-1} + \dots + \phi_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad t = 1, 2, \dots, T. \quad (13)$$

而自我迴歸整合移動平均 (ARIMA) 則是透過 d 階差分的 ARMA 模型。估計 ARIMA 模型同估計 ARMA 步驟相同，惟一不同，是在估計前需確定原序列差分階數 d ，對目標數列 y_t 進行 d 階差分，因此 ARIMA (p,d,q) 與 ARMA (p,q) 模型區別之處在於前者的自迴歸特性含有單位根的處理，差分的處理有助於將非平穩序列轉化為平穩序列。

參、向量自我迴歸模型

70 年代在凱因斯理論的指導下，大型的總體計量模型 (large-scale macroeconomic model) 被廣為運用，動輒百餘條的方程式充斥於模型，但是隨著總體經濟環境日趨複雜，大型的總體計量模型的績效表現越來越差，因為以往傳統計量分析方法中，通常會根據先驗理論建構出結構化的計量模型，再透過迴歸分析法求得模型中母數的估計值。但此種方法在結構模型的建立上，為了模型的認定 (identification) 於是模型中有許多限制，例如將變數視為外生變數，這對於變數性質的認定相當困難，因此 Sims (1980) 提出向量自我迴歸模型，其意義在於總體經濟變數會因時間而改變特性，其變化會反應在資料上，故以資料本身的特性來決定動態模式，亦即由資料本身去解讀經濟活動的本質，而本研究在進行民間消費預測時將以做為預測分析模型。

首先，因果分析有助於判斷指標間的因果情況，並透過衝擊反應及預測誤差拆解，則可去做預測分析。

向量自我迴歸模型的特色在於其可描繪變數間時序的動態變化，並可刻劃其因果結構。VAR 為學者 Sims (1980) 所提出，其認為大型計量模型存在模型設定問題，故 VAR 模型將所有變數都當成內生變數，避免任意限制變數間的關係 (陳旭昇，2014)。

在此本研究以考慮各指標自身落後項及其他指標落後項的 VAR 函數為討論主軸，考慮體系內跨變數的動態行為。模型概念如下：

$$\begin{bmatrix} \pi_t \\ \vdots \\ R_t \end{bmatrix} = \begin{bmatrix} \phi_1^{11} & \dots & \phi_1^{1n} \\ \vdots & \ddots & \vdots \\ \phi_1^{n1} & \dots & \phi_1^{nn} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ \vdots \\ R_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \dots \dots \dots (14)$$

係數 ϕ_p^{ij} 表示第 p 落後期的第 j 的變數對第 i 個變數的影響。矩陣型態

可寫為：

$$y_t = c + \phi_{1y_{t-1}} + \cdots + \phi_{py_{t-p}} + \varepsilon_t \dots\dots\dots (15)$$

各指標在經過 VAR 模型，瞭解各指標落後期間的相互影響狀況後，本研究將以衝擊反應去瞭解變數 π 如何因應外生衝擊下，隨時間對 R 演進的動態變化。在呈現上，以兩兩指標的關係作呈現，如民間消費受到經濟、檢索指標、綜合預測指標的影響期間狀況，以及何時開始趨緩等情況解釋，皆可透過衝擊反應函數去做進一步的瞭解。

為了瞭解各指標對民間消費預測關聯性，本研究透過預測誤差拆解的操作，可分析各指標在期間內（例如 10 期、20 期等）內的預測誤差變異，可被其他指標的解釋程度。

肆、社群網頁民間消費指標多變量分析

依據臉書的性質，其分析面向，可分為版主發言，民眾的回文或按讚或分享。分析重點如下：

1. 敘述統計

對於社群網站的大數據分析模型，在敘述統計的部份，可依據每年、每月或每日的關鍵字出現詞頻、按讚及分享數，去進行初步分析，此可呈現在何種時點下，版主及民眾對相關議題的關注力，並可與其他時點，其他對象來作比較分析。

而為瞭解欲分析主題的關鍵字出現頻率，是否具有相關性，如討論到各類品項的銷售是否會影響民間消費趨勢的情況，關鍵字的相關性分析則可提供分析者初步的判斷依據。

2. 多變量分析

在多變量分析的處理上，本研究會透過文字的集群分析及社會網絡分析，以了解議題、關鍵人、討論文字間的關係性，如消費品項間比較常與何種文字共同出現在同一篇發文或回文中。

第五章 研究成果

依據本研究之架構，對民間消費預測之指標設計，除傳統總體經濟指標之外，另以網路大數據指標輔助預測之。其中網路大數據指標則分為搜尋引擎、關鍵社群網站大數據兩面向建立指標，而在社群網站大數據的資訊擷取上，本研究設計並建置「社群網站資料擷取平台」，做為資料蒐集之工具，有關平台資訊請見附件四。

其中，社群網站指標因應發展趨勢，近三年各電商方開始經營相關粉絲團，故考量資料期間的面向，社群網站指標並未納入民間消費預測模型測試。本研究將透過相關性分析及社群網絡分析模組探索社群網站指標對民間消費之相關影響，並在後續本章第四節說明相關分析結果。

為將網路大數據納為預測民間消費之有效因子，故本研究透過主成份分析，其能將變數由繁至簡，濃縮精煉變數訊息，以便操作性分析。而為多面向探討改善民間消費之預測模組，本研究分別建立網路檢索指標，以及民間消費綜合指標，第一節則說明相關主成份分析內涵。後續則分別以第二節探討網路檢索指標對民間消費的預測效力，第三節則探討民間消費預測綜合指標之成效，以下則分別論述之。

第一節 大數據預測指標主成分分析

壹、網路檢索指標

在建立網路檢索指標上，係以商品品類關鍵字為基礎，觀察各消費者於品項在搜尋引擎 Google 的搜尋資料蒐集彙整，而本研究彙整多個電商（詳見表 3-1）品項⁵，並就電商網內各中類別之細目進行我國搜尋引擎大數據指標之建置。

以 Yahoo!奇摩來說明，其分為三大類別：購物中心、拍賣、服務+，整理出各類別中的分類項目後，將「中類」的分類項目，在上述三類中，共整理出的關鍵字清單為 529 項，購物中心佔了 42 項、拍賣佔了 370 項，服務+ 佔了 117 項，細目資訊請見圖 5-1。

我們以 15 家電商網之中類目去進行搜尋⁶，並依照電商網進行歸納，進行觀察，放入 Google Trends 搜尋趨勢，下載關鍵字的聲量數據資料，

⁵ 電商商品分類流程，以及網路大數據一搜尋指標篩選過程請見附件五。

⁶ 其中，因 PChome24 購物及 PChome 購物中心兩電商網中類目數較少，故合併視之。而 Yahoo 購物中心內的全商品分類已包含購物中心品牌旗艦之分類，故合併視之，總計共 15 家電商網。

再做後續的資料整理，資料總量達 1838 類商品檢索項目指標。

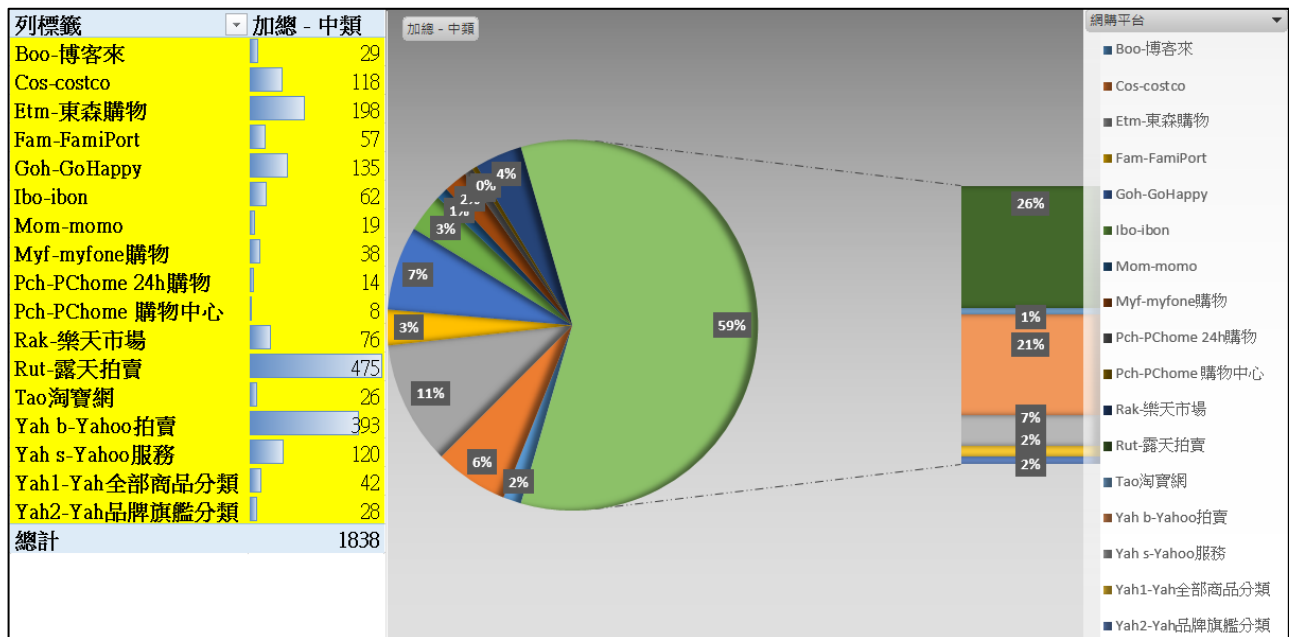


圖 5-1 15 大電商購物網站各網站中類的分類項目總計

依據其商品中類項目，並以 Google Trends 依據商品類目，建置自 2004 年以來各週搜尋聲量，調整其頻率為月資料，並依據各電商網，分別進行主成份萃取其電商網檢索指標，詳細主成份操作步驟請見附件五之說明。

後續則選擇各電商網第一主成份因子，再進行主成份分析，並依特徵值與因素解釋總變異量的能力決定因素萃取的個數，篩選原則同樣為選取特徵值大於 1 的因素或是累計解釋總變異量百分比達 70% 以上。表 5-1 顯示，若以特徵值大於 1 為篩選原則，由於自第三個因子開始之特徵值小於 1，故應萃取出兩個因子。此外，觀察變異量百分比及其累計值，第一個因子可解釋過半（61.2%）的總變異量，累積至第二個因子，可解說總變異量已達 94.5% 以上，且依圖 5-2 的陡坡圖，發現自第三個因子後特徵值差異趨緩⁷，因此，本研究選擇以兩個因子作為檢索指標進行分析⁸。

⁷ 相關分析顯示，兩者僅有微幅負相關（-0.0022），表示兩主成份皆有其代表性，且共同性小。並衡量主成份的 KMO 值，得 0.898。KMO 為取樣適當性衡量量數，值介於 0 至 1 之間，越大表示主成分效度越高。

⁸ 有關各電商網第一主成分組成商品資訊請見附件六。

表 5-1 檢索指標主成份分析結果之解釋總變異量

成份	特徵值	平方差變異量%	平方差累積變異量%
1	9.183	0.612	0.612
2	4.996	0.333	0.945
3	0.582	0.039	0.984
4	0.076	0.005	0.989
5	0.041	0.003	0.992
6	0.031	0.002	0.994
7	0.022	0.002	0.996
8	0.017	0.001	0.997
9	0.013	0.001	0.998
10	0.010	0.001	0.998
11	0.008	0.001	0.999
12	0.007	0.000	0.999
13	0.005	0.000	1.000
14	0.005	0.000	1.000
15	0.003	0.000	1.000

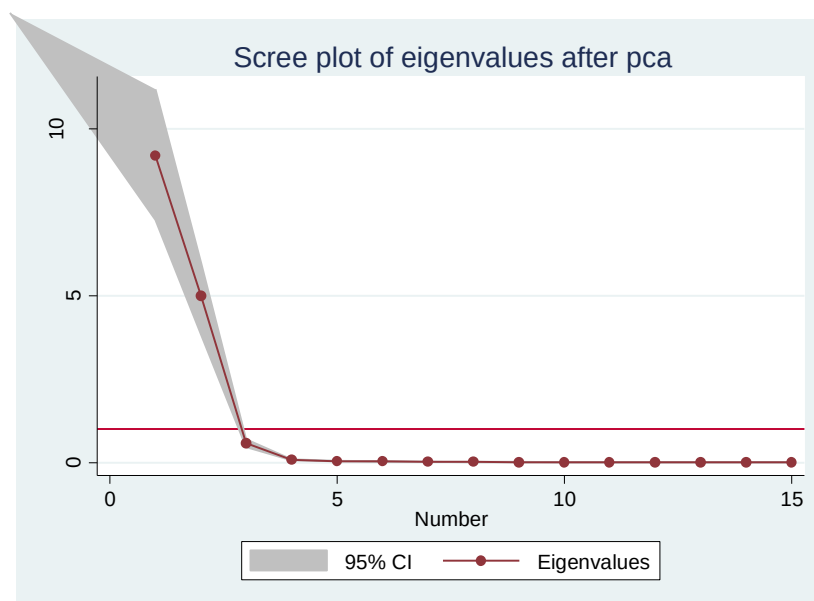


圖 5-2 網路檢索指標主成份分析結果之陡坡圖

表 5-2 則顯示，本研究所萃取出之兩項網路檢索指標之成份因子，所組成之主要解釋因子來源資訊，以及其平方差解釋變異量。

表 5-2 網路檢索指標平方差解釋變異量

15 大電商網站第一大主成份	網路檢索指標 第一主成分 (gt_fac_1)	網路檢索指標 第二主成分 (gt_fac_2)
Yah b -Yahoo 拍賣	0.992	
Rak-樂天市場	0.985	
Myf-myfone 購物	0.984	
Rut-露天拍賣	0.983	
Yah s -Yahoo 服務	0.961	
Cos-Costco	0.947	
Ibo-ibon	0.926	
Boo-博客來	0.901	
Pch-PChome 購物中心	0.889	
FAC_MOM_1	0.813	
Yah1-Yah 全部商品分類		0.992
Tao 淘寶網		0.991
Etm-東森購物		0.980
Goh-GoHappy		0.944
Fam-FamiPort		0.822

貳、民間消費預測綜合指標建置

相較於過去多以總體經濟相關概念對民間消費進行預測，本研究除探索納入網路檢索指標之預測效果之外，也欲建立民間消費綜合指標，期望透過較多樣化之大數據組合資訊，評估其對民間消費預測之相關效果，以下則分別就靜態萃取（PCA）及動態因素萃取模型（DFM）建立之。

為預測我國民間消費未來走勢，本研究搜納總體經濟指標作為民間消費預測之總體領先指標資料庫，資料期間彙整期間為 2004 年 1 月/季起。在總體經濟指標選擇考量，為能有效發揮消費預測之效果，本研究主要以經濟領先指標為目標，並參考 Vosen and Schmidt (2011)、Matsumoto, Matsumura and Shiraki (2013) 在進行消費預測時，納入我國消費者信心指數 CCI 為觀察變數。

表 5-3 總體經濟指標資料建置

序號	主要指標	頻率	來源
總體經濟領先指標			
1	外銷訂單指數 (Index 2011=100)	月	主計總處
2	貨幣總計數 M1B (十億元)	月	主計總處
3	股價指數 (Index 1966=100)	月	主計總處
4	工業及服務業受僱員工淨進入率 (%)	月	主計總處
5	核發建照面積 (住宅類住宅、商業辦公、工業倉儲) (千平方公尺)	月	主計總處
6	SEMI 半導體接單出貨比 (比率)	月	主計總處
市場信心			
7	消費者信心指數 CCI	月	中央大學

而為淬鍊民間消費預測綜合指標，在處理步驟上，除上述建立之網路檢索大數據指標，則先進行總體經濟領先指標之主成份分析，透過主成份分析，將變數由繁至簡，濃縮精煉總體經濟領先指標之變數訊息。

1. 總體經濟領先指標萃取

在萃取總體經濟領先指標時，考量消費者信心指數為展現消費者對市場目前看法，為主觀指標，與領先指標內之各項指標（如：外銷訂單指數、M1B、股價指數等）為市場實際表現，故在成份萃取時，則不將消費者信心指數納入經濟領先指標之操作範圍。

欲萃取之總體經濟領先指標，包含外銷訂單指數、貨幣總計數 M1B、股價指數、工業及服務業受僱員工淨進入率、核發建照面積及 SEMI 半導體接單出貨比進行成分萃取，依特徵值 (Eigenvalue) 與因素解釋總變異量的能力決定因素萃取的個數，篩選原則為選取特徵值大於 1 的因素或是累計貢獻數百分比達 70% 以上。表 5-4 顯示主成份分析解說成果累積變異量，由累積變異量並參考特徵值陡坡圖 (圖 5-3)⁹，本研究選擇前二成份因子作為總體經濟領先指標內涵¹⁰。

⁹ 本研究考量變量本身之特質來捕捉其成份數，故並未轉秩處理。

¹⁰ 總體經濟領先指標第一主成份 (buscyc_fac_1) 及第二主成份 (buscyc_fac_2) 之相關係數為 -0.0057，相關性小，顯示萃取出之因子有其一定的代表性。衡量主成份的 KMO 值為 0.626。

表 5-4 領先指標主成份分析結果之解說總變異量

成份	特徵值	平方差變異量%	平方差累積變異量%
1	2.682	0.447	0.447
2	1.594	0.266	0.713
3	0.754	0.126	0.838
4	0.584	0.097	0.936
5	0.297	0.050	0.985
6	0.089	0.015	1.000

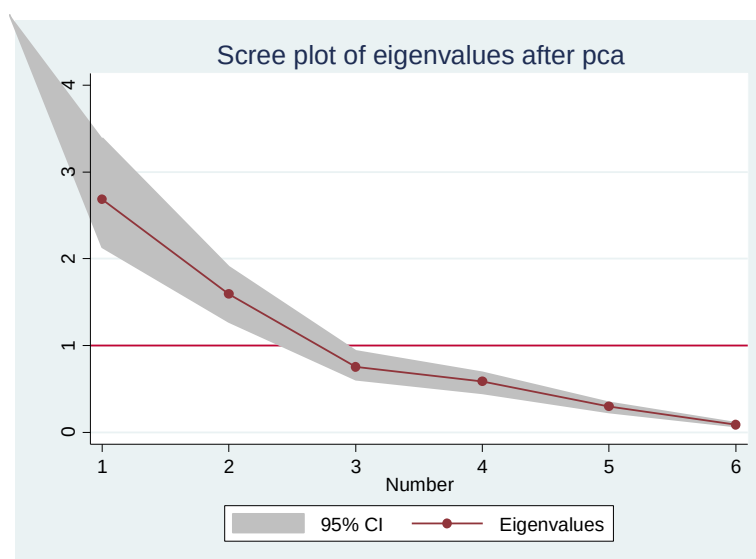


圖 5-3 總體經濟領先指標主成份分析結果之陡坡圖

下表 5-5 則顯示，本研究所萃取出之兩項總體經濟領先指標之成份因子，其所組成的因子主要來源資訊。

表 5-5 總體經濟領先指標平方差解釋變異量

領先指標	總體經濟領先指標 第一主成分 buscyc_fac_1	總體經濟領先指標 第二主成分 buscyc_fac_2
外銷訂單指數 (Index 2011=100)	0.911	
貨幣總計數 M1B (十億元)	0.894	
股價指數 (Index 1966=100)	0.874	
工業及服務業受僱員工淨進入率 (%)		0.782
核發建照面積 (住宅類住宅、商業辦公、工業倉儲) (千平方公尺)		0.759
SEMI 半導體接單出貨比 (比率)		0.484

2. 民間消費預測綜合指標建置

在建置民間消費預測綜合指標的部分，則納入總體經濟領先指標之成份要素、網路檢索指標之成份要素及消費者信心指數進行主成份分析，以建置民間消費預測綜合指標。其中消費者信心指數為民眾對消費市場的樂觀程度、而網路檢索指標則是購買前之瀏覽行為、代表其關注度，而總體經濟領先指標則為代表市場活絡度之客觀指標。以下則分別就靜態及動態萃取成果進行說明。

(1) 民間消費預測綜合指標主成份分析—靜態萃取

圖 5-4 為民間消費預測綜合指標靜態主成份分析操作示意，針對前文所萃煉之網路檢索指標（gt_fac_1、gt_fac_2）、消費者信心指數（CCI）及總體經濟領先指標（buscyc_fac_1、buscyc_fac_2），透過靜態主成份（PCA）萃取，以建立民間消費預測綜合指標。

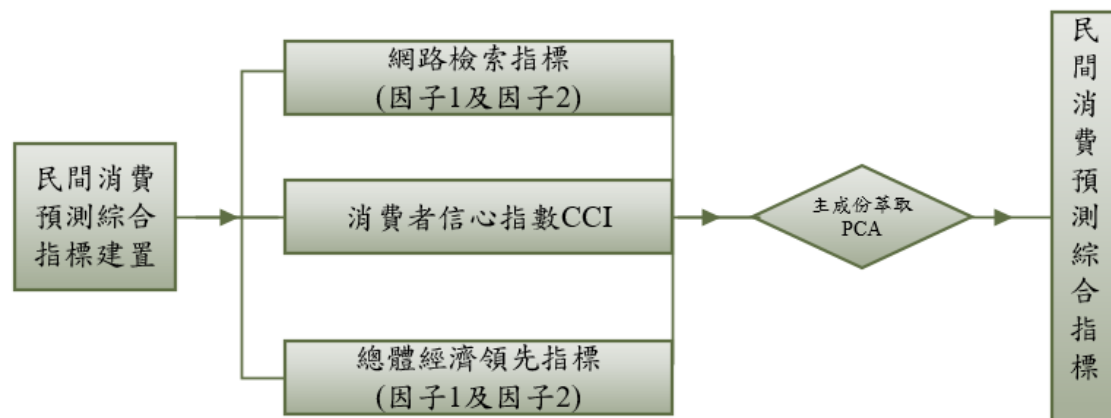


圖 5-4 民間消費預測綜合指標靜態主成份分析

透過主成份分析，依特徵值與因素解釋總變異量的能力決定因素萃取的個數，篩選原則同樣為選取特徵值大於 1 的因素或是累計變異量百分比達 70% 以上。若以特徵值大於 1 為篩選原則，由於自第三個因子開始之特徵值小於 1，故應萃取出兩個因子，作為民間消費綜合指標主成份分析模組之指標¹¹。（請參考表 5-6 及圖 5-5）

¹¹民間消費綜合指標（PCA）第一主成份（pca_factor1）及第二主成份（pca_factor2）之相關係數為-0.0457，相關性相當小，顯示萃取出之因子有其一定的代表性。此外衡量民間消費綜合指標主成份的 KMO 值為 0.387。

表 5-6 主成份分析模組之解說總變異量

成份	特徵值	差異質	平方差變異量%	平方差累積變異量%
1	2.388	1.040	0.478	0.478
2	1.348	0.372	0.270	0.747
3	0.976	0.719	0.195	0.943
4	0.258	0.228	0.052	0.994
5	0.029	.	0.006	1.000

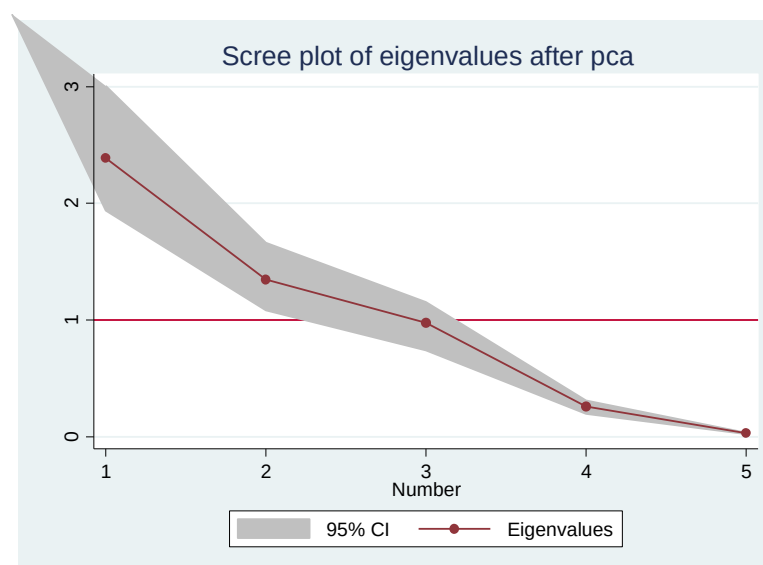


圖 5-5 主成份分析結果之陡坡圖

下表 5-7 則顯示，本研究所萃取出之兩項民間消費綜合指標（PCA）之成份因子，其所組成的因子主要來源資訊。

表 5-7 民間消費綜合指標（PCA）平方差解釋變異量

指標	民間消費綜合指標 (PCA) 第一主成份 pca_factor1	民間消費綜合指標 (PCA) 第二主成份 pca_factor1
buscyc_fac_2	0.948	
CCI	0.805	
gt_fac_2	0.725	
gt_fac_1		0.711
buscyc_fac_2		0.670

(2) 民間消費預測綜合指標主成份分析—動態萃取 (DFM 分析)

主成份分析能將變數由繁至簡，濃縮精煉變數訊息，而動態因素模型 (DFM) 則考量靜態因子及動態因子之落後期項所萃取出成份因子，而本研究在主成份分析之外，加入動態因素考量，比較並對照因素分析與加入動態因子後的動態因子模型 (DFM) 的異同，以更清楚了解臺灣民間消費情勢如何隨時間與外在衝擊而變化。

受限於 DFM 萃取方式會受到變項數量的影響，當數量過多且樣本數不足的情況下，則無法進行 DFM¹²，考量指標萃取的一致性¹³，故本研究則同樣先以靜態主成份分析萃取出總體經濟領先總指標與網路檢索總指標，後續則針對總體經濟領先總指標、消費者信心指數及網路檢索總指標施以 DFM 分析，萃取出民間消費預測綜合指標，來觀察其與前述之靜態分析模組之差異性。以下則針對 DFM 因子及分析模組成果進行說明。

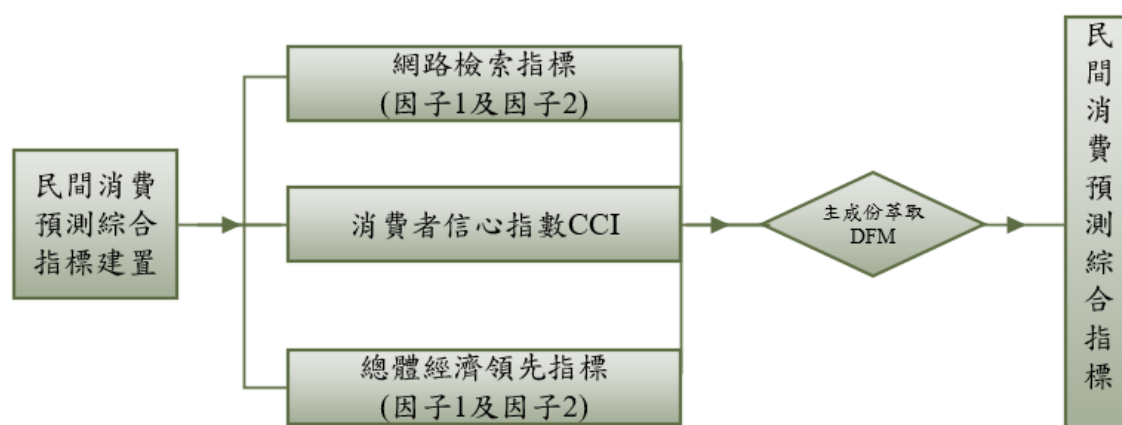


圖 5-6 DFM 模組分析

在操作上 DFM 運用兩種方式找出共同因子，方法 A 為直接運用總體經濟領先總指標 (buscyc_fac_1、buscyc_fac_2)、CCI 及網路檢索總指標 (gt_fac_1、gt_fac_2) 3 類別因經基準化 (反向基準化) 後找出共同因子；方法 B 為先針對此 3 類因子所含之變數群，個別進行主成份分析萃取一因子¹⁴，將此 3 個類別因子基準化及反向基準化後，找出共同因子。

¹² 如本研究所建構之 15 大電商網第一主成份萃取成網路檢索總指標即面對變數量過多的問題，無法進行 DFM 分析。

¹³ 避免僅有電商網使用靜態主成份分析萃取因子，而總體經濟領先指標卻利用 DFM 萃取，故統一透過主成份分析萃取各大類總指標，而民間消費預測綜合指標的萃取則施以靜態主成份 PCA 及 DFM 進行後續分析。

¹⁴ 在此以各主成份中，第一主成份進行萃取，主要考量第一主成分在平方差變異量為較高，較具該主指標之代表性。

本研究經過多種萃取模組組合嘗試¹⁵，並考量目標樣本數較短（配合網路檢索指標由 2004 年起），遂以一階差分為 DFM 模型衡量之設定，避免樣本遺失過多問題。本研究依據統計原則篩選最適模型，經由前述動態因素分析模型（DFM）之操作流程，篩選最終選定模型（模型組合 A1），並以估計之共同因子為民間消費預測綜合指標，命名為 dfm_factor1。

表 5-8 DFM 操作要素說明

編碼		類型	因子萃取數	AIC	ll (model)
方法 A	因子全一階差分				
	A1*	有常數項，共同因子落後一期	1	238.605*	-104.303
	A2	無常數項，共同因子落後一期	1	242.166	-111.083
	A3	無常數項	1	297.526	-143.7635
	A4	有常數項	無法收斂		
	水準值				
	A5	有常數項，共同因子落後一期	無法收斂		
	A6	無常數項，共同因子落後一期	無法收斂		
	A7	無常數項	1	1121.091	-555.5456
	A8	有常數項	1	935.9259	-457.963
方法 B	因子全一階差分				
	B1	有常數項，共同因子落後一期	1	310.1803	-146.0902
	B2	無常數項，共同因子落後一期	1	307.9399	-147.97
	B3	無常數項	無法收斂		
	B4	有常數項	無法收斂		
	水準值				
	B5	有常數項，共同因子落後一期	無法收斂		
	B6	無常數項，共同因子落後一期	無法收斂		
	B7	無常數項	無法收斂		
	B8	有常數項	無法收斂		

資料來源：本研究。

以下分別就網路檢索指標及民間消費預測綜合指標進行預測模型分析，評估應用大數據可否提升台灣民間消費預測。第二節實證分析為以網路檢索指標為基礎進行評估；第三節則以民間消費預測綜合指標為基礎進行評估。

¹⁵ 模型測試中，落後二期（含）以上，皆無法收斂，故以落後一期為模組設定。

第二節 應用大數據提升台灣民間消費預測實證分析

本研究的主要目的為應用大數據來提升民間消費預測之成效。本節以目前本研究所建置之網路大數據指標，其乃透過 15 大電商網，依據其商品中類項目，並以 Google Trends 依據商品類目，建置自 2004 年以來各週搜尋聲量，調整其頻率為月資料，並依據各電商網，分別進行主成份萃取其電商網檢索指標。並選擇各電商網第一主成份因子，再進行主成份分析，萃取出網路檢索指標，後續則透過 ARIMA 及 VAR 模型來進行民間消費的預測。進行模型檢測前，首先會透過單根檢定，來觀察及確認變項之適合性。

壹、變數說明

為進行預測模型的配適比較，本研究納入一般性總體變數（GDP、銀行利率、消費者物價指數等）、經濟領先指標。其中總體經濟領先變項為前述透過主成份萃取為總體經濟領先指標¹⁶。而網路檢索指標則是透過十五大電商網萃取而得（詳見表 5-9）。

表 5-9 變數說明

變數類型	變數中文名稱	資料來源	備註
應變數	台灣民間消費連鎖實質值	主計總處	
總體變數	台灣 GDP 連鎖實質值	主計總處	
	消費者物價指數年增率 ¹⁷	主計總處	
總體經濟領先指標	總體經濟領先指標因子 1	本研究	由外銷訂單指數、股價指數、工業及服務業受僱員工淨進入率、核發建照面積、SEMI 半導體接單出貨比，經由主成份萃取而得
市場信心指標	台灣消費者信心指數 CCI	中央大學	
網路檢索指標	網路檢索指標因子 1	本研究	
	網路檢索指標因子 2	本研究	

資料來源：本研究

¹⁶ 本研究另有納入台灣消費者信心指數 CCI，於非民間消費預測綜合指標模型中測試，惟較未顯著，且與消費者物價指數有相互影響，ARIMA 模型結果請見附件九。

¹⁷ 本文參考 Burke and Ozdagli. (2013) 作法，將消費者物價指數年增率視為通膨預期變數，將之納入作為影響民間消費的總體經濟變數之一。

貳、單根檢定

在進行預測模型評估前，首先針對各變項進行單根檢定，降低非定態之變項特性影響後續評估及預測，AR (p)、MA (q) 最大階次均設定為 4，並以 ADF 單根檢定進行評估。惟避免損失樣本及數列特性，本研究模型均以 I (0) 及 I (1) 進行模型設定、模型估計後檢視殘差均為定態，單根解檢定及變數處理請見表 5-10¹⁸。

表 5-10 單根檢定

變數中文名稱	變數代碼	ADF 單根檢定			以一階差分處理 (v)
		截距	截距與趨勢	無截距與趨勢	
台灣民間消費連鎖實質值	logpc	I (2)	I (2)	I (1)	v
台灣 GDP 連鎖實質值	rgdp	I (1)	I (0)	I (1)	
消費者物價指數年增率	cpiyoy	I (0)	I (0)	I (0)	
總體經濟領先指標因子 1	buscyc_fac_1	I (0)	I (0)	I (0)	
總體經濟領先指標因子 2	buscyc_fac_2	I (1)	I (1)	I (1)	v
消費者物價指數年增率	cpiyoy	I (0)	I (0)	I (0)	
台灣消費者信心指數 CCI	cci	I (1)	I (1)	I (1)	v
網路檢索指標因子 1	gt_fac_1	I (0)	I (2)	I (1)	v
網路檢索指標因子 2	gt_fac_2	I (2)	I (2)	I (0)	v

註：I (x)，表經過 x 階差分後為定態。

參、相關係數矩陣及敘述統計

表 5-11 呈現變數之敘述統計。而透過表 5-12 之相關分析可得知，總體景氣領先第二因子與第一因子相關係數達 0.788，避免影響模型評估，故模型中只放入總體景氣領先第一因子¹⁹。網路檢索指標第一、第二因子相關係數超過 0.8，可能造成共線性問題，惟此兩因子為本研究核心變數，因此建模之後進一步檢查變異數膨脹因素 (VIF, variance inflation faction)，經建模後檢視第一、第二因子 VIF 分別為 3.085、3.360，共線性問題應不嚴重。

¹⁸ 台灣民間消費連鎖實質值在納入截距與截距與趨勢為二階差分後為定態，惟兩模型在常數項、趨勢項不顯著，故以無截距與趨勢之一階差分為後續研究基礎。

¹⁹ 本文另有測試實質銀行業利率一個月期存款牌告利率，惟其與 CPIYOY 相關係數為-0.926，故模型放入消費者物價指數年增率。

表 5-11 變數敘述統計

變數	樣本數	平均數	標準差	最小值	最大值
dlog_pc	48	0.0023	0.0099	-0.0163	0.0220
dlog_rgdp	48	0.0038	0.0236	-0.0469	0.0291
cpiyoy	49	1.2626	1.3483	-1.3433	4.5333
buscyc_fac_1	49	-0.0013	1.0066	-3.1366	1.7996
dbuscyc_fa~2	48	0.0597	0.1550	-0.6043	0.3787
dgt_fac_1	48	-0.0666	0.2380	-0.7739	0.3861
dgt_fac_2	48	0.0382	0.2051	-0.4176	0.4824

表 5-12 相關係數矩陣

	dlog_pc	dlog_rgdp	cpiyoy	buscyc_fac_1	dbuscyc_fac_2	dgt_fac_1	dgt_fac_2
dlog_pc	1						
dlog_rgdp	-0.1086	1					
cpiyoy	-0.1682	-0.0601	1				
buscyc_fac_1	0.1171	0.2164	0.0387	1			
dbuscyc_fa~2	0.0946	0.2676	-0.1801	0.788	1		
dgt_fac_1	-0.3364	0.4023	-0.1864	-0.0212	-0.0221	1	
dgt_fac_2	-0.1509	0.5523	-0.2077	0.0074	0.0634	0.8042	1

備註：「d 變數名」表變數經過一階差分處理。

肆、ARIMA 模型分析

本研究透過 Eviews Automatic ARIMA Forecasting 功能，設定 ARIMA (p,d,q)；其中為避免損失過多樣本，因此 AR (p), MA (q) 最大階次均設定為 4，I (d) 則因民間消費經一次差分後已為定態序列，因此 d 設定為 0²⁰。並以 AIC 最小值作為挑選模型原則，經挑選最適 ARIMA(p,d,q) 階次為 ARIMA (4,1,0)，最終以 ARIMA (4,1,0) 為本研究最適 ARIMA 模型²¹。(請見表 5-13 之結果)

為進行模型預測績效比較，本文除了建構 ARIMA 模型，另建構一涵蓋與民間消費相關的傳統經濟變數之 ARIMAX 模型 (以下稱：ARIMAX-

²⁰ 由於以 (p,d,q) 形式呈現，則以 (p,l,q) 表示。

²¹ 本研究測試納入民間消費信心、民間消費預測總指標 (動態及靜態萃取)，惟模型檢測結果，故選擇目前模型為最適模型。

傳統模型)。此外，為探討大數據資料是否能提升預測我國民間消費的準確性，因此將傳統 ARIMAX 模型擴充，加入由 Google Trends 聲量資料中所萃取之第一、第二主成分因子，藉以評估加入大數據資料後的模型（以下稱 ARIMAX GT-Model），是否能有效提升預測我國民間消費之準確度。模型評估結果請見表 5-14。

表 5-13 ARIMA 最適落後期分析

Model	Log Likelihood	AIC	BIC	HQ
(3,1,3)*	170.7399	-7.23289	-6.9117	-7.11315
(4,1,3)*	171.0347	-7.20154	-6.84021	-7.06684
(3,1,4)*	170.8905	-7.19514	-6.8338	-7.06043
(4,1,2)*	169.4078	-7.17368	-6.8525	-7.05395
(4,1,0)	165.5358	-7.09048	-6.84959	-7.00068
(1,1,1)	163.4676	-7.08745	-6.92686	-7.02758
(3,1,2)	166.3965	-7.08429	-6.80325	-6.97952
(2,1,2)	164.8841	-7.06151	-6.82063	-6.97171
(4,1,1)	165.8758	-7.06115	-6.78011	-6.95638
(1,1,3)	164.8523	-7.0601	-6.81922	-6.9703
(1,1,4)	165.7287	-7.05461	-6.77357	-6.94984
(1,1,2)	163.6415	-7.05074	-6.85	-6.9759
(2,1,1)	163.575	-7.04778	-6.84704	-6.97294
(3,1,1)	164.408	-7.04036	-6.79947	-6.95055
(2,1,3)	165.0777	-7.02568	-6.74464	-6.92091
(2,1,4)	165.8679	-7.01635	-6.69517	-6.89662
(4,1,4)	167.1285	-6.98349	-6.58201	-6.83382
(3,1,0)	159.7368	-6.87719	-6.67645	-6.80236
(2,1,0)	156.4469	-6.77542	-6.61483	-6.71555
(1,1,0)	155.1245	-6.76109	-6.64065	-6.71619
(0,1,4)	155.5778	-6.6479	-6.40701	-6.5581
(0,1,2)	152.6132	-6.60503	-6.44444	-6.54516
(0,1,3)	152.6196	-6.56087	-6.36013	-6.48604
(0,1,1)	148.8204	-6.48091	-6.36046	-6.43601
(0,1,0)	139.4765	-6.11007	-6.02977	-6.08013

備註：*(3,1,3)、(4,1,3)、(3,1,4)、(4,1,2)於估計時模型不穩定，故未選取之。

於模型預測的部分，本研究針對 ARIMA 模型分析結果，分別進行樣本內及樣本外預測，且其中並就靜態與動態預測分別檢測。

表 5-14 ARIMA 分析結果

ARIMA Model			ARIMAX-傳統模型			ARIMAX GT-Model			ARIMAX GT-Model		
ARIMA (4,1,0)			ARIMAX (1,1,0)			ARIMAX (3,1,0)			ARIMAX (1,1,0)		
變項	估計係數	Std. Error	變項	估計係數	Std. Error	變項	估計係數	Std. Error	變項	估計係數	Std. Error
C	0.002 ***	0.001	C	0.003 ***	0.000	C	0.002 ***	0.000	C	0.003 ***	0.001
AR (1)	-0.291 *	0.151	AR (1)	-0.931 ***	0.059	AR (1)	-0.809 ***	0.159	AR (1)	-0.933 ***	0.058
AR (2)	0.024	0.151				AR (2)	-0.399 *	0.227			
AR (3)	-0.153	0.138				AR (3)	-0.554 ***	0.136			
AR (4)	0.44 ***	0.139									
			DLOGRGDP	0.176 ***		DLOGRGDP	0.152 ***	0.045	DLOGRGDP	0.176 ***	0.033
			CPIYOY	-0.001 ***	-0.001	CPIYOY	-0.001 ***	0.000	CPIYOY	-0.001 ***	0.000
			BUSCYC_FAC_1	0.001	0.001	BUSCYC_FAC_1	0.001 **	0.000	BUSCYC_FAC_1	0.000	0.000
						DGT_FAC_1	-0.012 ***	0.003	DGT_FAC_1	-0.007 *	0.004
						DGT_FAC_2	0.013 ***	0.003	DGT_FAC_2	0.004	0.005
R-squared			0.693	R-squared			0.85	R-squared			0.902
Adjusted R-squared			0.654	Adjusted R-squared			0.83	Adjusted R-squared			0.876
Log likelihood			169.827	Log likelihood			180.134	Log likelihood			189.155
											183.34

一、樣本內預測

針對樣本內預測，本研究依據前述設計之四模型，就靜態與動態模型分別比較，依據表 5-15 樣本內靜態預測績效比較可發現，納入大數據資料之模型(ARIMAX(3,1,0)GT Model 及 ARIMAX(1,1,0)GT Model)，預測效果在 RMSE、MAE、Theil U 等三指標，皆相較 ARIMA Model 及 ARIMAX-傳統模型表現為佳。而在樣本內動態預測績效比較的部分，雖與 ARIMA Model 各有優劣，但在多項指標中，顯示 ARIMAX (3,1,0) GT Model 有較好的績效表現。圖 5-7 及 5-8 分別呈現樣本內靜態與動態之預測序列趨勢。

表 5-15 樣本內靜態及動態預測績效比較

Forecast Evaluation statistics	RMSE	MAE	MAPE	Theil U
靜態預測				
ARIMA (4,1,0)	0.0056	0.0043	389.7643	0.2863
ARIMAX (1,1,0) 傳統模型	0.0038	0.0029	115.2349	0.1817
ARIMAX (3,1,0) GT Model	0.0032	0.0026	211.6075	0.1548
ARIMAX (1,1,0) GT Model	0.0037	0.0030	114.4096	0.17976
動態預測				
ARIMA (4,1,0)	0.0072	0.0059	83.1975	0.4649
ARIMAX (1,1,0) 傳統模型	0.0090	0.0081	288.8734	0.5754
ARIMAX (3,1,0) GT Model	0.0067	0.0057	139.2236	0.4079
ARIMAX (1,1,0) GT Model	0.0090	0.00811	277.7358	0.57862

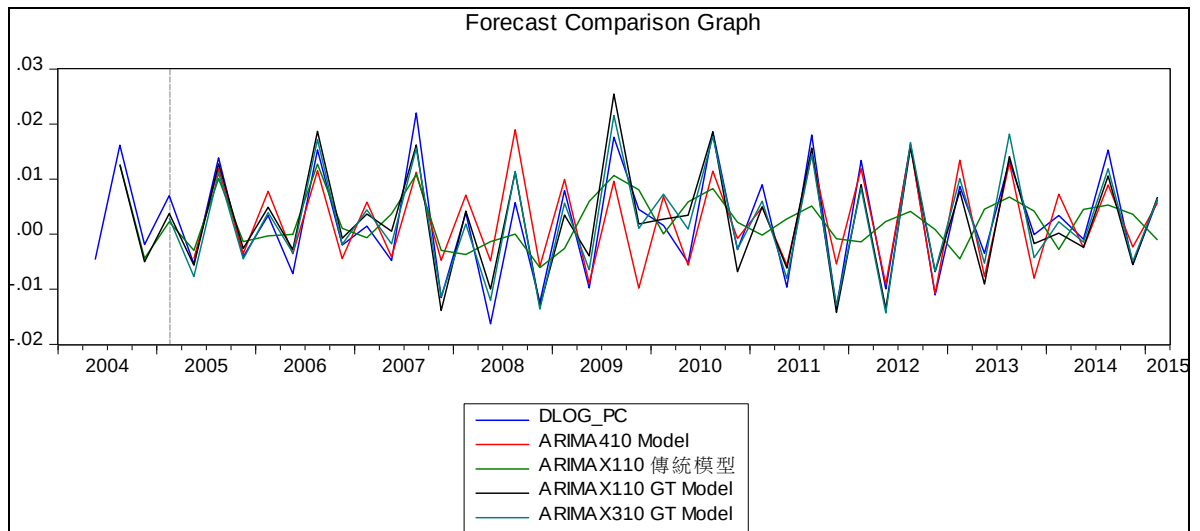


圖 5-7 樣本內靜態預測趨勢

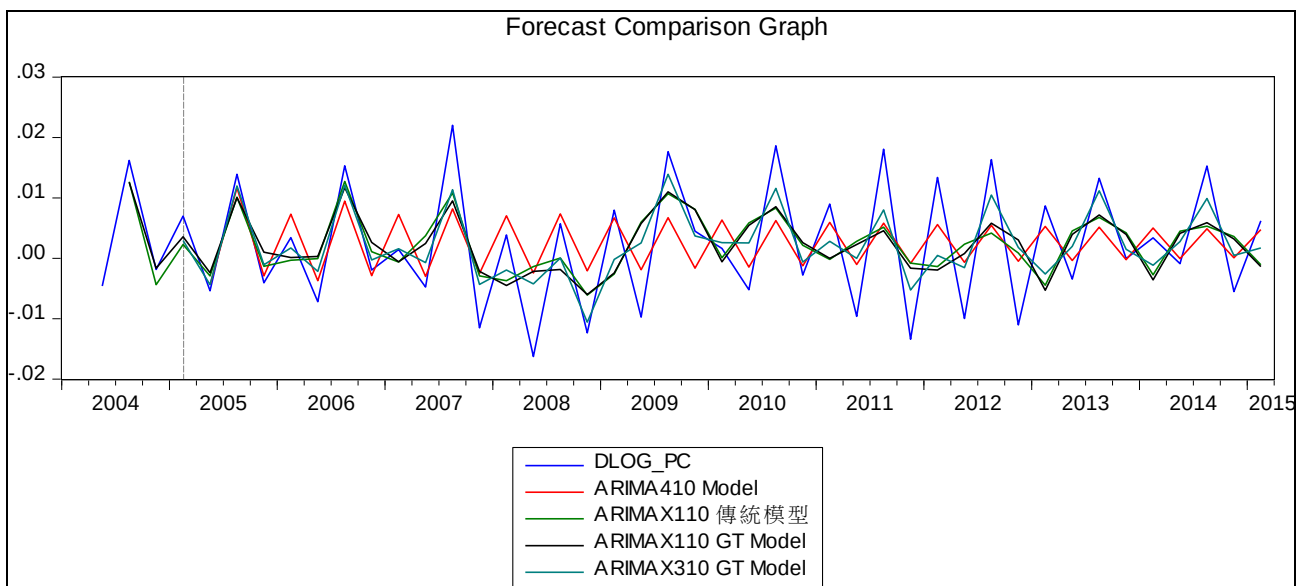


圖 5-8 樣本內動態預測趨勢

此外，本文針對研究中所建構的 ARIMA、ARIMAX-GT 模型與 ARIMAX-傳統模型的樣本內預測績效，進行 Diebold-Mariano (DM) Test，檢定結果可發現 ARIMAX (3,1,0) GT Model 與兩者有顯著差異。

表 5-16 ARIMA 樣本內預測 DM Test

模型	ARIMA (4,1,0)	ARIMAX (1,1,0) 傳統模型
ARIMAX (1,1,0) 傳統模型	-4.2069 ***	6.3660 *** 3.8347 ***
ARIMAX (3,1,0) GT Model	-3.0509 ***	
ARIMAX (1,1,0) GT Model	0.8365	

二、樣本外預測：

於樣本外預測之設計，考量樣本期間過短，故本研究以 2004Q1 至 2015Q1 為估計區間，對 2015Q2 及 2016Q1 等四個樣本點進行預測分析，以避免樣本過少影響估計。於樣本外靜態預測，ARIMA 模型預測教具績效，而在動態預測部份，ARIMA 與 ARIMAX (1,1,0) 傳統模型互有領先，不過若觀察四組模型的調整後判定係數 ($\text{adj } R^2$)，則以 ARIMAX-GT (3,1,0) 模型最佳。(見表 5-17)

表 5-17 樣本外靜態及動態預測績效比較

Forecast Evaluation statistics	RMSE	MAE	MAPE	Theil U
靜態預測				
ARIMA (4,1,0)	0.003	0.003	334.635	0.318
ARIMAX (1,1,0) 傳統模型	0.005	0.005	525.090	0.509
ARIMAX (3,1,0) GT Model	0.007	0.006	572.855	0.520
ARIMAX (1,1,0) GT Model	0.006	0.005	631.137	0.5562
動態預測				
ARIMA (4,1,0)	0.004	0.003	464.803	0.339
ARIMAX (1,1,0) 傳統模型	0.005	0.004	300.378	0.437
ARIMAX (3,1,0) GT Model	0.007	0.005	307.851	0.532
ARIMAX (1,1,0) GT Model	0.005	0.004	367.683	0.480

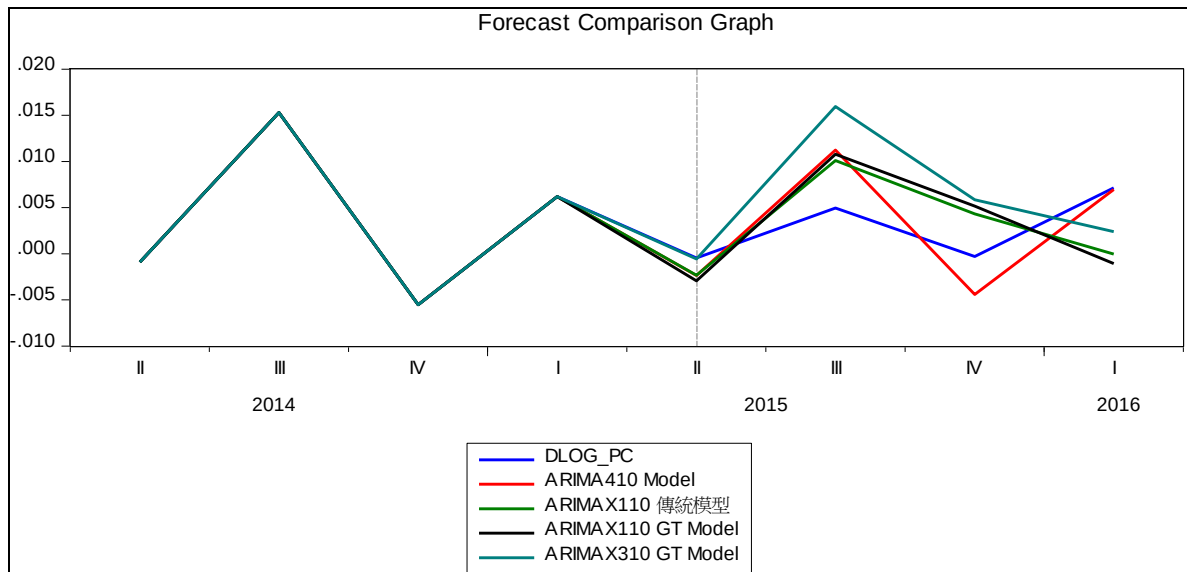


圖 5-9 樣本外靜態預測趨勢圖

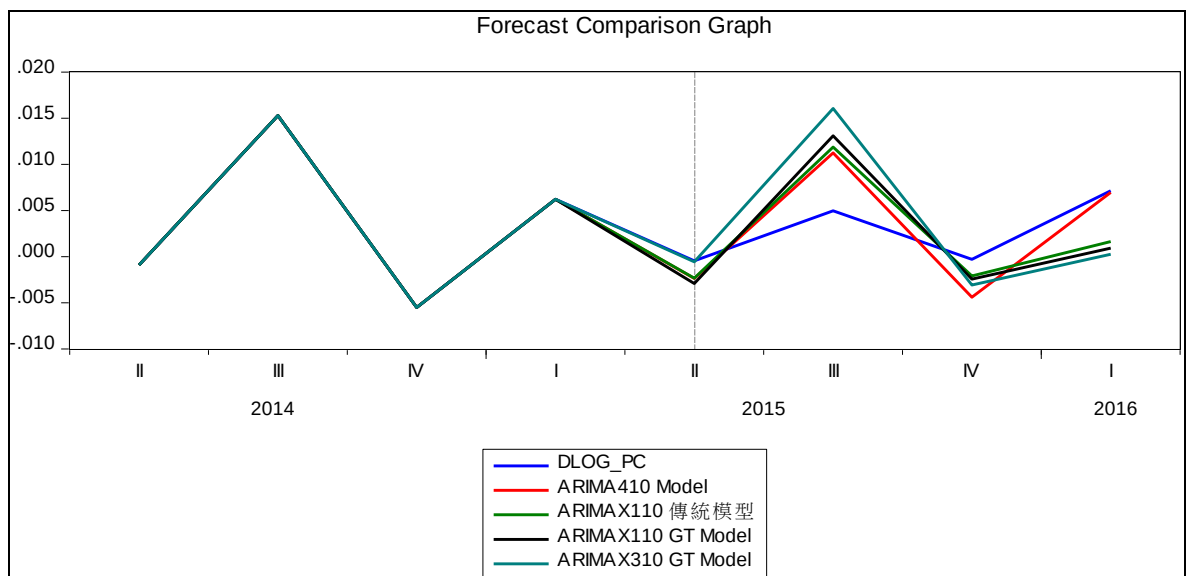


圖 5-10 樣本外動態預測趨勢圖

本文亦針對研究中所建構的 ARIMA、ARIMAX-GT 模型與 ARIMAX-傳統模型的樣本外預測績效，進行 Diebold-Mariano (DM) Test，檢定結果可發現 ARIMAX (3,1,0) GT Model 並未與 ARIMA 及 ARIMAX-傳統模型有顯著差異。

表 5-18 ARIMA 樣本外預測 DM Test

模型	ARIMA (4,1,0)	ARIMAX (1,1,0) 傳統模型
ARIMAX (1,1,0) 傳統模型	-0.7794	-1.374 -2.2478 ***
ARIMAX (3,1,0) GT Model	-1.3936	
ARIMAX (1,1,0) GT Model	-1.54641	

伍、VAR 模型分析

在進行 VAR 評估時，考量樣本數，以及參考前述 ARIMA 分析結果，以落後 3 期為最大階次，並以 AIC 作為判斷依據（見表 5-19），選擇落後 3 階之 VAR 模型。為進行預測績效比較，本研究同樣建置 VAR-傳統模型，以及 VAR-GT 模型（納入網路檢索指標），作為後續預測分析之基準。

表 5-19 VAR 落後期設定篩選表

模型	落後期	AIC	log likelihood
VAR-GT 模型	1	-11.840	296.555
	2	-14.174	375.657
	3*	-16.065	443.337
VAR-傳統模型	1	-8.686	206.741
	2	-10.692	260.523
	3*	-11.940	296.772

1. VAR-傳統模型

在建置 VAR-傳統模型，同樣以實質 GDP (drgdp)、消費者物價指數年增率、總體經濟領先指標因子為研究標的，設定落後 3 期，並以 2004Q1 至 2015Q1 為估計區間，對 2015Q2 及 2016Q1 進行樣本外預測分析。表 5-20 及 5-21 分別為兩模型估計結果。

VAR 模型皆經由 Lagrange-multiplier test，確認內有落後期內殘差無序列相關，並由 Jarque-Bera test，認定符合常態分配，而因果分析、衝擊反應及預測誤差拆解結果請見附件七。

表 5-20 VAR-傳統模型

		dlog_pc	dlog_rgdg	cpiyoy	buscyc_fac_1
		估計係數	估計係數	估計係數	估計係數
_cons		0.007 ***	0.015 ***	0.183	0.197 *
dlog_pc	L1.	-0.581 ***	0.151	80.789 *	7.548
	L2.	0.344 *	0.373	12.863	1.276
	L3.	0.167	0.766 **	-78.204 *	3.108
dlog_rgdg	L1.	0.080	-0.758 ***	-22.769 *	3.016
	L2.	-0.187 ***	-0.976 ***	-10.821	0.455
	L3.	-0.118 **	-0.817 ***	3.113	6.248 **
cpiyoy	L1.	-0.003 ***	-0.001	1.004 ***	-0.076 *
	L2.	0.001	-0.001	-0.048	-0.029
	L3.	-0.001	-0.001	-0.048	-0.088 *
buscyc_fac_1	L1.	-0.002	0.012 **	1.319 **	1.490 ***
	L2.	0.005	-0.001	-1.225	-1.167 ***
	L3.	-0.003	-0.004	0.548	0.329 **
Log likelihood = 296.7716		AIC= -11.94008			

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

表 5-21 VAR-GT 模型

		dlog_pc	dlog_rgdg	cpiyoy	buscyc_fac_1	dgt_fac_1	dgt_fac_2
		估計係數	估計係數	估計係數	估計係數	估計係數	估計係數
_cons		0.009 ***	0.014 ***	-0.171	0.183	0.058	0.046
dlog_pc	L1.	-0.734 ***	0.354	84.947 **	8.282	-10.706 *	4.425
	L2.	-0.004	0.424	12.257	0.088	-16.900 **	6.368
	L3.	-0.100	0.364	-57.118	0.570	-4.895	6.307 *
dlog_rgdg	L1.	0.091	-0.787 ***	-9.379	3.595	-8.409 ***	-7.556 ***
	L2.	-0.197 **	-0.819 ***	30.489	3.676	-0.488	-5.443 ***
	L3.	-0.166 **	-0.608 ***	12.714	9.184 *	2.843	-0.858
cpiyoy	L1.	-0.002 ***	-0.001	0.973 ***	-0.088 *	-0.038	-0.031 **
	L2.	0.000	0.000	0.136	-0.009	-0.028	0.031 *
	L3.	-0.002 **	-0.003 ***	-0.074	-0.106 *	0.023	0.002
buscyc_fac_1	L1.	-0.003	0.007 *	1.440 ***	1.447 ***	0.182 **	0.090 **
	L2.	0.004	0.003	-1.851 **	-1.146 ***	-0.160	-0.049
	L3.	-0.001	-0.006 *	0.862	0.319 **	0.080	-0.008
dgt_fac_1	L1.	-0.009 *	-0.013	1.770	-0.277	-0.089	-0.051
	L2.	-0.009 *	0.005	1.025	0.280	0.006	0.228 **
	L3.	-0.010 *	-0.015 **	-0.880	-0.087	-0.078	-0.060
dgt_fac_2	L1.	0.002	0.016	1.979	0.487	0.535 *	0.514 ***
	L2.	0.006	-0.015	-4.461 **	-0.511	0.466	-0.073
	L3.	0.015	-0.018	1.358	-0.116	0.000	0.247
Log likelihood = 443.3366		AIC = -16.0652					

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

圖 5-11 顯示 VAR 模型中，樣本外動態之預測序列趨勢。

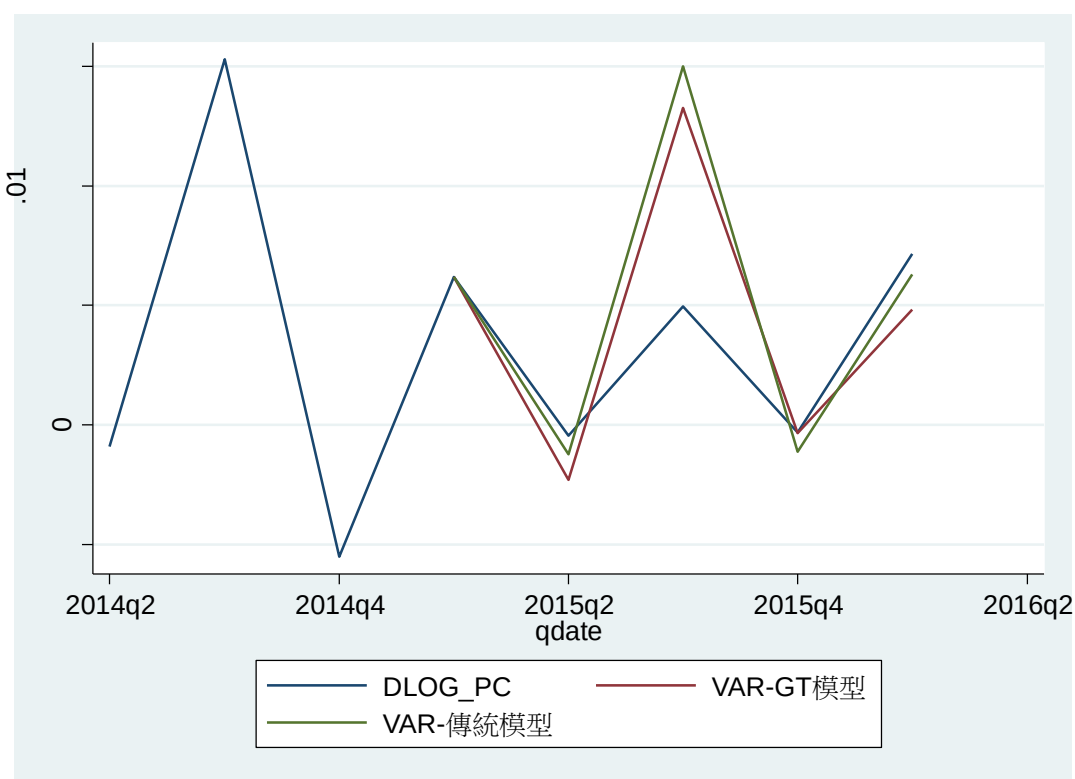


圖 5-11 VAR 樣本外預測趨勢圖

以 MSE、MAE、MAPE 等三指標比較兩模型之預測效果。此外，本文針對研究中所建構的 VAR-GT 模型在 MSE 與 MAPE 指標中，皆顯示其較 VAR-傳統模型具有績效。而在進行 Diebold-Mariano (DM) Test，檢定結果未顯著，顯示無法證明兩模型有顯著差異。

表 5-22 VAR 模型樣本外預測績效比較

Forecast Evaluation statistics	MSE	MAE	MAPE
VAR-傳統模型	0.0000256	0.003114	1.597
VAR-GT 模型	0.0000194	0.003119	1.520
DM Test Statistic: -0.01864			

陸、小結

本節以由十五大電商網之商品別所淬煉出之網路檢索指標，探討網路大數據對於民間消費預測之改善情況。在 ARIMA 分析模組中，透過 AIC 指標挑選最式模型，並建置 ARIMA (4,1,0) Model、ARIMAX (1,1,0)-傳統模型、ARIMAX(3,1,0) GT-Model 及 ARIMAX(1,1,0) GT-Model。且各模組均透過靜態與動態預測評估，並就樣本內及樣本外分別檢視及比較其預測效力。

於 ARIMA 分析下，可發現在樣本內靜態預測中，涵蓋網路檢索指標之 (ARIMAX (3,1,0) GT Model 及 ARIMAX (1,1,0) GT Model) 預測績效多優於 ARIMA Model 及 ARIMAX-傳統模型表現。而在樣本內動態預測的部分，也顯示 ARIMAX(3,1,0)GT Model 有相對不錯之預測績效。惟透過 DM Test 顯示 ARIMAX (1,1,0) GT Model 並未明顯異於 ARIMA (4,1,0) 模型。

而在 ARIMA 樣本外預測評估中，靜態預測模組顯示 ARIMA(4,1,0) 預測績效優於其他三者。在動態預測模組中，則顯示 ARIMA (4,1,0) 與 ARIMAX (1,1,0) 傳統模型各有優劣，但 DM Test 檢定下可發現，其他三模型並未與 ARIMA (4,1,0) 有顯著差異，顯示樣本外模組之預測績效比較未能確切成果。

此外，本研究亦有透過 VAR 樣本外預測，比較 VAR-GT 模型與 VAR-傳統模型之預測績效，結果 VAR-GT 模型較 VAR-傳統模型更具預測績效，惟透過 DM Test 並未能證明兩估計模型有顯著差異。

整體而言，在 ARIMA 模型中，納入網路檢索指標在樣本內的預測績效具有較好的預測效力，無論是透過動態或靜態的預測，且 ARIMAX -GT Model 與 ARIMA 經檢定後顯示其有顯著的差異性。

第三節 民間消費綜合指標預測效力實證研究

本研究為透過多面向的大數據指標，期望建置提升民間消費預測之指標，故透過靜態及動態的主成份分析，分別建置民間消費預測綜合指標，後續同樣透過 ARIMA 及 VAR 模型來進行民間消費的預測。在進行模型檢測前，首先會透過單根檢定，來觀察及確認變項之適合性。

壹、變數說明

為進行預測模型的配適比較，本研究納入一般性總體變數（GDP、銀行利率、消費者物價指數等），而在民間消費綜合指標的部份，透過靜態 PCA 萃取出兩項因子，而在動態 DFM 萃取出一項指標，後續則進行相關比較評估。

表 5-23 變數說明

變數類型	變數中文名稱	資料來源	備註
應變數	台灣民間消費連鎖實質值	主計總處	
總體變數	台灣 GDP 連鎖實質值	主計總處	
	實質銀行業利率一個月期存款牌告利率	中央銀行	
	消費者物價指數年增率	主計總處	
民間消費預測綜合指標 (靜態萃取)	民間消費預測總指標 1	本研究	由總體經濟領先指標、市場信心指標、網路檢索指標等，透過靜態主成分萃取而成，並依據徵值 (Eigenvalue) 與因素解釋總變異量的能力決定因素萃取的個數。
	民間消費預測總指標 2		
民間消費預測綜合指標 (動態萃取)	民間消費預測總指標 dfm	本研究	同樣由上述變數，經由動態因子模型 (DFM)，萃取因子

資料來源：本研究

貳、單根檢定

在進行預測模型評估前，首先針對各變項進行單根檢定，降低非定態之變項特性影響後續評估及預測。惟避免損失樣本及數列特性，本研究模型均以 $I(0)$ 及 $I(1)$ 進行模型設定、模型估計後檢視殘差均為定態。(詳見表 5-24)。

表 5-24 單根檢定

變數中文名稱	變數代碼	ADF 單根檢定			以一階差分處理(v)
		截距	截距與趨勢	無截距與趨勢	
台灣民間消費連鎖實質值	logpc	$I(2)$	$I(2)$	$I(1)$	v
台灣 GDP 連鎖實質值	rgdp	$I(1)$	$I(0)$	$I(1)$	
消費者物價指數年增率	cpiyoy	$I(0)$	$I(0)$	$I(0)$	
民間消費預測總指標(pca) 1	pca_factor1	$I(1)$	$I(0)$	$I(0)$	
民間消費預測總指標(pca) 2	pca_factor2	$I(1)$	$I(1)$	$I(0)$	v
民間消費預測總指標(dfm)	dfm_factor1	$I(0)$	$I(0)$	$I(0)$	

註： $I(x)$ ，表經過 x 階差分後為定態。

參、相關係數矩陣及敘述統計

表 5-25 呈現變數之敘述統計。而透過表 5-26 之相關分析可得知，由 PCA 法所萃取之民間消費綜合指標第一因子 (dpca_factor1) 與 DFM 之因子 (dfm_factor1) 相關係數達 0.730²²。

表 5-25 變數敘述統計

變數	樣本數	平均數	標準差	最小值	最大值
dlog_pc	48	0.0023	0.0099	-0.0163	0.0220
dlog_rgdp	48	0.0038	0.0236	-0.0469	0.0291
cpiyoy	49	1.2626	1.3483	-1.3433	4.5333
dpca_factor1	48	0.0823	0.2677	-0.6981	0.7020
pca_factor2	49	0.0391	1.2250	-3.0262	3.5060
dfm_factor1	48	0.0005	1.2890	-5.5951	2.6530

²² 本文另有測試實質銀行業利率一個月期存款牌告利率，惟其與 CPIYOY 相關係數為-0.926，故模型放入消費者物價指數年增率。

表 5-26 相關係數矩陣

	dlog_pc	dlog_rgdg	cpiyoy	dpca_factor1	pca_factor2	dfM_factor1
dlog_pc	1					
dlog_rgdg	-0.1086	1				
cpiyoy	-0.1682	-0.0601	1			
dpca_factor1	0.1268	0.2122	-0.267	1		
pca_factor2	0.098	0.2000	0.1884	0.3447	1	
dfm_factor1	0.1643	0.2208	0.0516	0.7303	0.5295	1

備註：「d 變數名」表變數經過一階差分處理。

肆、ARIMA 模型分析

本研究透過 Eviews Automatic ARIMA Forecasting 功能，設定 ARIMA (p,d,q)，同前述設定，為避免損失過多樣本，因此，AR (p)、MA (q) 最大階次均設定為 4。

為評估納入網路大數據之民間消費預測綜合指標之預測效果，本文除建構 ARIMAX-PCA 模型與 ARIMAX-DFM 模型，另建構一涵蓋與民間消費相關的傳統經濟變數之 ARIMAX-PCATR 及 ARIMAX-DFMTR 模型。並與本章第二節之 ARIMA 模型進行比較，觀察民間消費預測綜合標是否有效提升預測我國民間消費之準確度，表 5-27 為相關估計結果。

表 5-27 民間消費綜合指標 ARIMA 模型分析

ARIMAX PCA 模型			ARIMAX PCATR 模型			ARIMAX DFM 模型			ARIMAX DFMTR 模型		
ARIMAX (1,1,3)			ARIMAX (4,1,0)			ARIMAX (1,1,3)			ARIMAX (4,1,0)		
變項	估計係數	Std. Error	變項	估計係數	Std. Error	變項	估計係數	Std. Error	變項	估計係數	Std. Error
C	0.00148 ***	0.00022	C	0.00367 ***	0.00082	C	0.00207 ***	0.00043	C	0.00322 ***	0.00076
AR (1)	-0.9106 ***	0.1012	AR (1)	-0.5737 ***	0.1958	AR (1)	-0.9136 ***	0.0981	AR (1)	-0.4524 **	0.1693
			AR (2)	-0.0776	0.2215				AR (2)	0.0608	0.1915
			AR (3)	-0.1111	0.1988				AR (3)	-0.0605	0.1772
			AR (4)	0.3737 **	0.1663				AR (4)	0.4044 **	0.1704
MA (1)	0.452463	208.5334				MA (1)	0.77736	7.537298			
MA (2)	-0.563869	577.0686				MA (2)	-0.1703	3.987701			
MA (3)	-0.888513	656.6669				MA (3)	-0.7585	7.476103			
			DLOGRGDP	0.1854 ***	0.0537				DLOGRGDP	0.1964 ***	0.0427
			CPIYOY	-0.0014 ***	0.0003				CPIYOY	-0.0012 ***	0.0003
DPCA_FACTOR1	0.008137 ***	0.00255	DPCA_FACTOR1	-0.0010	0.0015	DFM_FACTOR	0.00152	0.0009	DFM_FACTOR	0.0002	0.0003
PCA_FACTOR2	0.001	0.00044	PCA_FACTOR2	0.0010 **	0.0005						
R-squared		0.789	R-squared		0.909	R-squared		0.770	R-squared		0.894
Adjusted R-squared		0.747	Adjusted R-squared		0.885	Adjusted R-squared		0.733	Adjusted R-squared		0.870
Log likelihood		170.325	Log likelihood		190.581	Log likelihood		168.284	Log likelihood		187.337

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

一、樣本內預測

針對樣本內預測，本研究依據前述設計之三模型，就靜態與動態模型分別比較，依據表 5-28 樣本內靜態預測績效比較可發現，以納入民間消費綜合指標之 ARIMAX(4,1,0) PCATR 模型及 ARIMAX(4,1,0) DFMTR 模型，其預測效果在 RMSE、MAE、MAPE 及 Theil U 等四指標，皆相較 ARIMA Model 表現為佳，且 ARIMAX(4,1,0) PCATR 模型於 RMSE 指標標準下，也較 ARIMAX-傳統模型表現為佳。

而在樣本內動態預測績效表現可發現，ARIMAX(4,1,0) PCATR 模型之預測效果，於 RMSE、MAE 及 Theil U 等三指標中，皆優於 ARIMA Model 及 ARIMAX-傳統模型。圖 5-12 及 5-13 分別呈現樣本內靜態與動態之預測序列趨勢。

表 5-28 樣本內靜態及動態預測績效比較

Forecast Evaluation statistics	RMSE	MAE	MAPE	Theil U
靜態預測				
ARIMA (4,1,0)	0.0056	0.0043	389.7643	0.2863
ARIMAX (1,1,0) 傳統模型	0.0038	0.0029	115.2349	0.1817
ARIMAX (3,1,0) GT Model	0.0032	0.0026	211.6075	0.1548
ARIMAX (1,1,0) GT Model	0.0037	0.0030	114.4096	0.17976
ARIMAX (1,1,3) PCA 模型	0.0088	0.0073	187.5757	0.5739
ARIMAX (4,1,0) PCATR 模型	0.0037	0.0030	187.8162	0.1892
ARIMAX (1,1,3) DFM 模型	0.0091	0.0076	148.0954	0.6480
ARIMAX (4,1,0) DFMTR 模型	0.0038	0.0031	173.3152	0.1937
動態預測				
ARIMA (4,1,0)	0.0072	0.0059	83.1975	0.4649
ARIMAX (1,1,0) 傳統模型	0.0090	0.0081	288.8734	0.5754
ARIMAX (3,1,0) GT Model	0.0067	0.0057	139.2236	0.4079
ARIMAX (1,1,0) GT Model	0.0090	0.00811	277.7358	0.57862
ARIMAX (1,1,3) PCA 模型	0.0049	0.0038	277.6152	0.2503
ARIMAX (4,1,0) PCATR 模型	0.0031	0.0026	241.2499	0.1496
ARIMAX (1,1,3) DFM 模型	0.0051	0.0041	214.2914	0.2677
ARIMAX (4,1,0) DFMTR 模型	0.0032	0.0027	230.7892	0.1556

註：顏色顯示之模型乃為本章第二節所建置之。

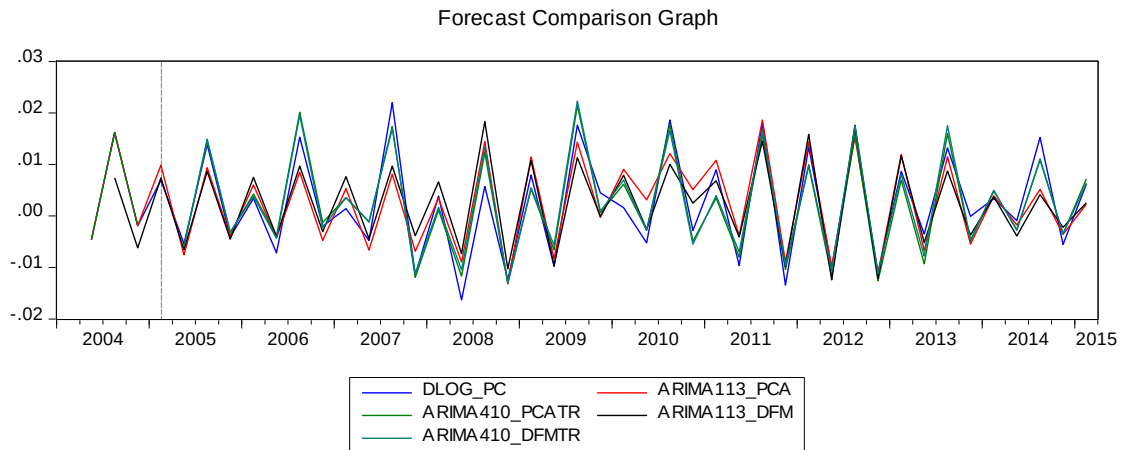


圖 5-12 樣本內靜態預測趨勢圖

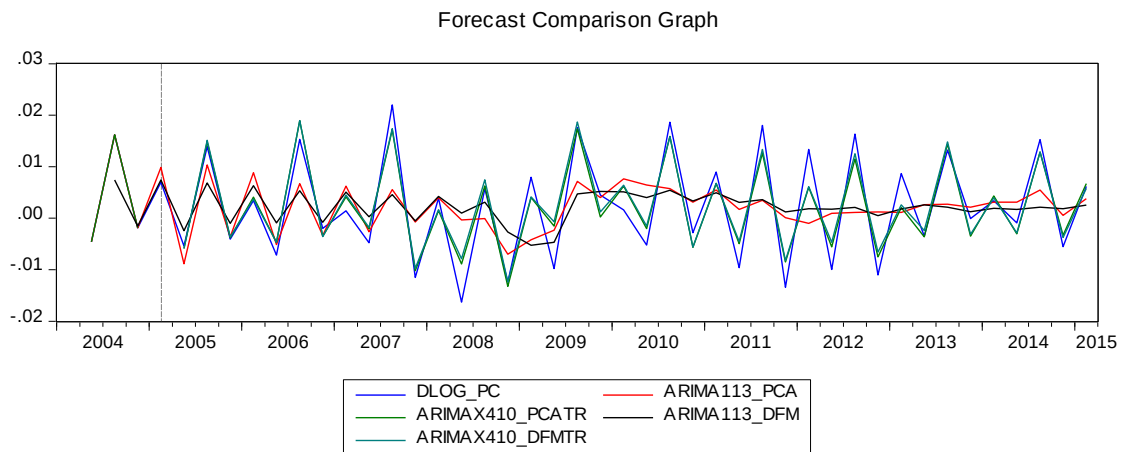


圖 5-13 樣本內動態預測趨勢圖

本文亦針對研究中所建構的 ARIMA 模型與民間消費綜合指標各項模組進行 Diebold-Mariano(DM)Test，檢定結果可發現，各模組與 ARIMA 皆有顯著差異。

表 5-29 民間消費綜合指標 ARIMA 樣本內預測 DM Test

模型	ARIMA (4,1,0)
ARIMAX (1,1,3) PCA 模型	-3.770386 ***
ARIMAX (4,1,0) PCATR 模型	4.649846 ***
ARIMAX (1,1,3) DFM 模型	4.597816 ***
ARIMAX (4,1,0) DFMTR 模型	-4.604661 ***

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

二、樣本外預測

針對樣本外預測，本研究依據本章第二節所設計之預測模組，就靜態與動態模型分別比較。依據表 5-30 樣本外靜態預測績效比較可發現，以納入民間消費綜合指標之 ARIMAX (4,1,0) DFMTR 模型及 ARIMA 模型各有領先，若以 MAPE 觀察，則以 DFMTR 模組預測績效為佳。

而在樣本內動態預測績效表現可發現，以 ARIMA Model 及 ARIMAX—傳統模型各有領先。圖 5-14 及 5-15 分別呈現樣本內靜態與動態之預測序列趨勢。

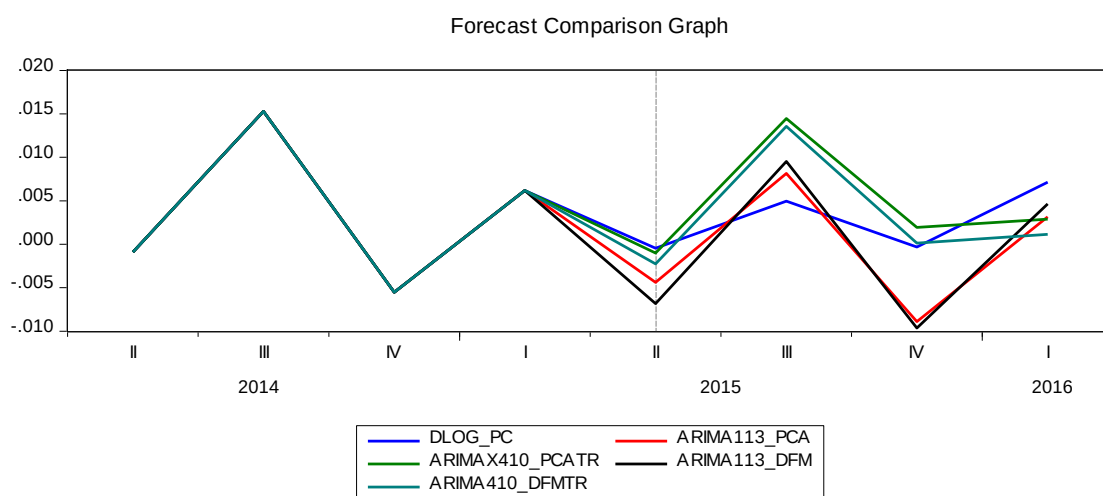


圖 5-14 樣本內靜態預測趨勢圖

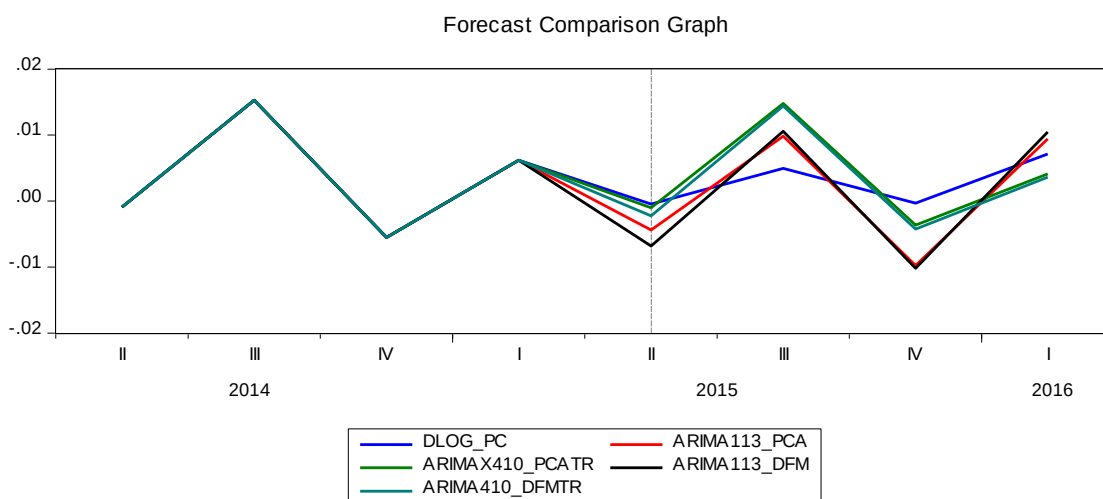


圖 5-15 樣本內動態預測趨勢圖

表 5-30 樣本外靜態及動態預測績效比較

Forecast Evaluation statistics	RMSE	MAE	MAPE	Theil U
靜態預測				
ARIMA (4,1,0)	0.003	0.003	334.635	0.318
ARIMAX (1,1,0) 傳統模型	0.005	0.005	525.090	0.509
ARIMAX (3,1,0) GT Model	0.007	0.006	572.855	0.520
ARIMAX (1,1,0) GT Model	0.006	0.005	631.137	0.5562
ARIMAX (1,1,3) PCA 模型	0.005	0.005	931.976	0.490
ARIMAX (4,1,0) PCATR 模型	0.005	0.004	273.079	0.451
ARIMAX (1,1,3) DFM 模型	0.006	0.006	1126.361	0.506
ARIMAX (4,1,0) DFMTR 模型	0.005	0.004	197.120	0.473
動態預測				
ARIMA (4,1,0)	0.004	0.003	464.803	0.339
ARIMAX (1,1,0) 傳統模型	0.005	0.004	300.378	0.437
ARIMAX (3,1,0) GT Model	0.007	0.005	307.851	0.532
ARIMAX (1,1,0) GT Model	0.005	0.004	367.683	0.480
ARIMAX (1,1,3) PCA 模型	0.006	0.005	1003.813	0.444
ARIMAX (4,1,0) PCATR 模型	0.005	0.004	360.290	0.442
ARIMAX (1,1,3) DFM 模型	0.007	0.006	1178.285	0.480
ARIMAX (4,1,0) DFMTR 模型	0.005	0.005	473.960	0.451

本文亦針對研究中所建構的 ARIMA 模型與民間消費綜合指標各項樣本外預測模組進行 Diebold-Mariano (DM) Test，檢定結果可發現，僅 ARIMAX (1,1,3) DFM 模型檢定結果為顯著，顯示兩者具有差異性。

表 5-31 民間消費綜合指標 ARIMA 樣本外預測 DM Test

模型	ARIMA (4,1,0)
ARIMAX (1,1,3) PCA 模型	-1.12542
ARIMAX (4,1,0) PCATR 模型	-1.13263
ARIMAX (1,1,3) DFM 模型	-1.82079 *
ARIMAX (4,1,0) DFMTR 模型	-1.46228

2. VAR-傳統模型

在建置 VAR-傳統模型，同樣以實質 GDP (drgdp)、消費者物價指數年增率為研究標的，並設定落後 3 期，估計結果見表 5-33 至 5-36。有關 VAR 各模型之因果檢定、衝擊反應及預測誤差結果，見附件八。

表 5-32 VAR 落後期設定

模型	落後期	AIC	log likelihood
VAR-PCA 模型	1	-7.0754	164.121
	2	-7.3887	176.1625
	3*	-8.0798	195.6363
VAR-PCATR 模型	1	-4.0873	93.87642
	2	-4.4228	102.8783
	3*	-4.9732	115.9516
VAR-DFM 模型	1	-10.518	256.1452
	2	-12.287	313.0362
	3*	-12.836	343.1285
VAR-DFMTR 模型	1	-7.1281	173.2536
	2	-8.3994	212.3875
	3*	-9.2705	242.0462

表 5-33 VAR-PCA 模型

		dlog_pc	dpca_factor1	pca_factor2
		估計係數	估計係數	估計係數
_cons		0.005 ***	0.028	-0.164 **
dlog_pc	L1.	-0.620 ***	6.199	10.491 *
	L2.	-0.232	15.809 ***	27.882 ***
	L3.	-0.553 ***	10.996 **	19.797 ***
dpca_factor1	L1.	0.000	-0.083	-0.489
	L2.	0.013 **	0.168	0.236
	L3.	-0.001	0.033	0.000
pca_factor2	L1.	0.014 ***	0.454 ***	1.806 ***
	L2.	-0.022 ***	-0.584 ***	-1.452 ***
	L3.	0.010 ***	0.123	0.455 ***
Log likelihood = 195.6363		AIC = -8.07982		

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

表 5-34 VAR-PCATR 模型

		dlog_pc	dlog_rgdp	cpiyoy	dpca_factor1	pca_factor2
		估計係數	估計係數	估計係數	估計係數	估計係數
_cons		0.010 ***	0.027 ***	0.409	0.043	0.079
dlog_pc	L1.	-0.729 ***	-0.196	98.409 **	5.370	24.205 ***
	L2.	-0.019	-0.793 *	-33.766	15.569 *	25.011 ***
	L3.	-0.156	-0.019	-129.240 ***	24.202 ***	15.183 *
dlog_rgdp	L1.	0.045	-0.753 ***	-17.479	5.578 **	-0.798
	L2.	-0.132 **	-0.738 ***	9.648	-1.532	-5.957 **
	L3.	-0.095	-0.562 ***	27.945 **	3.172	5.447 **
cpiyoy	L1.	-0.003 ***	-0.001	1.063 ***	-0.065 **	-0.124 ***
	L2.	0.001	-0.002	0.104	-0.008	0.065 *
	L3.	-0.002 *	-0.002	-0.358 *	0.030	-0.062
dpca_factor1	L1.	-0.001	-0.006	-1.799 *	0.002	-0.621 ***
	L2.	0.006	0.000	1.699	-0.211	-0.248
	L3.	-0.002	0.002	0.058	0.158	0.037
pca_factor2	L1.	0.004	0.031 ***	3.199 ***	0.238	1.659 ***
	L2.	-0.004	-0.024 *	-4.704 ***	-0.190	-0.879 ***
	L3.	0.002	0.003	1.929 **	-0.061	0.113
Log likelihood =343.1285		AIC = -12.83554				

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

表 5-35 VAR-DFM 模型

		dlog_pc	dfM_factor1
		估計係數	估計係數
_cons		0.004 ***	-0.306 **
dlog_pc	L1.	-0.509 ***	14.405
	L2.	-0.019	77.185 ***
	L3.	-0.443 ***	55.741 ***
dfM_factor1	L1.	0.003 ***	1.090 ***
	L2.	-0.004 **	-0.946 ***
	L3.	0.001	0.295 **
Log likelihood =115.9516		AIC = -4.973248	

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

表 5-36 VAR-DFMTR 模型

		dlog_pc	dlog_rgdp	cpiyoy	dfM_factor1
		估計係數	估計係數	估計係數	估計係數
_cons		0.007 ***	0.017 ***	0.154	-0.141
dlog_pc	L1.	-0.657 ***	0.058	99.838 **	-30.150
	L2.	0.243	0.568	56.974	63.408 **
	L3.	0.096	1.158 ***	-45.706	109.356 ***
dlog_rgdp	L1.	0.077	-0.666 ***	-22.518 *	21.176 ***
	L2.	-0.161 ***	-0.962 ***	-19.131 *	-1.195
	L3.	-0.107 **	-0.808 ***	-1.167	-1.911
cpiyoy	L1.	-0.003 ***	0.000	0.995 ***	-0.036
	L2.	0.001	-0.003	-0.100	-0.361 ***
	L3.	-0.001	-0.002	-0.095	0.236 **
dfM_factor1	L1.	0.001 *	0.005 **	0.238	0.960 ***
	L2.	-0.001	-0.001	0.003	-0.792 ***
	L3.	0.000	-0.001	-0.053	0.294 ***
Log likelihood =242.0462		AIC = -9.270547			

圖 5-16 呈現各樣本外 VAR 預測模組之趨勢圖。

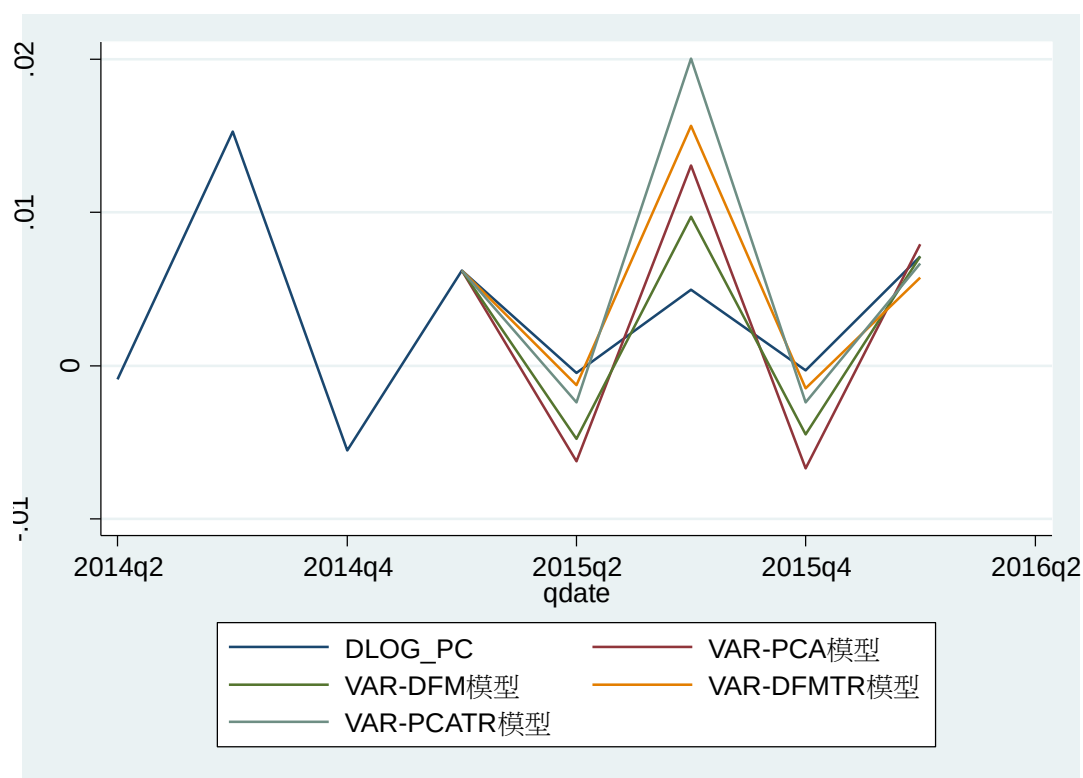


圖 5-16 VAR 樣本外預測趨勢圖

比較各模型之預測效果（詳見表 5-37）在 MSE、MAE、MAPE 等三指標此外，本文針對研究中所建構的 VAR-PCATR 模型在 MSE 指標中，顯示其較 VAR-傳統模型具有績效。然而，於進行 Diebold-Mariano (DM) Test，檢定結果未顯著（詳見表 5-38），顯示無法證明兩模型有顯著差異。

表 5-37 VAR 模型樣本外預測績效比較

Forecast Evaluation statistics	MSE	MAE	MAPE
VAR-傳統模型	0.0000256	0.003114	1.597
VAR-GT 模型	0.0000194	0.003119	1.520
VAR-PCA 模型	0.0000351	0.005264	8.713
VAR-PCATR 模型	0.0000147	0.003312	5.94
VAR-DFM 模型	0.0000588	0.004894	3.506
VAR-DFMTR 模型	0.0000295	0.003522	1.976

表 5-38 VAR 樣本外預測 DM Test

模型	VAR-傳統模型	VAR-GT 模型
VAR-PCA 模型	-4.132 ***	-3.741 ***
VAR-PCATR 模型	-0.3391	-0.3681
VAR-DFM 模型	-3.586 ***	-2.369 **
VAR-DFMTR 模型	-3.677 ***	-1.307

陸、小結

為透過多樣化網路大數據組成指標，探索有效提升民間消費預測之重要因子，故除網路檢索指標之外，本研究建置民間消費綜合指標進行相關研究。民間消費綜合指標乃透過總體經濟領先指標、消費者信心指數及網路檢索指標，分別以靜態及動態的主成份進行淬鍊。

在 ARIMA 分析模組中，本研究透過 AIC 指標挑選最適模型，並建置 ARIMAX-PCA 模型與 ARIMAX-DFM 模型，此外，為據以比較，亦建構一涵蓋與民間消費相關的傳統經濟變數之 ARIMAX-PCATR 及 ARIMAX-DFMTR 模型。且各模組均透過靜態與動態預測評估，並就樣

本內及樣本外分別檢視及比較其預測效力。

於 ARIMA 分析下，可發現在樣本內靜態預測中，ARIMAX (4,1,0) PCATR 模型及 ARIMAX (4,1,0) DFMTR 模型相較 ARIMA (4,1,0) 模型有較好的預測績效。而在樣本內動態預測的部分，ARIMAX (4,1,0) PCATR 模型之預測效果，於 RMSE、MAE 及 Theil U 等三指標中，皆優於 ARIMA Model 及 ARIMAX –傳統模型，且透過 DM Test 進行模型比較，四模組皆與 ARIMA Model 有顯著差異。

而在 ARIMA 樣本外預測評估中，靜態預測模組顯示 ARIMA(4,1,0) 與 ARIMAX(4,1,0)DFMTR 各有優劣。在動態預測模組中，則顯示 ARIMA (4,1,0) 與 ARIMAX (1,1,0) 傳統模型各有優劣。但 DM Test 檢定下發現，ARIMAX (4,1,0) PCATR 模型及 ARIMAX (4,1,0) DFMTR 並未與 ARIMA (4,1,0) 有顯著差異，顯示加入民間消費綜合指標之模組，在樣本外之預測績效與簡單 ARIMA 模型無顯著差異。

此外，本研究亦透過 VAR 樣本外預測，比較四模組與 VAR-傳統模型模型之預測績效，結果發現 VAR-PCATR 模型、VAR-GT 模型、VAR-傳統模型各有優劣。惟透過 DM Test 檢定發現，VAR-PCATR 模型與兩者並未有顯著差異。

整體而言，在 ARIMA 模型中，納入民間消費預測綜合指標在樣本內的預測績效具有較好的預測效力，無論是透過動態或靜態的預測，且其與 ARIMA 經檢定後顯示其有顯著的差異性。

第四節 關鍵社群網站民間消費預測指標分析

在關鍵社群網輿情大數據資訊蒐集上，本研究電商臉書粉絲團經營狀況，觀察較為合適的目標電商，做為資料蒐集依據，本研究彙整 10 個粉絲團(詳見表 5-39 細目)。有鑑於各電商在社群網站經營時間仍在開發期，資訊的累積也相對搜尋指標少，故本研究以 2014 年 6 月至今的日資料去進行整合研究。

表 5-39 關鍵社群網站大數據觀測站

電商名稱	粉絲團名稱	電商名稱	粉絲團名稱
博客來	博客來	PChome 購物	PChome 商店街
GoHappy	GoHappy 快樂購物網	淘寶網	淘寶網官方 Taobao
7-11 集團 ibon	ibon mart 線上購物	Yahoo!購物	Yahoo 奇摩購物中心
momo 富邦購物網	momo 購物網	EHS 東森購物	森森購物網
Myfone 購物	myfone 購物	樂天市場	樂天市場

資料來源：本研究彙整。

社群網站指標因應發展趨勢，近三年各電商方開始經營相關粉絲團，惟考量樣本期間較短，因此，社群網站指標並未納入民間消費預測模型測試。本研究將透過相關性分析及社群網絡分析模組探索社群網站指標對民間消費之相關影響。

於網路大數據指標中的「關鍵社群網站大數據指標」，本研究透過六種視覺化圖表的方式，呈現出社群網站之經營在民間消費之相關表現。其中包含：聲量圖、情緒指標圖、盒鬚圖與分配圖、相關性圖、樹狀圖以及集群網絡圖。

透過聲量圖，可以觀察出消費者是否對該商品有興趣並且在網路上進一步的查詢或做詢問；情緒指標圖，可以看出對該商品的正面或負面評價與喜好程度；盒鬚圖與分配圖，可以看出對該商品的正面或負面評價與喜好程度，且盒鬚圖涵義為五數彙總，針對第一位數、中位數、第三位數、最小值和最大值及超過三個標準差的極端值，有助於進一步的深入分析，各資料的配適狀態；相關性圖，可以看出關鍵字在集群間的相關性；樹狀圖，可以看出個別關鍵字它們彼此的相關性；集群網絡圖，首先透過分群的效果，可以知道哪些商品的搜尋頻率具有集群效應，在集群中又可以發現到，某些商品的搜尋程度彼此的連結度較大。

由於資料是以季的方式呈現，故每家網購粉專僅有 12 筆數值資料，不足以做出預測，因此研究團隊採取視覺化的方式去做分析及商品相關度比對，此數據分析有助於探討消費與模型的連結度，但 8 家網購粉專中的「日本樂天市場-台灣」資料量略為不足，故我們只採用 7 家網購粉專作為分析對象。

最後將 15 大電商購物網站在 Google Trends 追蹤出來的關鍵字，取出 Google Trends 主成份分析中的第一大主成份，平方差累積變異量前 10 名的商品挑選出來，與該網購粉專的集群網絡圖做對應，即可以觀察出電商的用詞與網友在 FB 上的用語不一定相同，所以針對集群網絡圖的複雜性，作出相關係數對應表：Google Trends 主成份分析與 FB 集群網絡圖之商品相關性分析，當集群網絡圖的外圍商品超過 10 個以上，複雜度較高，僅列出線條較粗、商品相關性較高的商品；複雜度低的集群網絡圖，則列出所有相關連的商品。

將各個網購粉專的資料做彙整與資料整理後，目前以七家網購粉專的「月頻率」資料結果，來做社群指標的初步理解與分析，分別為：博客來、Yahoo 奇摩購物網、ibon mart 線上購物、淘寶網、momo 購物網、myfone 購物網、PChome 購物中心。整體分析架構如圖 5-17。以下則以博客來之相關分析結果為例呈現²³，並於小結說明各電商網之主要分析結果。

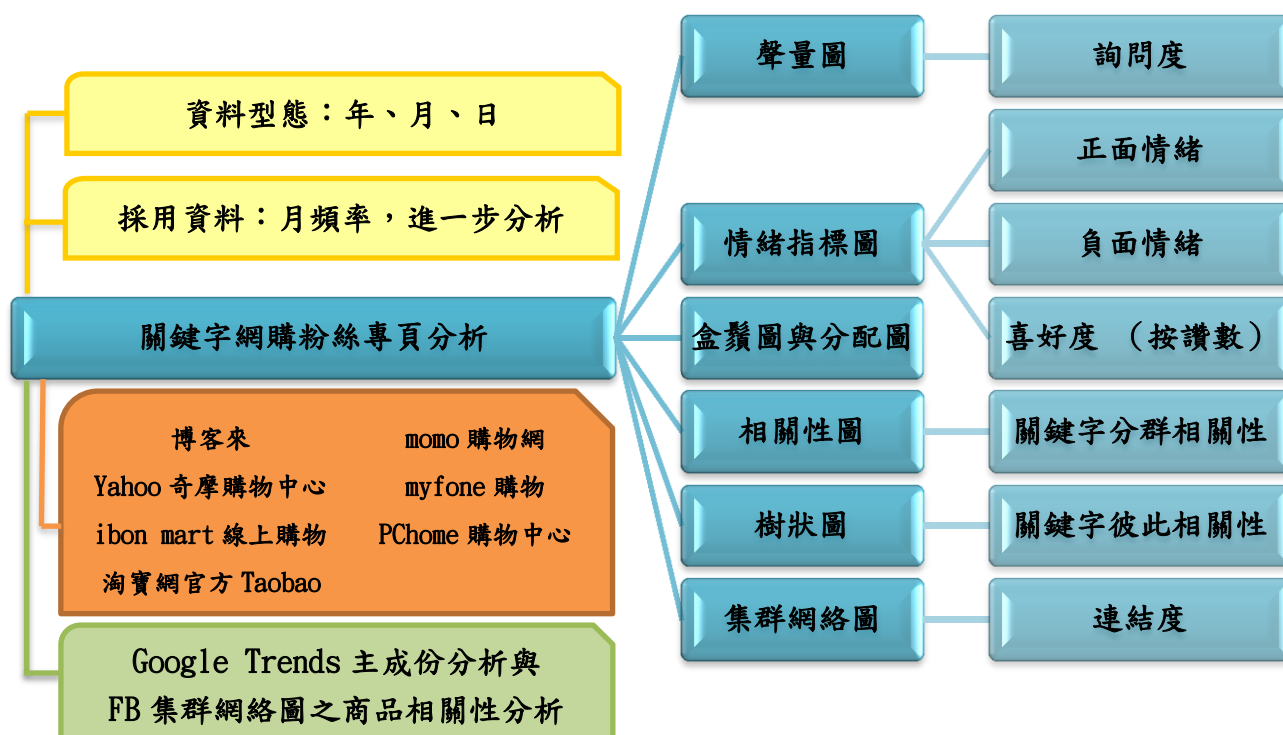


圖 5-17 關鍵社群網站民間消費分析架構

²³ 其他電商網之分析結果，請見本報告之光碟資料檔。

一、博客來聲量時間走勢分析

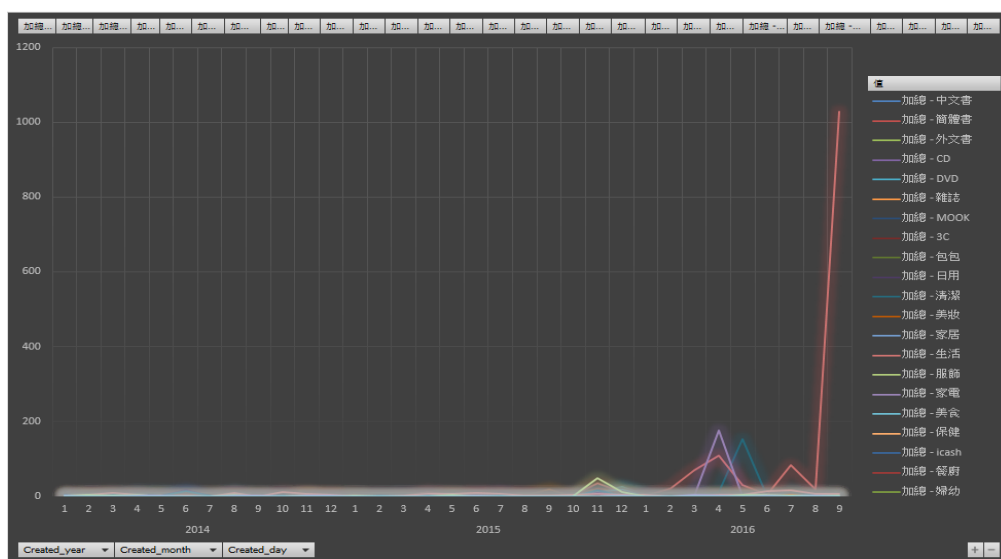


圖 5-18 博客來-聲量圖

在博客來的粉絲專頁中，從聲量圖的月頻率走勢可以發現（見圖 5-18），2016 年 9 月「生活」詢問度最大，高達 1000 次以上，第二大聲量為「家電」，出現在 2016 年 4 月，而第三大聲量為「清潔」。

二、博客來按讚及情緒指標

在博客來的粉絲專頁中，從情緒指標圖的月頻率走勢可以發現（見圖 5-19），最高點的正面情緒分數與負面情緒分數，皆是出現在 2016 年的 9 月，而對於該關鍵字喜好度（網友點擊讚數的頻率）最高峰出現在 2015 年 8 月。

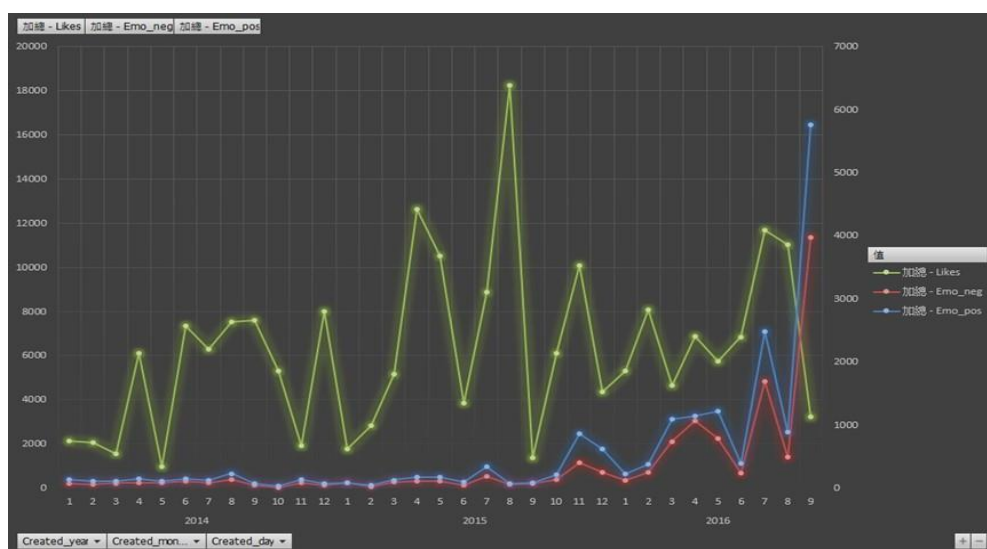


圖 5-19 博客來-情緒指標圖

三、博客來分佈與分配

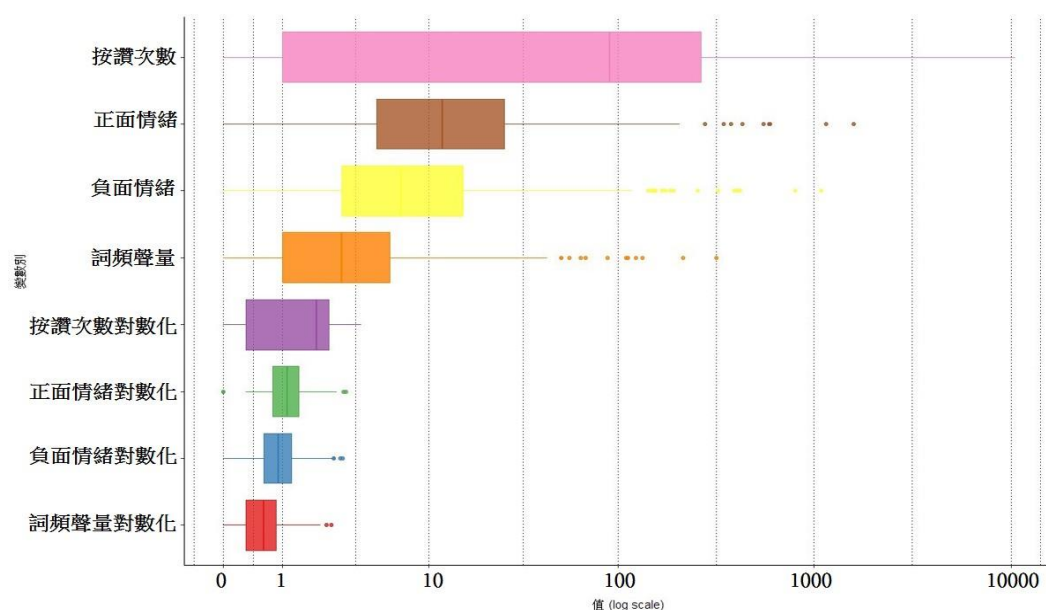


圖 5-20 博客來-盒鬚圖

在博客來的粉絲專頁中，從盒鬚圖的月頻率走勢可以發現（見圖 5-20），「按讚次數」約集中在 1-400 之間，「正面情緒」約集中在 7-40 之間，「負面情緒」約集中在 4-30 之間，「詞頻聲量」約集中在 1-7 之間，「按讚次數對數化」約集中在 0.4-4 之間，「正面情緒數化」約集中在 0.8-2 之間，「負面情緒數化」約集中在 0.6-1.5 之間，「詞頻聲量對數化」約集中在 0.45-0.9 之間。

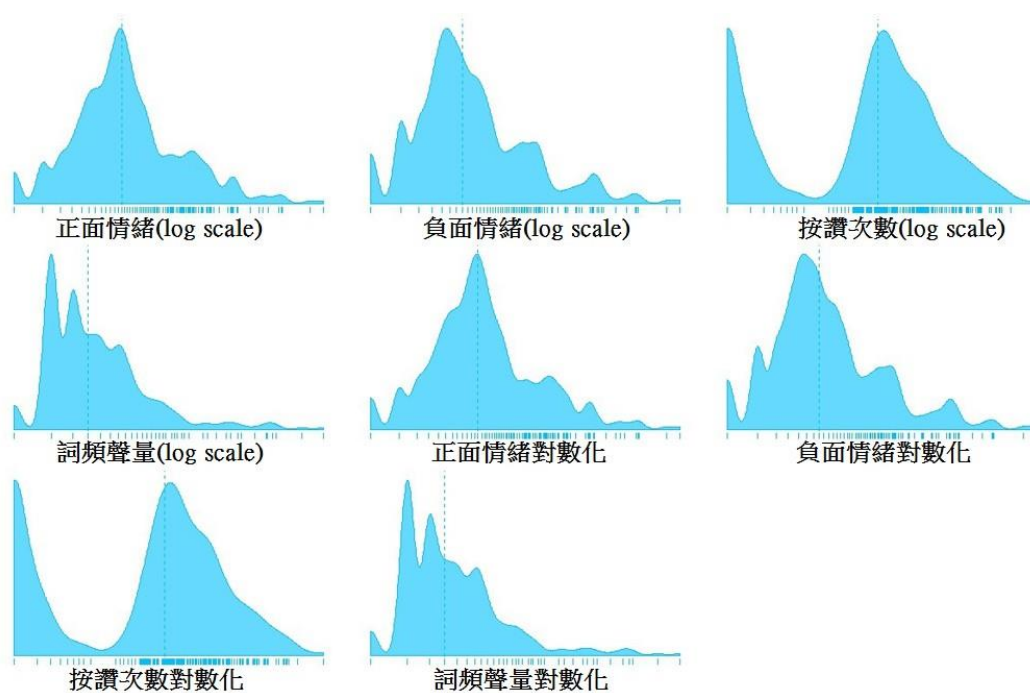


圖 5-21 博客來-分配圖

在博客來的粉絲專頁中，從分配圖的月頻率走勢可以發現（見圖 5-21），「正面情緒」、「負面情緒」、「詞頻聲量」、「正面情緒對數化」、「負面情緒對數化」、「詞頻聲量對數化」皆為右尾分配，另外「按讚次數」、「按讚次數對數化」皆為雙峰分配，資料型態呈現出來的多為右尾分配，表示多數不為常態分配，且資料顯示為右尾分配，可以從分配圖的結果知道該筆資料是否為常態的集群分配。

四、博客來相關性、集群、網絡分析

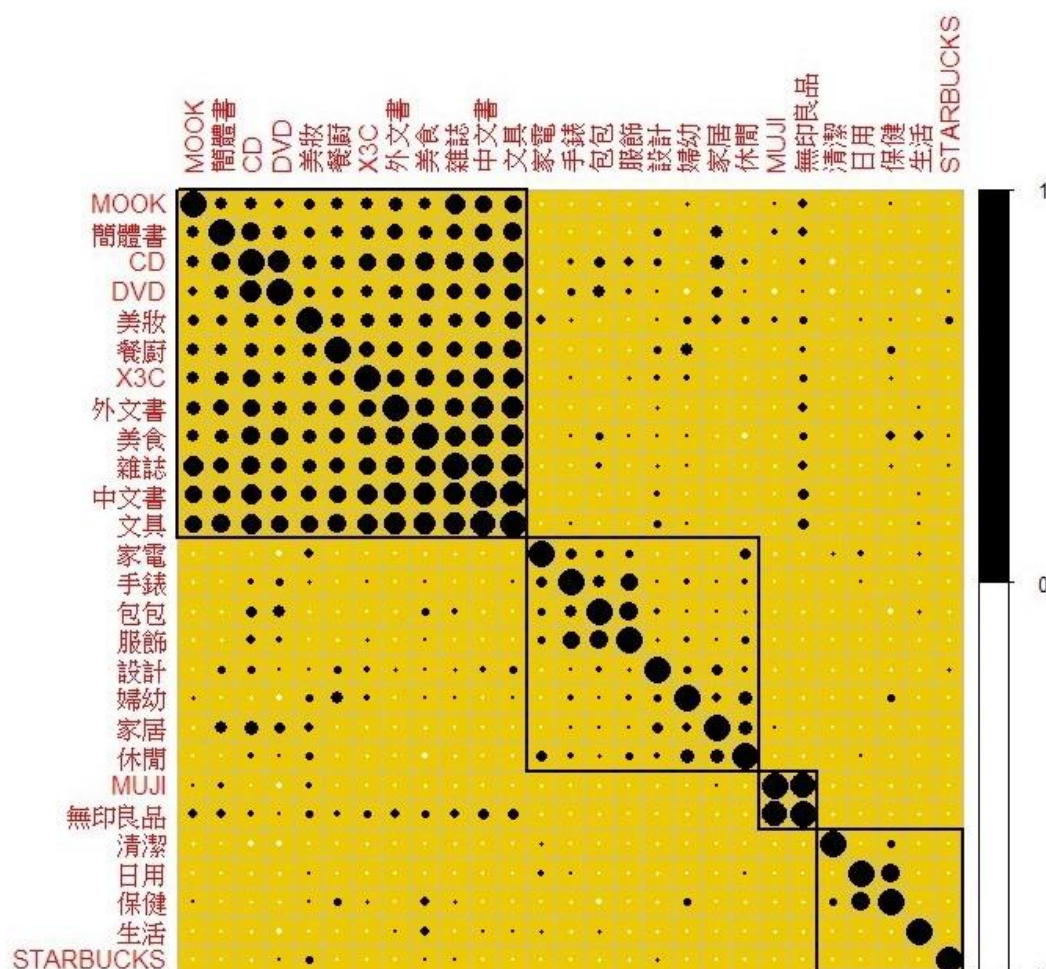


圖 5-22 博客來-相關性圖

在博客來的粉絲專頁中，從相關性圖的月頻率視覺化圖可以發現（見圖 5-22），以「MOOK」、「簡體書」、「CD」、「DVD」、「美妝」、「餐廚」、「3C」、「外文書」、「美食」、「雜誌」、「中文書」、「文具」為同一群組，其中「文具」對其他關鍵字最具相關性。

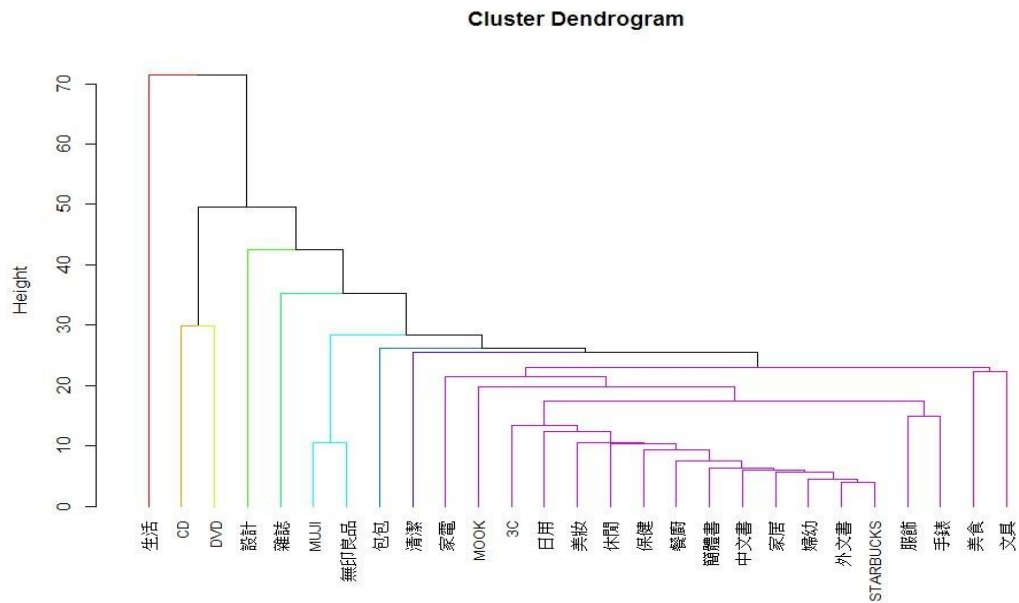


圖 5-23 博客來-樹狀圖

在博客來的粉絲專頁中，從樹狀圖的月頻率視覺化圖可以發現（見圖 5-23），以「生活」做為關鍵字樹狀圖的主軸，關鍵字樹狀圖中的內圈為「3C」、「日用」、「美妝」、「休閒」、「餐廳」、「簡體書」、「中文書」、「家居」、「婦幼」、「外文書」、「STARBUCKS」，而其中「外文書」、「STARBUCKS」和其他關鍵字最具相關。

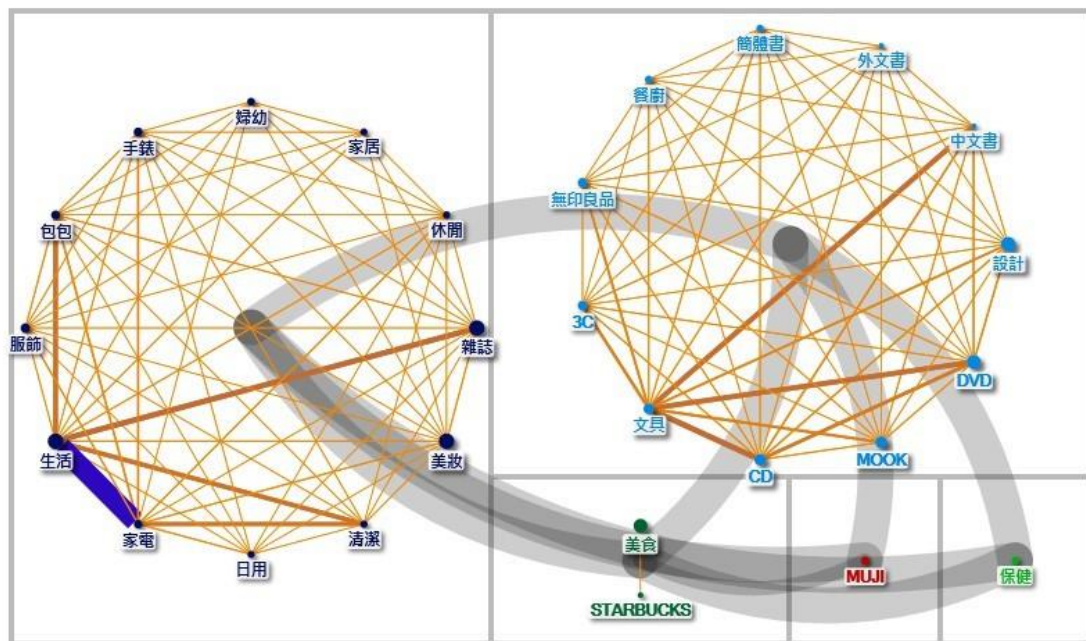


圖 5-24 博客來-集群網絡圖

在博客來的粉絲專頁中，從集群網絡圖的月頻率視覺化圖可以發現（見圖 5-24），該資料被分為二大群組，群組一以「生活」為主點，和「包包」、「雜誌」、「家電」、「清潔」連結性較高，並且與「家電」的連結性最高，群組二則以「文具」為主點，和「3C」、「中文書」、「DVD」及「CD」的連結性較高。

表 5-40 博客來-Google Trends 主成份分析與 FB 集群網絡圖之商品
相關性分析

Boo-博客來					
商品	GT 主成份 平方差累 積變異量	網絡相關性 1	網絡相關性 2	網絡相關性 3	網絡相關性 4
CD	.948	<u>文具</u>	<u>DVD</u>		
家居	.947	無			
保健	.946	無			
DVD	.929	<u>文具</u>	<u>CD</u>		
STARBUCKS	-.922	無			
包包	.898	<u>生活</u>			
文具	.898	<u>中文書</u>	<u>CD</u>	<u>DVD</u>	<u>3C</u>
生活	.892	<u>家電</u>	<u>清潔</u>	<u>雜誌</u>	<u>包包</u>
設計	.835	無			
雜誌	.819	<u>生活</u>			

將博客來在 Google Trends 主成份中，第一大主成份平方差累積變異量前 10 大商品，與 FB 集群網絡圖（見圖 5-24）做對應，因為該網站的集群網絡圖複雜度較高，僅列出線條較粗、商品相關性較高的商品，可以發現到第一主成分中的「CD」平方差累積變異量為 0.948，為最高的累積變異量，對應到網絡圖中的「文具」、「DVD」；「DVD」平方差累積變異量為 0.929，對應到網絡圖中的「文具」、「CD」；「包包」平方差累積變異量為 0.898，對應到網絡圖中的「生活」；「文具」平方差累積變異量為 0.898，對應到網絡圖中的「中文書」、「CD」、「DVD」、「3C」；「生活」平方差累積變異量為 0.892，對應到網絡圖中的「家電」、「清潔」、「雜誌」、「包包」；「雜誌」平方差累積變異量為 0.819，對應到網絡圖中的「生活」，當電商的用詞與網友在 FB 上的用語對應上時，可以知道該商品具有強烈的集群效應，因此可以推論出當商品在 Google Trends 做迴歸有影響力，且在 FB 時仍有商品關聯，代表可以提升消費預測能力。

本研究透過博客來之分析結果為範例操作，其它電商網頁之分析請見本報告之光碟資料檔。

壹、小結

分析七家網購的粉絲專頁「月頻率」資料結果：博客來、Yahoo 奇摩購物網、ibon mart 線上購物、淘寶網官方 Taobao、momo 購物網、myfone 購物、PChome 購物中心，本研究以視覺化圖表方式可以歸納出消費者對該商品的搜尋行為分析。

從聲量圖來看（見表 5-41），可以觀察出網友對七家粉絲專頁的商品詢問度高低，各家網購擁有自己的特色商品，因此消費者有商品需求時，會依據網站的特色去做進一步的搜尋，可以觀察到消費者在「博客來」與「PChome 購物中心」較常搜尋的關鍵字皆為「生活」。

表 5-41 七家網購的粉絲專頁-聲量圖之綜合結果

粉絲專頁	關鍵字	詢問度最高之時間	粉絲專頁	關鍵字	詢問度最高之時間
博客來	生活	2016/9	momo 購物網	團購美食	2014/7
Yahoo 奇摩購物中心	童鞋	2014/11	myfone 購物	手機	2016/5
ibon mart 線上購物	icash	2015/4	PChome 購物中心	生活	2015/3
淘寶網官方 Taobao	辦公	2013/3			

從情緒指標圖來看（見表 5-42），可以觀察出網友對七家粉絲專頁的情緒喜好，大多網站的正面情緒、負面情緒、喜好度皆在相同時間出現最高峰，像是「momo 購物網」、「PChome 購物中心」。

表 5-42 七家網購的粉絲專頁-情緒指標圖之綜合結果

粉絲專頁	正面情緒分數最高之時間	負面情緒分數最高之時間	喜好度（按讚數）最高時間	粉絲專頁	正面情緒分數最高之時間	負面情緒分數最高之時間	喜好度（按讚數）最高時間
博客來	2016/9	2016/9	2015/8	momo 購物網	2014/7	2014/7	2014/7
Yahoo 奇摩購物中心	2014/11	2014/8	2015/1	myfone 購物	2015/12	2014/9	2014/9
ibon mart 線上購物	2014/11	2014/8	2014/8	PChome 購物中心	2015/3	2015/3	2015/3
淘寶網官方 Taobao	2014/10	2014/1	2015/10				

從盒鬚圖與分配圖來看（見表 5-43 及 5-44），可以觀察出網友對七家粉絲專頁的正面情緒、負面情緒、按讚次數、詞頻聲量，集中的區間與配適狀況，是以「聲量圖」和「情緒指標圖」為基礎再做進一步的分析，大多網購的各種指標皆是以右尾分配、雙峰分配為主。

表 5-43 七家網購的粉絲專頁-盒鬚圖與分配圖之綜合結果-1

粉絲專頁	正面情緒 集中區間 配適狀況	負面情緒 集中區間 配適狀況	粉絲專頁	正面情緒 集中區間 配適狀況	負面情緒 集中區間 配適狀況
博客來	7-40 右尾分配	4-30 右尾分配	momo 購物網	7-200 雙峰分配	0-40 右尾分配
Yahoo 奇摩購物中心	9.5-150 雙峰分配，且右肥尾	3-400 雙峰分配，右峰居高	myfone 購物	8-300 雙峰分配	1-40 雙峰分配，右峰居多
ibon mart 線上購物	4.5-10 右尾分配	1-6 雙峰分配，右峰居高	PChome 購物中心	8-60 常態分配	3-40 雙峰分配，左峰居多
淘寶網官方 Taobao	1-12 雙峰分配，左峰居高	0-9 雙峰分配，右峰居高			

表 5-44 七家網購的粉絲專頁-盒鬚圖與分配圖之綜合結果-2

粉絲專頁	按讚次數 集中區間 配適狀況	詞頻聲量 集中區間 配適狀況	粉絲專頁	按讚次數 集中區間 配適狀況	詞頻聲量 集中區間 配適狀況
博客來	1-400 雙峰分配	1-7 右尾分配	momo 購物網	1-7 右尾分配	4-80 右尾分配
Yahoo 奇摩購物中心	0-70 右尾分配	5.3-7 雙峰分配	myfone 購物	0-80 右尾分配	4-120 右尾分配
ibon mart 線上購物	20-70 右尾分配	1-3 右尾分配	PChome 購物中心	0-10 右尾分配	1-9 雙峰分配，右峰居高
淘寶網官方 Taobao	0-3.5 雙峰分配	0-3 雙峰分配，左峰居高			

從相關性圖來看（見表 5-45），可以觀察出網友對七家粉絲專頁的商品詢問度高低，當關鍵字被高度搜尋時，相關文字也將連帶的被高度搜尋，因此可以推斷出消費者在做消費時，哪些商品的需求可能具有連帶關係，實際消費時可能會將該同質性商品同時購買。

表 5-45 七家網購的粉絲專頁-相關性圖之綜合結果

粉絲專頁	關鍵字高度搜尋，彼此間的關鍵性強烈	粉絲專頁	關鍵字高度搜尋，彼此間的關鍵性強烈
博客來	MOOK、簡體書、CD、DVD、美妝、餐廚、3C、外文書、美食、雜誌、中文書、文具	momo 購物網	精品、生活用品、健康食品
Yahoo 奇摩購物中心	家電、視聽、電玩	myfone 購物	保養、戶外、香氛、精品、影音、居家、美妝、婦幼、食品、保健
ibon mart 線上購物	紅利點數、網路遊戲、加油站、證券、第三方支付、銀行、電信、保險	PChome 購物中心	日用、3C、周邊、數位
淘寶網官方 Taobao	汽車、家電、戶外、用品		

從樹狀圖來看（見表 5-46），可以觀察出網友對七家粉絲專頁的商品相關性，各家網購以不同的關鍵字作為主軸，再從樹狀圖發散出更多相關的商品，將關鍵字做分群以達到精簡資料，做進一步的分析，各個網購業者有自己的特色商品，其中「momo 購物網」、「myfone 購物」同時都出現「傢俱」與其他商品的連結。

表 5-46 七家網購的粉絲專頁-樹狀圖之綜合結果

粉絲專頁	關鍵字最具相關性	粉絲專頁	關鍵字最具相關性
博客來	外文書、STARBUCKS	momo 購物網	美妝、保健食品、傢俱、團購美食
Yahoo 奇摩購物中心	戶外、票券	myfone 購物	內著、電腦週邊、傢俱
ibon mart 線上購物	數位、悠遊卡	PChome 購物中心	周邊、鞋子
淘寶網官方 Taobao	母嬰、箱包		

從集群網絡圖來看（見表 5-47），可以觀察出網友對七家粉絲專頁的商品連結性高低，在關鍵字相互影響下，哪些商品具有集群效果，哪些商品的連結性較高，哪些關鍵字的連結性最強烈，大多的網購業

者，皆是以 3C 產品為主軸，再與其他周邊商品做連結。

表 5-47 七家網購的粉絲專頁-集群網絡圖之綜合結果

粉絲專頁	關鍵字連結性最高	粉絲專頁	關鍵字連結性最高
博客來	生活、家電	momo 購物網	家電、3C
Yahoo 奇摩購物中心	演唱會、音樂	myfone 購物	平板、OUTLET、手機 平板、手機
ibon mart 線上購物	週邊、童鞋	PChome 購物中心	生活、3C
淘寶網官方 Taobao	用品、家電		

從 Google Trends 主成份分析與 FB 集群網絡圖之商品相關性分析來來看(見表 5-48)，可以觀察出電商的用詞與網友在 FB 上的用語不一定相同，當兩者的用語對應上時，可以知道該商品具有強烈的集群效應且相互影響力強；兩者用語未對應上時，代表該商品較無集群效應，綜合七家粉絲專頁的商品集群結果，可以觀察出「3C」、「生活」、「家電」這三項商品的集群效應較好。

表 5-48 七家網購的粉絲專頁-Google Trends 主成份分析與 FB 集群網絡圖之商品相關性分析表之綜合結果

粉絲專頁	商品	平方差累積變異量	Google Trends 主成份分析與 FB 集群網絡圖之商品-網絡相關性(僅列出明顯集群性高的相關商品)
博客來	CD	.948	文具、DVD
	DVD	.929	文具、CD
	包包	.898	生活
	文具	.898	中文書、CD、DVD、3C
	生活	.892	家電、清潔、雜誌、包包
	雜誌	.819	生活
Yahoo 奇摩購物中心			無
ibon mart 線上購物			無
淘寶網官方 Taobao	戶外	.854	家電、用品
momo 購物網	3C	.677	家電
myfone 購物	配件	.965	手機
	手機平板	-.874	Outlet、平板、手機
PChome 購物中心	生活	.928	3C

第六章 結論

壹、研究結論

在網路世代，民眾食、衣、住、行皆離不開網路使用，而使用行為的資訊累積數量也日益龐大。近年有相當多的研究，透過全球使用率最高的檢索系統 Google 所釋出的 Google Trends 資訊，作為相關民間消費、房市、股市等各面向的預測基準資訊，主要原因在於其具有相對即時性，透過週資料的分析，比過去月頻率總體經濟指標更能掌握市場變動，如：Choi and Varian (2012) 分析汽車銷售與失業補助等；Herzog (2014) 研究公司資產價格及股價；Toth and Hajdu (2012) 分析房市、車市等。而透過 Google Trends 對民間消費進行預測之研究也相當多 (Vosen and Schmidt, 2011；Kholodilin, Podstawski, and Siliverstovs, 2010；Vosen and Schmidt, 2012)，顯示檢索指標的預測研究驗證已有相當之累積。

本研究透過主成份及動態因素模型，萃取出網路檢索指標及民間消費綜合指標，並施以時間序列預測分析模組 ARIMA 與 VAR，探討納入網路大數據資訊對民間消費預測的相關效益。

其中，檢索指標的建構乃透過國內 15 大電商網站內商品分類項目，作為檢索指標的建構內涵，彙整資料總量達 1838 類商品檢索項目時間序列資訊，再依照各電商網各別進行成份萃取，建置各電商的商品檢索數列指標。而民間消費綜合指標則是以總體經濟領先指標、消費者信心指數及網路檢索指標萃取而成。

在網路檢索指標對民間消費的預測研究部分，結果發現在 ARIMA 模型中，納入網路檢索指標在樣本內的預測績效較佳，無論是透過動態或靜態的預測，且透過 DM 檢定，發現其與 ARIMA 模型有顯著的差異性。而在樣本外預測模組中，模型間並無顯著差異。

而於民間消費綜合指標的預測研究部分，結果發現同樣在樣本內預測中，納入民間消費預測綜合指標之模型相較於自迴歸模組有較佳的預測能力。

整體而言，在樣本內預測模組中，考慮網路檢索指標、民間消費預測綜合指標之民間消費預測模型是相較於自迴歸模型，在模型配適度上更好，惟受限於檢索資料的累積是由 2004 年起，樣本數有限，

未來在資訊不斷累積下，期望能透過多樣態模型檢視及開發檢索指標在民間消費預測的效益。此外，檢索資訊雖具有即時性，可助於瞭解民間消費行為，惟可能有較多的雜訊，非資料可呈現，如最後是否購買，亦可能影響模型預測能力，此為大數據資料分析下所需特別注意之處。

而針對關鍵社群網站對民間消費預測之分析，受限於資訊累積量的不足，多數關鍵社群網站經營消費粉絲團多由 2014 年起，未能透過前述時間序列之預測分析。本研究目前透過聲量、情緒、分配及分佈、網絡、樹狀、相關圖等視覺化分析模組對其進行相關探索，並將 15 大電商購物網站在 Google Trends 追蹤出來的關鍵字，取出 Google Trends 主成份分析中的第一大主成份，平方差累積變異量前 10 名的商品挑選出來，與該網購 FB 粉專的集群網絡圖做對應，觀察不同社群網站的銷售行為及消費者反映情況，期望待未來資料量累積，則可納入消費預測分析之。

貳、未來研究建議

- 一、目前本研究以 15 大電商網商品類目建構之檢索關鍵字，並進行後續研究。受限於研究時程，未能評估不同關鍵字變化下的預測能力，未來研究建議可透過更多樣化的關鍵字建構模式，探索不同組成內容的檢索指標對民間消費的預測效果。
- 二、目前本研究受限於研究期程，未能處理強外生性（strong exogeneity）問題，未來則可透過外生性判讀，期望能夠強化考量網路大數據的民間消費預測指標效益，作為政策評估及建議的重要基準。
- 三、目前本研究為使用主成份及動態因素模型萃取相關指標，未來建議可透過更多樣化的分析模組，如 Ridge regression、LASSO（Least Absolute Shrinkage and Selection Operator）等，探索及比較不同具有預測效度的綜合指標。
- 四、目前配合檢索指標建置民間消費預測指標，時間起始點設定於 2004 年起，且目標項目民間消費為季資料，樣本數有限。未來研究建議以混和頻率的計量方法（如 Giannone et al., 2008），運用高頻月資料（甚至是周資料）預測低頻的季資料（民間消費），可進一步充分運用網路大數據資料及時提供最新訊息的優點。

參、研究限制

- 一、透過關鍵字設定進行的檢索指標及社群網路消費資訊同樣為時間序列的方法進行相關分析，指標具有資訊累積的即時性，然而在分析上，為捕捉搜尋關鍵字的相關性，並非用於解釋搜尋行為的相關性。搜尋行為相關性為針對個人之各項資訊檢索行為來建置，目前在技術上非本研究所能克服。
- 二、目前本研究建置消費型檢索指標，消費類別的分類採用電商網之商品項目，惟商品類項目較難以涵蓋服務類，然而，若要採用國民所得帳類別項目來做民間消費項目別來建置檢索指標，則恐有未能配合消費者使用習慣的問題，如餐廳及旅館，範疇過廣，非本研究在研究期程內所能彙整，期望透過未來研究加以評估。
- 三、由於社群網站消費資訊累積量，受限於其發展，僅能就 2014 年起進行分析，就消費預測上資料量仍顯不足，待未來資料累積量較足夠，方能納入消費預測分析模組中。

參考文獻

一、中文文獻

- 林玳瑩 (2014)。台灣民間消費決定因素之研究。台灣大學經濟學碩士論文，未出版。
- 林俊宏 (譯)，Mayer-Schonberger, Viktor and Kenneth Cukier (原著) (2013)。《大數據》。台北：天下。
- 洪淑玲 (2010)，「信用卡簽帳金額與總體經濟因素關聯性之研究」碩士論文，世新大學金融學研究所。
- 陳旭昇 (2014)。時間序列分析-總體經濟與財務金融之應用。台北：東華書局。
- 陳畔麗 (2007)，「台灣民間消費成長潛力與政策研究」，綜合規劃研究：96 及 97 年，行政院經濟建設委員會，277-296。
- 劉欣姿 (2012)，「使用 Google 搜尋工具改善臺灣消費模型推估能力」，《經濟研究》，第 15 期，頁 129-144。
- 詹惠珠 (2016)，「15 檔寶可夢概念 喊 GO」，《經濟日報》，8 月 7 日。<http://udn.com/news/story/7251/1878415>。
- 聯合報 (2016)。〈台灣人愛滑手機 每天逾 3 小時 全球第一〉，《聯合報》，7 月 21 日。<http://goo.gl/UaPu8v>。
- 動腦新聞 (2015)。〈2016 消費者消費型態 虛實無界限〉，12 月 28 日。<http://goo.gl/CE80q4>。
- INSIDE (2015)。〈電商數據平台 EagleEye 公佈 2014Q4 台灣購物網站排行榜〉，《INSIDE》，2016/06/25。<http://goo.gl/jLUOqt>。

二、英文文獻

- Askatas, Nikos; Zimmermann, Klaus F. (2009) “Google econometrics and unemployment forecasting. German Council for Social and Economic Data.” (RatSWD) Research Notes 41.
- Bai, Jushan, and Serena Ng. (2002). “Determining the number of factors in approximate factor models.” *Econometrica*, 70.1: 191-221.
- Boik, Andre, Shane Greenstein, and Jeffrey Prince. (2016) . “The

- Empirical Economics of Online Attention.” website: <http://goo.gl/hkIyoH>. [assessed July 02,2016].
- Burke, Mary A., and Ali Ozdagli. (2013) . “Household inflation expectations and consumer spending: evidence from panel data .” No. 13-25. Working Papers, Federal Reserve Bank of Boston.
- Choi, Hyunyoung, and Hal Varian. (2012) . “Predicting the present with google trends.” *Economic Record*, 88 (s1) :2-9.
- Davenport, Thomas H., and Gilbert J. Probst. (2002) . Knowledge management case book: Siemens best practices. John Wiley & Sons, Inc..
- Da, Zhi, Joseph Engelberg, and Pengjie Gao. (2011) “In search of attention.” *The Journal of Finance* 66.5: 1461-1499.
- DePaolo, Concetta A., and Kelly Wilkinson. (2014) . “Recurrent online quizzes: Ubiquitous tools for promoting student presence, participation and performance.” *Interdisciplinary Journal of E-Learning and Learning Objects*, 10:75-91.
- Hamid, Alain, and Moritz Heiden (2015) . “Forecasting Volatility with Empirical Similarity and Google Trends.” *Journal of Economic Behavior & Organization*, No. 117: 62–81.
- Herzog, Bodo. (2014) . “Google’s search data and its Application in finance.” *Sciences*, 2 (1) :1-7.
- Godfrey, Daniel., Caley Johns, Meyer, Carol Sadek, Carl Meyer, and Shaina Race (2014) . “A Case Study in Text Mining: Interpreting Twitter Data From World Cup Tweets.” arXiv preprint arXiv:1408.5427.
- Kantrowitz, Alex (2015) . “Facebook Takes Big Step Forward On Commerce, Builds Shops Into Pages.” BuzzFeedNews. (accessed February 1, 2016) .
- Kholodilin, Konstantin A., Maximilian Podstawski, and Boriss Siliverstovs. (2010) “Do Google searches help in nowcasting

- private consumption? A real-time evidence for the US.” KOF Swiss Economic Institute working Paper 256.
- Matsumoto, Azusa, Kohei Matsumura, and Noriyuki Shiraki (2013) . “Potential of Search Data in Assessment of Current Economic Conditions. ” No. 2013-04-18. Bank of Japan.
- McLarren, N. and Shanbhogue, R. (2011) “Using Internet Search Data as Economic Indicators,” Bank of England Quarterly Bulletin, Bank of England, volume 51, No.2, pp.134-140.
- Pavlicek, Jaroslav (2014) *Google Econometrics: Unemployment in Visegrad Countries*. Unpublished doctoral dissertation, Charles University Faculty of Social Sciences Institute of Economic Studies Prague.
- Sims, Christopher A. (1980) “Macroeconomics and reality.” *Econometrica: Journal of the Econometric Society* : 1-48.
- Vosen, Simeon, and Torsten Schmidt. (2011) “Forecasting private consumption: survey-based indicators vs. Google trends.” *Journal of Forecasting*, 30.6: 565-578.
- Simon, Herbert A. (1971) . *Designing Organizations for an Information-Rich World*. Computers, Communication, and the Public Interest, ed. Martin Greenberger. The Johns Hopkins Press.
- Stock, James H., and Mark W. Watson. (2002) “Macroeconomic forecasting using diffusion indexes.” *Journal of Business & Economic Statistics*, 20.2: 147-162.
- Stock, James H., and Mark W. Watson. (2005) “Implications of dynamic factor models for VAR analysis.” National Bureau of Economic Research, No. w11467.
- Toth, Istvan Janos, and Miklós Hajdu. (2012) “Google as a tool for nowcasting household consumption: estimations on Hungarian data.” 31th CIRET Conference, Vienna. [https:// www. ciret. org/ conferences/ vienna](https://www.ciret.org/conferences/vienna).

Vosen, Simeon and Torsten Schmidt (2012) . “A monthly consumption indicator for Germany based on Internet search query data.” Applied Economics Letters 19.7: 683-687.

三、網站資訊

(一) 大數據擷取資料網

Google Trends : <https://www.google.com.tw/trends/>

「fbCrawler」: <http://informaship.cf/>

(二) 15 大購物網

Boo-博客來 : <http://www.books.com.tw/>

Cos-Costco : <https://warehouses.costco.com.tw/>

Etm-東森購物 : <http://www.etmall.com.tw/Pages/Home.aspx>

Fam-FamiPort : <http://www.famiport.com.tw/index.asp>

Goh-GoHappy : <https://www.gohappy.com.tw/>

Ibo-ibon : <http://www.ibon.com.tw/>

Mom-momo : <http://www.momoshop.com.tw/main/Main.jsp>

Myf-myfone : <http://www.myfone.com.tw/buy/>

Pch-PChome 線上購物 : <http://shopping.pchome.com.tw/>

購物中心 : <http://24h.pchome.com.tw/>

Rak-樂天市場 : <http://www.rakuten.com.tw/>

Rut-露天拍賣 : <http://www.ruten.com.tw/>

Tao-淘寶網 : <https://world.taobao.com/>

Yah 1.Yah 2-Yahoo 購物中心 : <https://tw.buy.yahoo.com/>

Yah b-Yahoo 拍賣 : <https://tw.bid.yahoo.com/>

Yah s-Yahoo 服務+ : <https://tw.serviceplus.yahoo.com/>

附件一 網路搜尋關鍵字設定討論會議會議紀錄

「中央銀行 105 年委託研究計畫-應用大數據提升台灣民間消費預測」

網路搜尋關鍵字設定討論會議

會議時間：105 年 6 月 20 日（一） 15:30-17:00

會議地點：中央銀行 6 樓會議室

參與人員：計畫主持人 郭迺鋒 副教授

兼任助理 林哲緯 碩士

兼任助理 黃筱雯 學士

會議主題：網購消費商品搜尋關鍵字與粉絲專頁報告與討論

委託單位意見	研究團隊回應
1. 消費者對於進口品的消費，是否包含在 15 個網購的關鍵字中？	1. 淘寶網、GoHappy、ibon...等都是國際連鎖的網站，具有代表性，在網購的品項中都含有各種進口商品，因此進口商品的關鍵字也包含在其中，央行可以考慮與統一超商作簽約，統一超商內的掃描系統、ibon 機器，每個月約有 35 萬筆的資料，透過消費者實際的消費紀錄，可以理解出消費者的行為如何變化。
2. 在 15 個網購中，每一個購物網中的分類有一致性嗎？	2. 各家網購的大類有相似性，在中類的分類上有些許的差異。
3. 購物網站的選擇上，有沒有搜尋的依據？	3.此為參考鷹眼數據中，他們統計 2014Q4 以來，台灣前 15 大的購物網站。
4. GOMAJI 團購、Yahoo 拍賣，也常常被大家廣為使用，要不	4.謝謝貴行提供的建議，我們會再將 GOMAJI 團購、Yahoo 拍賣

要將這兩個網購再加入資料中，一起做分析呢？	的網站，加入資料庫中。
5. 利用這些關鍵字，比較總體變數時，會使用那些官方資料做比較？	5. 會利用國民所得、國內經濟部統計處的零售資料(月資料)...等，一起做分析，Google Trends 下載出來的資料為週資料，經由轉換後變成月資料，將可以作分析比較，且 Google Trends 涵蓋著所有人的搜尋紀錄，是具有指標性的資料
6. 怎麼結合數據資料，達到最後總體的指標，在消費預測上做出最適當的決策呢？	6. 利用視覺化圖表，表達出各資料間的關聯性，其關聯性高的，可以當作最後的指標依據，例如：XX 商品的指數，可以作為消費預測的指標，但這只是消費預測中的其中一個方法，不代表最後的決策。
7. 如果搜尋出來的關鍵字包含「疫苗」，這似乎跟消費牽扯不上關係，如何解決此種雜訊問題？	7. 這涵蓋著消費者搜尋資料時的情緒問題，當消費者在搜尋時，不一定要實際去購買，想要但是不一定會實際去做，但是可以利用這個偏好，了解消費者是否對此東西是否有興趣，另外，財務心理學中也提到「注意力假說」：當人們產生注意力時，其行動力就會提高，因此消費者會如何做消費決策，可以透過網路「輿情」指標，來分析彼此間是否有關連性，「輿情」指標，具有獨特性，且是人們自然而然留下的紀錄，並非一般的問卷調查，我們搜尋「輿情」資料的來源包含：Google Trends 週資

	料，從 2004 年至今；FB 爬文，日資料，兩者的頻率相當高。
8. 當景氣不好時，消費者只會看看，不會實際去購買，所以其搜尋資料的結果，可能無法真實表達出消費預測？	8. 在不同的燈號下，會有不同的行為決策，因此已經將此項因素考慮在其中了，燈號的變化也不能夠代表實際的購買行為會發生改變，例如：「我習慣到統一超商去買咖啡，不會受他家便利商店的促銷影響」，所以景氣好壞、店家的促銷策略，都不能真實表達出實際的消費行為，因此團隊採用「輿情」資料，來觀察消費者潛在的消費情形，跟傳統的統計資料也許會有落差，但這個「指標」並非要具有代表性，而是提供一個新的來源，未來貴行在做決策時，可以當作一個參考的因素。
9. 在現有的成果中，各資料間彼此的關係有什麼關聯性？	9. 目前團隊以「博客來」的資料作分析，就發現到它們在大類的部分，就呈現正相關的線性圖形，因此可以推論出，光是用「輿情」資料作分析，就具有相當的影響力，而政府提供的「消費者信心指數」只有統計數據，並沒有包括「輿情」的因素。
10. 當用不同關鍵字搜尋時，可能會造成不同的結果，像是「無印良品、MUJI」，要如何反應出資料的精確性？	10. 每個人會依據自己的習慣去搜尋資料，因此不管是使用英文還是中文，都包含在消費輿情的訊息中，會使用中文搜尋的通常為台灣、中國大陸、使用英文搜尋的可能是美國人，

	這是一個大數據資料庫，利用前十五大的網購店家分類來探索關鍵字，搜尋出來的結果本來就無法很精確，因此可以將最後分析出來的結果，與實際狀況做比對，就能夠知道這樣的「輿情分析」是否有準確度。大數據可能只有 5 年的景氣，現在仍處於大數據資料的過渡期，未來會有像是 TEJ 資料庫，將這些資訊做彙整後，再將資料賣出
11. 可以將消費者信心指數與股市指數或外匯市場做連結嗎？	11. 在操作上是可以執行的，但是這可能需要開另外一個專案計畫來做，目前團隊先把輿情指標的因素完成，未來可以加入央行的資料做比對，來預測實質的消費。另外可以用同樣的模式，將各家金控的 FB，針對財富管理、外匯、TRF、期貨等作搜尋，做為未來預測參考的項目之一。外匯市場中，大家會注意何時去日本、美國旅遊時較便宜、貿易的匯差，通用可以利用這種 Google Trends、FB 爬文這種模式，來觀察大家關心的議題、對未來的看法、市場的反應情形。

附件二 期中審查意見及回應表

「應用大數據提升台灣民間消費預測」委託研究計畫 期中報告審查會

時 間：民國 105 年 8 月 26 日下午 3 時至 4 時 30 分

地 點：中央銀行第 2 大樓第 1102 會議室

主 席：林處長宗耀

報告人：郭副教授迺鋒（世新大學財務金融學系）

出 席：

評論人：林教授馨怡（政治大學經濟學系）

連教授勇智（台灣大學國際企業學系）

經研處：林處長宗耀、林副處長淑華、吳副處長懿娟、廖研究

員俊男、張副研究員天惠、鄭副研究員漢亮、吳副研

究員俊毅、陳專員佩玗、林專員依伶、朱專員浩榜、

蕭宇翔、李宗憲

金檢處：潘稽核雅慧（代理侯研究員德潛出席）、盧專員月雲

外匯局：曹副研究員體仁

業務局：張專員國興

壹、評論人意見與報告人答覆：

評審意見	研究團隊回應
壹、林教授馨怡： 個人對本研究的想法給予肯定，以下針對期中報告之內容分別提出簡評、建議與未來研究方向： 一、 期中報告簡評： 1.在網路世代，注意力的產生會帶來搜尋力，並進而影響購買力。因此，藉由網路檢索搜尋量的觀測，可即時預測民間消費。 2.社群媒體上有巨量資訊量，若能量化及分析，並轉為可用的參考數據，將有助於預測民間消費。 3.本研究課題重要且有助於政府部門即時掌握民間消費資訊。 4.針對關鍵字設計爬文系統以及擷取套件等，搜集民間消費相關巨量資料，對於民間消費預測有貢獻。 二、 期中報告建議： 1.如何透過注意力的形塑，進而帶動民間消費購買力，當中的傳遞機制，並未在報告中說明，建議可以增加文獻探討，以及管道說明。 2.既有文獻大多是針對失業率、旅遊、股市、或特定產業的實證分析，但這些分析是否能適用於民間消費此類總合變數，需要解釋及討論。 3.社群媒體的巨量資料，如何清除雜訊，並轉換成有用資訊，以及其與民間消費之間的關聯性，可以增加文獻探討及方法說明。 4.由於變數皆時間序列資料，可增加實證變數的定態性確認，再進行模型設定及估計，否則可能有虛假迴歸疑慮。例如：期中報告的水準值模型結	研究團隊回應： 二、期中報告建議： 1.本計畫將補充文獻，說明相關傳遞機制與管道。 2.感謝評審建議，已進行相關調整。 3.感謝評審建議，已進行相關調整。 4.在進行模型評估前，本研究針對各變相進行單根定態檢定，並針對非定態數列已差分處理，謝謝評審建議。 5.謝謝評審的建議，在因子命名上已做

果和變動值模型分析結果並不一致。 5.實證分析中主成份之分類名稱（一般消費與進階消費），似乎與一般認知不同。例如：文具、飲料或洗沐－沐浴類，民眾多認為其屬於一般消費，但本研究將其歸類於進階消費；建議可以思考不加名稱或更改為適當名稱（p.39）。 三、未來研究方向： 1 由於大數據資料量龐大，雜訊也相當多，建議在預測前先進行因果關係檢定，排除雜訊，過濾有用訊息。 2.本研究當中有高維度問題，期中報告僅使用主成份分析方法萃取主要因子，建議可以使用其他挑選變數的方法，例如：Ridge regression、LASSO（Least Absolute Shrinkage and Selection Operator）等，以協助預測。 3.由於未來將進行的預測能力評估等量化分析需要設定落後項，或是樣本內、樣本外模型比較等計量處理，而樣本數目前僅有 48 筆，可能產生自由度問題。建議尋找更高頻率資料。例如：經濟部統計處的零售額月資料等。 4.關鍵字設定是否能更為標準化，涵蓋範圍更廣泛，例如：主計總處之定義。某些民間消費項目未列入關鍵字，例如：教育、水電瓦斯及其他燃料、醫療、交通、通訊、餐廳及旅館、民間非營利機構消費等。	相關調整。 三、未來研究方向 1.感謝評審建議，已先進行相關因果檢定，詳細資訊請見附件四。 2.感謝評審的建議。受限於研究期程，未來有關計量上的處理，將可嘗試 Ridge regression 與 LASSO 等方法。此外，總體經濟變數，本研究考量指標時程的預測力，故本研究目前採納總體經濟領先指標做為目標變數。 3.謝謝評審的建議。然而受限於網路檢索指標之期間限制，本研究目前設計由 2004 年第一季起之資訊為主要分析內涵，後續研究待資料量的累積，將可透過更長期間的資訊觀察其預測效力。 4.感謝評審的建議，關於關鍵字設定的標準化問題，由於 15 個購物網下各有各的分類，原設計從中找出 3 個分類，分別為：GDP 中民間消費有 12 個子項、家庭收支調查中的 12 個子項與物價指數中的細項，進而分成 11 大類。然而，商品類許多項目多對照到單一項目（休閒及文化）之中，影響序列累積。考量前述原因，本研究採用 15 個購物網中之關鍵字依照此 11 中類目進行分類對應，並做後續分析，未來期望朝總類與分類皆嘗試的方向處理。
---	---

5.目前預測標的與零售資料較接近，而與總體民間消費有落差。	5.本研究目前透過更詳盡之模型設計來預測我國民間消費走勢。
貳、連教授勇智 大數據為近年之趨勢，本計畫嘗試利用巨量資料提升民間消費預測上之即時性與準確性，其中即時性尤為重要，而準確性的部分則可逐步修正與改進。此外，本計畫非常細心，執行的過程面面俱到，目前亦往正確方向前進。以下針對期中報告的成果，給予幾點意見與未來研究方向： 一、從機械學習與人工智慧的角度，經濟預測應為一種漸進式的學習過程，可使預測準確性逐漸地提高，這樣學習過程與進程，可嘗試在本研究中展現。譬如說，目前的模型納入不同的變數，且可以分別討論各個變數效果。現有研究結果顯示變數之解釋能力還不錯，甚至 R-square 高達 96%，未來或可進一步分開驗證這些變數的預測效果，以找出各變數解釋力所占之比例。對於解釋力較低的變數，可嘗試進行修正，以提升它在預測上之準確性。據此即可展現出一種修正與進化的預測過程，建議可嘗試往此方向發展。 二、變數的建置過程中，運用到聲量的概念，惟文中並未給予具體的定義，應係指討論或成交的聲量。假設為討論的聲量，則消費者在搜索某項欲購買之商品時，會在不同的購物網站中關注與比價，且皆留下軌跡，但是最後只在某個購物網站消費。在這樣的情況下，可能會造成本文變數指標的建構產生膨脹與過度計算的現象，未來如何做修	一、預測部分依循連教授的意見，進一步對各變數的解釋力與預測力加以檢測與分析，感謝評審委員的建議。 二、有關聲量的定義，如果是 Google Trends 即為搜尋的次數，而如果是從 Facebook 或是 PTT，比較像是情境下之討論，可能是文章、轉貼或互動，所以該情形下聲量類似關注度，彼此間有些微的差異。本計畫後續以 Facebook 與 PTT 的聲量資料整理完成後，亦將進行模型配適，以分析 Google Trends 的搜尋次數以及 Facebook 與 PTT 情境下的討論何者較具預測能力。

正與精進為可思考的方向。 三、本研究的消費品項分類大致分為一般消費與進階消費。從行銷學角度來看，不同消費的模式（如消費與服務商品），可以從不同的品項建構消費的規模，再進一步觀察各品項與景氣循環間之關係，如飲食支出的消費可能就較不受景氣之影響，但是珠寶類消費（奢侈品）就會受景氣的影響。因此，可根據各品項受景氣影響的程度，考慮不同的權重，重分配各品項商品之角色，以強化模型的預測能力。 四、本研究從網路資料搜尋消費關鍵字，惟跨境消費可能已包含在內，並非全為國內消費。未來的研究中，將跨境消費與國內消費連動納入考量，是可以思考的方向之一。國貿局目前也正在嘗試研究電商的行為，以掌握跨境消費的脈動。 五、郭副教授提及未來可嘗試 online 與 offline 資料的結合，目前本計畫使用的資料多為 online 資料。國內各家大型賣場與商店均可蒐集相關的數據，如果能將兩邊的資料一併考慮，應可提升民間消費的預測能力。 六、未來可透過網路共享的概念，嘗試加強模型預測能力，例如，每個人對景氣的看法皆不同，有人去餐廳消費，發現人潮相當熱絡，認為景氣未必那麼差；有人去採買年貨，發現跟去年相比人潮似乎真有受到影響。透過網路共享開放的平台與管道，讓消費者抒發自己的看法，利用這些即時普遍的共享資料，或可進一步強化模型的預測能力。	三、謝謝委員的建議，有關品項的分類，目前在品項的分類上，物價指數的分類是比較詳細，本研究將納入未來研究建議之中，期望透過後續研究能夠做相關分析。 四、在目前的研究架構下，本研究恐無法處理跨境電商與大型賣場資料的部分。如果央行能夠掌握賣場資料的話，在預測上的確會有很大的幫助。另外，如果未來能夠建立共享的平台，也將有助於預測。 二、謝謝評審的建議，受限於本次研究期間、資源有限，未來將嘗試結合 online 與 offline 兩者資訊。 三、謝謝評審的肯定。受限於社群網站之電商經營仍屬近三年的新興活動，此亦影響本研究在資訊蒐集累積的量度，故本研究透過聲量分析及社群網路分析模組來探索社群網站對民間消費影響之內涵，進行其與民間消費之關係性分析。
---	--

參、侯研究員德潛（潘稽核雅慧代為宣讀）

- 一、民間消費模型的部分，民間消費的需求函數一般學理上大多係以實質變數來設定，可否嘗試於目前模型中加入 CPI 或是民間消費的平減價格，以補充遺漏的解釋變數。
- 二、解釋變數型態的選取，利率方面目前本計畫是採用 10 年期中央公債，但一般學理上在消費函數中多採短期利率，商品均衡部分通常係考慮實質利率。
- 三、有關方程式設定的部分，同時考慮每人平均 GDP 與總體民間消費，恐產生一致性的問題，建議或可直接採用人口與 GDP 作為解釋變數。
- 四、有關民間消費的預測部分，目前的內容皆為樣本內的事後模擬，似乎無法進行預測，建議或可進一步釐清前期項聲量與當期消費間之關係，檢測其是否具顯著的影響，或可提升模型的預測能力。
- 五、有關 Pokemon Go 與民間消費間關聯性的部分，目前的估計式中，股價與民間消費並無同期顯著性，但在 Pokemon Go 的聲量主成份分析中，對中華電之股價報酬率卻有影響。從聲量的搜尋影響股價，進而影響民間消費，這樣的關係是否可進一步以文字補充說明。
- 六、期中報告中最後有提及因果檢定與 VAR 理論說明，期末報告是否會進一步提供相關的實證結果？
- 七、與林教授建議類似，第一與第二主成份之組成差異不大，彼此間的關係為何？當中皆有包括消費財、一般消耗品與耐久財等。是否可針對

一、謝謝侯研究員的意見，民間消費已採用平減價格來進行評估。

二、謝謝侯研究員的意見，本研究考量民間消費的預測力，故目前採用之重要變項為領先經濟指標作為後續預測模型之變項。

三、同上。

四、謝謝侯研究員的意見，本研究已透過 ARIMA 及 VAR 模型探索關鍵變數落後期對我國民間消費的影響。

五、Pokemon Go 之相關研究主要呈現搜尋關注力之於股市的影響效力，由於並未與本文民間消費預測有直接關係，參考計量分析科之建議，相關內容已刪減。

六、請參考期末報告第五章之研究成果。

七、在期末報告中，於進行模型測試前，各變項已先進行相關分析，相關內涵請見附件四。

此分類進一步說明或調整。 至於其他的細部意見，書面資料會提供郭副教授參考。	
肆、張專員國興： 一、有關主成份分類的部分，第一與第二類的名稱尚可理解，但是第三或第四較不容易理解，且第三與第四很多部分皆包含於第一與第二類中。據此，第三與第四類名稱似有不妥。 二、本計畫透過關鍵字的搜索蒐集網路消費行為大數據資料，屬於非結構化的資料，通常這樣的資料具備數量龐大與更新速度相當快的特點，所得到的結果亦非常多樣、複雜且充滿變化，需要相當高明的洞察力，或是特別的敏感度，在亂序中整理出其中的相關性與規則，以協助決策判斷。而在資料處理方面，大數據通常為相當籠統的觀念，較適用於個體面的應用，例如：商場產品交叉銷售、紅利積點行銷策略。但是如果應用到總體經濟領域，則要非常慎重。這種非結構化的資料，資料的生成往往充滿不可預測性、前後不一致性與不穩定性，但總體經濟中的變數通常為結構化變數。因此，本計畫每期所篩選出的變數必須檢測其是否為穩定、定態與可預測的變數，如果為不穩定的變數，卻拿來當作幫助民間預測消費的變數，恐將造成偏誤。 三、大數據於民間消費的即時預測，通常必須滿足強、弱外生性的檢測，但是本文並沒有進行相關的檢測。一般而言，大數據的非結構化	一、感謝專員建議，主成份之命名已加以修改，。 二、有關總體資料是否適用於文字探勘，本研究團隊於央行第一次報告時，大致已將國際上央行利用文字探勘的相關研究進行整理。英國英格蘭銀行在此議題上推動較多，而本委託研究計畫則係參考日本央行的相關研究，其是依循 Google 首席經濟學家 Varian 的文章進行，當時 Varian 係討論耐久財中之汽車與房屋銷售，分析 Google Trends 的相關資料是否具有影響力，結果相當不錯。英格蘭銀行則是將文字探勘列入研究主題之一，同時期 BIS 也針對各國央行進行文字探勘的接受度調查，發現各國央行對於文字探勘的接受度皆相當高，大致支持文字探勘的方法論。此外，再進行變數處理時，就時間數列資訊處理模式，先做相關單根檢定，詳請見附件四。 三、有關強、弱外生性檢定的部分，受限於研究時程，本次研究尚未處理，未來或可進行相關的檢測。本研究係以 Yahoo 購物中心為例做說明，目前共有 42 種分類，如以使用

資料應可符合強、弱外生性的要求。個人認為大數據資料在總體經濟的應用恐相當有限。	者的經驗來看，或可看到更深層的經濟意涵，但本研究計畫恐無法延伸討論到這部分。利用傳統時間序列的方法只能捕捉搜尋關鍵字的相關性，並不能解釋搜尋行為的相關性。央行可以嘗試大數據的有關應用，例如之前提到的分享平台。在此，本計畫只是初步的嘗試，試圖提供央行不同的參考資訊。至於模型的預測能力，或許可經過幾季的觀察後，再來檢驗該模型的實用性。
伍、廖研究員俊男 一、計畫欲採用 17 家電商資料進行民間消費預測的相關分析，惟目前期中報告僅提供 Yahoo 購物中心的分析結果。後續的研究係採各家分析的方式，從中找出最佳電商的關鍵字組合，或將全部電商關鍵字彙整從中找出主成份再進行分析？ 二、Google Trends 的資料如何與民間消費中之細項對照？其中是否考慮租金支出與醫療等變數。	一、目前本研究以為各家電商之商品中分類商品類目進行指標蒐集及分析，並經由主成份萃取出檢所指標並進行後續預測分析。 二、Google Trends 的資料中皆有囊括該分類。Google 為模糊搜尋，並非精確搜尋，不管是錯誤或正確的字，在一般情形下 Google 均可提供精確的搜尋結果。另外，Google 本身有其產業分類，但本計畫並未使用，而改採用電商分類，因其比較符合民眾用語。Google 在文字探勘與模糊類比的搜尋力上應為全球第一，此應為無庸置疑的事實。而大數據的資料會不會有誤差？當然會有，大數據的其中一個特性即為容許錯誤存在，傳統問卷也不一定會調查到精確的結果，普查資料中也可能存在諸多錯誤，各種方式蒐集的資料非完美無瑕，也不會因為是大數據的資料而使不完整性有所提

三、就模型設定來看，本研究同時放入民間消費與 GDP 在消費估計式的左邊與右邊，惟從資料的角度觀察，民間消費與 GDP 應是同時產生，不宜以此式進行預測。建議考慮以實質薪資月頻資料取代 GDP，且消費函數右邊放入消費的前期項進行預測。 四、本計畫後續或可嘗試混頻資料計量模型的相關運用。 五、建議本計畫將常見的總體變數一併納入考慮。 六、建議本計畫建置民間消費預測的基準模型（baseline model），如傳統的民間消費模型或時間序列模型，並進行樣本外預測與 D-M 檢定，其他意見會詳列在書面資料，並於會後提供郭副教授參考。	高。 三、本計畫後續會考慮 AR (p) 模型進行相關分析。亦考慮相關的月資料，以總體經濟領先指標分析之，詳細內容請見第五章，感謝廖研究員的建議。 四、本計畫目前嘗試提供較著重於基礎面的分析與相關網路資料指標的建構，後續相關資料將全部移轉，央行如有需要，可再進行相關的計量分析。 五、本計畫已加入總體變數，並且在同時考慮網路資料與傳統總體變數下，或可透過嘗試不同的組合，以提升模型民間消費的預測能力。 六、本研究目前已進行修正與補充說明，謝謝廖研究員的建議。
陸、業務局 一、p36-38 有關表 5-3 中 42 項關鍵字清單的排列順序最好能與圖 5-3 消費關鍵字 Google trends 每週聲量走勢圖、圖 5-4 消費關鍵字每月聲量走勢圖及圖 5-5 消費關鍵字 Google trends 每季聲量走勢圖的排列順序對應，因為作者雖已紅框標示，但個別關鍵字的圖標題太小，難以辨識，如果排列順序再不對應，則圖文的說明會讓讀者產生	一、謝謝您的建議，相關圖表已做調整。

不一致之錯覺。 二、 p38 表 5-4 消費搜尋引擎主成份分析，作者將關鍵字搜索所得的消費大資料（Big Data）歸納為：一般消費、進階消費、鞋子與其他項目、內睡衣等 4 大主成份因子，感覺有點怪，前兩項消費名目還容易懂，後兩項到底屬於甚麼性質之消費名目？似可包括在一般消費項目下，沒必要個別獨立成為主成份因子。	二、 有關主成份名稱選取的部分，並非如傳統以品項的角度分類，而是必須跳出傳統的思維改採搜尋的角度分析，因此，本研究是以搜尋關鍵字的相關性進行分析，而非商品的相關性。另外，Google 亦有相關性的分析：Google Correlate，但這部分資料並未公開提供，如果去觀察這些相關性，會發現以品項的角度來看，很難理解彼此的關聯性。是故，本研究係以搜尋的相關性找出主成份，這樣的主成份是否具有預測力，就交由模型找出最合適的組合，本研究在期末報告中依循時間序列的方法修正與補充本模型的結果。
三、 透過關鍵字爬梳（key word mining）所篩選出來網路消費行為的 Big Data，這類非結構化資料通常具備龐大數量（volume）、電子速度存取（velocity）及複雜多樣變化（variety）等特性，它需要高明的洞察力或獨特的敏感度，亂序中理出相關性，從而促成更強的決策、判斷與處理能力。Big Data 分析通常是語意籠統的相關概念，作為個體面應用，或有助於商家產品交叉銷售或搭配紅利積點的行銷策略；但應用於總體經濟，必須慎重，因為這類非結構化資料的生成，往往欠缺可預測性、一致性與穩定性。	三、 有關總體資料是否適用於文字探勘，本研究團隊於央行第一次報告時，大致已將國際上央行利用文字探勘的相關研究進行整理。英國英格蘭銀行在此議題上推動較多，而本委託研究計畫則係參考日本央行的相關研究，其是依循 Google 首席經濟學家 Varian 的文章進行，當時 Varian 係討論耐久財中之汽車與房屋銷售，分析 Google Trends 的相關資料是否具有影響力，結果相當不錯。英格蘭銀行則是將文字探勘列入研究主題之一，同時期 BIS 也針對各國央行進行文字探勘的接受度調查，發現各國央行對於文字探勘的接受度皆相當高，大致支持文字探勘的方法論。
四、 作者將大資料用來協助台灣民間	四、 謝謝提供建議，有關強、弱外生

消費的即時預測，只要該數據的生成過程（Data Generation Process）符合強外生性（strong exogeneity）之條件，大致上不會有問題。若想進一步進行政策分析，則必須符合超外生性（strong exogeneity）之條件。換言之，具有強外生性的解釋變數，一般不能再作為政策工具干預的超外生變數。因此，大數據分析在總體經濟的應用相當有限。	性檢定的部分，如果時間允許，本研究會進行相關的檢測。
柒、計量分析科 一、研究主題 本文主旨為應用大數據提升台灣民間消費預測，重點應為「網路大數據指標之建構」，以及「大數據指標能否顯著提升預測績效」。惟就期中報告來看，尚看不出加入大數據指標後，對於民間消費預測能力提升之成效，亦未說明將來如何應用大數據提升台灣民間消費預測。建議未來可以傳統民間消費預測模型或時間序列模型作為基準模型，再比較加入大數據指標的模型預測能力，俾利了解大數據指標是否確能有效提升民間消費之預測能力。 二、段落內容 1. P.5 提及近來熱門的手機遊戲 Pokemon Go，除非作者認為此資料對消費預測很重要，否則不須耗費篇幅說明 2. 本文目的為「應用大數據提升民間消費預測」，文獻回顧，除非有助於本文實證操作或研究，否則建議刪除細節及無關。例如： (1) 宜先介紹並定義大數	一、研究主題 謝謝提醒，本研究目前評估多樣態模型並進行相關比較，探索納入網路大數據之後對於我國民間消費預測的影響。 二、段落內容 1. 謝謝評審的指導，本計畫在執行計畫時，發現實務上的案例可能與本次的消費預測方式有相似處，將手機熱門遊戲 Pokemon Go 進行進一步研究時，所發現之特殊的現象呈現參考，目前已將相關內涵調整。 2. 內文已增加大數據概念的介紹，以便評審更加了解大數據的思維及研究方式。 (1) Godfrey et al. (2014) 一文可聚焦於討論社群網站使用者的想法、資料

據，以及探討大數據與民眾經濟活動的關聯，以便了解為何大數據有助提升經濟預測結果。 (2) Godfrey et al. (2014) 為探討社群網站指標的文獻，但內容與本文主題差距甚遠，可刪除此文；或可聚焦於討論社群網站使用者的想法、資料適用性。 (3) 「注意」、「注意力」、「關注力」、「關注度」似同為英文字 attention，建議於首次提到時補充英文原文 3. P.35 提及目前成果僅以 yahoo! 奇摩為例。惟 P.18 亦提及本文已篩選出 17 家電商平台（表 3-1）及相關關鍵字，且 P.35 說明考量電商商品類定義有重疊，以及關鍵字與項目別過多，將影響大數據指標之建置；兩者說法不一致，似有矛盾。因此，期末成果是否將擴展至 17 家電商平台？直接以人次占比最高的 Yahoo! 奇摩購物中心搜尋引擎的民間消費指標關鍵字作出結果，是否真較其他 16 家電商為佳，仍有待驗證。 4. 第三章建構大數據民間消費指標的說明不夠清楚： (1) P.20 表 3-3 中提及第二步驟是將中類的分類項目放入 Google Trend，查詢每周被人搜尋的頻率。而	適用性，雖未與民間消費有直接性關係，但仍可作為觀念釐清之相關依據。 (3) 已補充英文原文內容，謝謝您的建議。 3. 本計畫在期中報告時，已經把 15 家電商（相關篩選原則請見附件五之說明）的關鍵字數據資料都收集完成，將 Yahoo 的三大平台做關鍵字的整合：Yahoo 購物中心、Yahoo 拍賣、Yahoo 服務，原本分別有 70 項、393 項、120 項，考量 Yahoo 商品類定義重疊，整理出商品的唯一清單共 529 項，分別為 42 項、370 項、117 項。先從購物網站中的分類，作為關鍵字搜尋，再將關鍵字放入 Google Trends，確定消費者在搜尋時的文字用語後，再放入 Fb 做搜尋，關鍵字通過三階段的設定，將使本研究可以做進一步的分析。 4. (1) 在第三章 P.20 第三步驟放入的關鍵字選取，是由 Google Trends 搜尋到的所有關鍵字直接放入爬蟲平台中做搜尋。有些關鍵字不符合搜尋的
--	---

第三步驟放入的關鍵字，是否為放入頻率最高的關鍵字？	用語或是字句太長，在 Google Trends 搜尋不到結果，因此在 Google Trends 做搜尋時，就已經將 15 個電商平台的分類詞經過篩選。相關內容已在內文進行調整。
(2) 關於社群網站的指標，是否為選定數個關鍵字，建立關鍵字出現次數的時間序列資料？又是否應與最後選取的網路檢索指標關鍵字相同？而建置預測民間消費的社群網站指標，與多變量分析（集群分析與社會網絡分析）或文字雲的分析方式又有何關聯？ (3) P.23 民間消費指標處理架構的說明不夠清楚，建議將處理流程、篩選標準，以及後續將如何應用於預測等步驟，一步步明確寫下。 (4) 在每次預測時，關鍵字是否需定期更換？	(2) 關鍵字的結構變化，受限於計畫執行時間，目前暫不做處理。 (3) 謝謝評審的指導，已將民間消費指標處理架構、流程及篩選標準，做文字上的修正。 (4) 關鍵字的結構變化，受限於計畫執行時間，目前暫不做處理。
5. 第四章研究方法，許多內容為估計式的介紹，似可簡化；而集群分析、社會網絡分析，KMO 及 Bartlett 顯著性檢定等卻未多做說明，另請補充相關估計式及說明。此外，文中未說明將如何應用上述計量方法，建議在各模型敘述段落簡單說明如何應用於本文研究，以及相關的模型設定。在段落安排上，由於動態因子模型要求其觀察值為定態數	5. 本計畫在處理時間性質資料，完全採時間序列資料處理法，例如：單根，使研究方法具連貫性。

列，建議將 ADF 單根檢定方法前移至主成份分析，使研究方法具連貫性。 6. 第五章第一節，表 5-1 與表 5-2 列出總體經濟指標與民間消費支出細項清單，建議未來補充說明兩者對民間消費模型之角色為何。另外，民間消費支出細項走勢與本文所用的關鍵字聲量走勢是否有相關或對應關係？可否使用關鍵字，分別預測民間消費支出細項，再加總各預測值？如果不估計民間消費細項，是否有必要列出（包括 P.77），請斟酌。 7. 第五章第二節之實證結果宜多加論述，使讀者了解相關經濟意涵。 8. 第五章第三節敘述社群網路資料擷取平台之建立，惟本文目的係建置大數據民間消費指標，關於擷取社群網站文字的爬蟲軟體操作可移至附錄。正文應加強說明如何從社群網路使用者的發文建置預測指標的步驟。	6.有關消費預測的指標，本計畫將會以總體的消費支出取做預測，不分細項。相關圖示，僅是用作觀測我國民間消費樣態內涵。附錄資訊也進行刪減，謝謝您的建議。 7.謝謝指導，本計畫已在第五章實證結果多加論述。 8.謝謝指導，由於建置爬蟲軟體為期初規劃之重點之一，故將期置於正文說明，在期末報告中已作順序調整，而本計畫也在正文加強說明社群網路使用者的發文建置預測指標。
三、資料處理 1. 本文目的係提升民間消費預測績效，建置大數據指標時，所用的網路搜尋關鍵字宜與民間消費支出細項吻合（如表 2-2 美國 BEA 之作法）。否則，建置的大數據指標僅與實體商品相關性較高，較無法涵蓋服務部分，或是其他占民間消費比重甚大的租金、交通、餐廳支出等。 2. 部分資料說明仍待補充：	1.所言高見，本計畫將實證結果做成政策建議時，會將其限制性做說明。 2.

(1) P.33 表 5-1 總體經濟指標之選取標準為何？其中，序號 2 重要總體經濟變數似又與文中引用之 GDP 等指標有關？是否各指標之細目別資料皆用於實證？若否，則建議將附件三中，實證時未採用的變數刪除。事實上，表 5-1 有些指標與民間消費關聯較小（如營造工程），而有些與民間消費關聯較大之指標（如零售業、餐飲業營業額等）卻未包含其中。建議另可參考主計總處新聞稿表 8 之「民間消費相關指標」。 (2) P.33 民間消費支出資料為名目或實質？估計式右邊解釋變數為名目或實質？兩邊是否一致宜先確認。 (3) 建議說明如何將日資料與月資料轉換成季資料（如加總、平均等），而關鍵字由週聲量轉月聲量或季聲量，亦應說明如何轉換；成長率係以 YoY 或以 MoM 計算，亦需說明。 (4) 在繪聲量走勢圖時，為何「每週」及「每季」均為時間序列走勢圖，但「每月」卻為「歷年同月」的走勢圖？建議統一。 (5) 觀察圖 5-1、圖 5-3 等，可發現本文數據多有季節性因素，但文中並無提及如何處理。建議說明相關變數是否先經季調，再進行迴歸預測（如 Schmidt and Vosen（2009）的建議）？ (6) P.38 將 Yahoo 的消費關鍵字分	(1) 因為本計畫要做民間消費的預測，所以只採取具有領先性的指標，故本研究用國發會的領先指標做為短期預測。 (2) 兩邊都是採實質的變數。 (3) 轉換資料的方式，是採取 EViews 中的轉頻方法，指數的數值採用平均法，量的數值採用加總法。 (4) 謝謝指導，本計畫將會將聲量走勢圖的時間序列做統一。 (5) 由於各項預測變數皆透過季頻率的調整，目前暫不處理季節調整之項目，以觀察原始變項之特性。 (6) 本計畫用陡坡圖的方式去判斷四
---	--

析出 4 大主成份因子。為何選擇選取 4 個因子？是否依據 Bai and Ng (2002) 建置的資訊判定準則選取？是否經過 factor rotation 後才得到 4 個主成份？ (7) P.38 提及細目資訊可見「附件五」，惟「附件五」第肆部分之「KMO 及 Bartlett 檢定」(P.95) 與解說總變異量表 (P.96) 並未與本文討論連結。此外，第二主成份呈現下降趨勢，建議稍加說明 (似會影響後續實證結果)。 3. 若分別以水準值及變動率模型進行估計，則無單根檢定的必要 (p.77)。 4. P.40 關鍵社群網站輿情大數據指標資料起始日為 2014 年 6 月，若資料長度與 Google Trends 資料長度不同，如何建立民間消費預測總指標？如何以計量模型檢視社群網站輿情大數據指標有助預測民間消費？ 5. 本文在實證時，對利率進行取 log、差分等處理，惟理論上，較少對利率變數這樣處理。 四、實證模型與結果 1. 本文在民間消費模型估計後，進行樣本內預測，惟是否應為樣本外預測？因為 OLS 在估計係數時已將殘差最小化，樣本內預測意義不大。另請補充樣本外預測方式 (固定法、滾動法或遞迴法)，以及	大主成份的個數，按照多變量的判准：未轉後之統計量、陡坡圖、最大變異數法做取捨，當作其主成份。 (7) 謝謝指導，將在正文加強說明「KMO 及 Bartlett 檢定」與解說總變異量表，以及第二主成份下降的趨勢。 3. 謝謝指導，本研究已在單根上做處理。 4. 在做 Google Trends 時會將時間做交集，再做後續的分析，統一由 2004 年起進行分析。而社群網站輿情大數據受限於資料期間過短，則透過網絡分析等方式探索其與我國民間消費之相關影響，以供後續研究參考。 5. 處理的原因為，本計畫要做民間消費的預測，能預測就沒有解釋力，能解釋就無法做預測。在期末報告的研究中對於總體變項的挑選已做相關調整。 四、實證模型與結果 1. 期末報告已依照原先的計畫做 DFM 模型、並進行 ARIMA 及 VAR 模型。
--	---

與基準模型(如傳統民間消費預測模型，或時間序列模型)比較之 MAE 與 RMSE 大小。另外，期末報告請進一步進行 Diebold-Mariano 檢定，以確定模型間預測績效是否有統計上的顯著差異。 2. 本文的研究目的係以當期經濟指標即時預測(nowcasting)當期民間消費或預測未來 1 季以後的民間消費？事實上，文獻多發現 Google Trends 僅能提供短期預測的資訊，對長期預測助益不大，建議本文對預測期間的長短略作說明。 3. 本文以季資料進行預測，Google Trends 資料起始點為 2004 年，致樣本數只有 48 期。建議本文可考慮混和頻率的計量方法(如 Giannone et al., 2008)，運用高頻月資料(甚至是周資料)預測低頻的季資料(民間消費)，充分運用網路大數據資料能及時提供最新訊息的優點。 4. 本文建構之社群網站指標、網路檢索指標等，多屬高頻率的週或月資料，建議建置可用於每月更新民間消費支出的預測模型。否則，每月耗費心力搜集大數據資料，卻無法用於民間消費的即時預測，實屬可惜。 5. P.41、P.43 民間消費模型如何設定？是否參考相關文獻？表 5-1 與表 5-2 列出總體經濟指標與民間消費支出項目清	2.經過本計畫的實證結果，才能做判斷預測能力的好壞，與國外文獻上的預測是否為可行性。 3.本計畫目前暫不考慮此方法論，會在結論的時候，將混合頻率的方法，作為結論上的建議，期望未來後續研究嘗試相關作法。 4.因為目前實質消費的頻率為季資料，而且為延宕 2 季，不存在月頻率資料，所以本研究以週頻率資料，來彌補此落差。 5.這些變數完全是採國外文獻回顧上的變數。文獻為：Torsten Schmidt and Simeon Vosen (2009), Forecasting Private Consumption-Survey-based
---	--

單，為何最後模型的總體變數只有人均 GDP、市場利率、股價指數及消費者信心指數？相關變數之意涵及其對民間消費之影響為何？建議補充說明。另外，模型亦可考慮零售業、餐飲業營業額等與民間消費關聯性較高的統計資料（月資料）。 6. P.41、P.43 模型設定並無下標時間，無法了解本文係以同期季資料預測同期民間消費，或未來民間消費。另應注意模型目的係為預測民間消費，因此不應包含與民間消費同時發布的國民所得帳中同期相關變數（如 GDP、人均 GDP 等），或其利用多變量分析得到的萃取因子。因為這些變數公布時，同期的民間消費也公布了，也沒有預測的意義。因此，總體經濟變數之 GDP 是否改為月頻率的資料（如實質薪資等）較為恰當？請再斟酌。 7. P.42 表 5-7 實證結果顯示，進階消費對民間消費有顯著之負向影響，似明顯不合預期，且就其係數的絕對值來看，甚至比一般消費的正向效果還大，是否可提供合理解釋？ 8. P.43 最後一段前兩行說明總體經濟面向各要素皆為明顯影響，但表 5-9 卻顯示無影響。 五、文字敘述與格式 1. 本文許多用詞不一致，建議重新檢視並修正，以利閱讀。例如：	Indicators vs. Google Trends，而在期末報告的內涵，則進行相關調整。 6. 本計畫在方程式的解說會加強下標的說明，在總體經濟面項，也透過相關經濟領先指標去進行模型分析。 7. 期末報告內之預測研究已進行 ARIMA 及 VAR 之模型驗證，並做相關說明。 8. 同上。 五、 1. 感謝您的建議，已對內文進行相關調整。
---	--

(1) 應用大數據所建立的指標名稱用詞不一。例如，P.7 最後一段將利用搜尋引擎及社群網站資料建置之指標稱為「民間消費預測綜合分析指標」，P.17 亦以此稱之；P.18、P.20 標題僅稱為「民間消費指標」；惟其餘又有部分稱為「民間消費預測指標」（如圖 3-1 等）。若這些指標皆相同，建議用詞一致，以免混淆；若不同，宜避免混用，必要時則應於首次出現時定義清楚。 (2) P.7 「臉書」、「Facebook」等。 (3) P.23 最後一段，找出共同因子的方法 A、方法 B，應即是 P.24 圖 3-4 的 DFM1、DFM2？請直接與內文對應。 2. 文中如出現英文專有名詞，應於第一次出現時完整列出全名以及簡稱，之後再用該簡稱即可。例如： (1) P.4 出現 VAR 模型的簡稱，但 P.29 才列出全名（且有重複）。 (2) P.12 最後一段，請註明美國經濟諮商會消費者信心指數（Conference Board Consumer Confidence Index）、密西根消費者情緒指數（Michigan Consumer Sentiment Index）分別簡	2. 本計畫在文字敘述及格式上，依照評審的建議做修正(第 3-5 點亦同)。
---	---

稱為 CCI、MCSI，以便與 P.14 再度出現時對照。 (3) P.42、P.43 MAE 中文應為「絕對均差」。且 P.43 中，MAE、RMSE 之中文與全文不必重複。 3. 本文章節呈現方式過於混亂，如第三章有第一節，而第四章僅用壹表示，第二部分則為二；字型亦不統一，建議修正。 4. 文中出現的任何圖表，建議於文中直接標示出與內文之關係，以利讀者閱讀。例如，表 1-1、表 1-2、圖 1-1、圖 1-2、圖 3-1、表 3-3、圖 3-2、圖 3-4、圖 5-2、表 5-3、圖 5-3 等，內文均未提及；另 P.22 文中指的圖 4 應為圖 3-3，P.23 圖 5 應為 P.24 圖 3-4。建議重新檢視並修正。 5. 文中數學式排版之大小、位置、格式與在內文中之表示符號皆不一致，建議進一步修改，以利閱讀。例如： (1) P.26 第 (3) 式，文中均未說明 j。 (2) P.27 第 2 段出現 (3.25) 式，惟並無此式，請確認其對應性；且第 2 段數學式與文字排列混亂，不易閱讀。 (3) P.30 第 (17) 式之變數代表為何，文中未說明；且 $\begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ R_{nt} \end{bmatrix} \text{ 應為 } \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}。$ 6. 建議重新檢視並修正文中之	6. 感謝評審的建議，各項目依照評審
---	---------------------------

筆誤或錯別字。例如： (1) P.4 倒數第 2 行，「個國家進行調查」；倒數第 8 行，「更具準確性」。 (2) P.6 第 2、3 行，「不可否認地」。 (3) P.11 第 2 段第 2 行，「衡量」；第 2 段第 3 行，「關注力」；倒數第 5 行，「...指標的代表性」。 (4) P.14 第 3 行，「總體經濟遭受」；倒數第 3 行，「...（以 S&P 500 為代表）」。 (5) P.15 倒數第 5 行，「世界盃粉絲專頁」。 (6) P.25 第 4 段倒數第 2 行，「Stock and Watson（2005）」。 (7) P.28 倒數第 3 行，「predictive causality」。 (8) P.29 標題，應為「向量自我迴歸（Vector Autoregression, VAR）模型」，且註腳 7 應上標。 (9) P.30 第 4 段，提出 VAR 者應為 Sims（1980）；倒數第 2 段最後一行，「反應函數」。 (10) P.30 標題，應為「自我迴歸整合移動平均（Autoregressive Integrated Moving Average, ARIMA）模型」。P.31 相關敘述亦同。 (11) P.34 倒數第 2 行，是否為「皆通過單根檢定」？惟部分變數取差分後，仍未	的建議做修正
--	---------------

通過單根檢定 (P.77)。 (12) P.75 資料頻率轉換的第一步驟與第二步驟敘述重複，應刪除其一。 7. 建議修正部分文獻引用格式，並檢視是否列於文末參考文獻： (1) 引用格式：P.4 最後一段及表 1-1→Brown (2014)；P.8 第 1 段第 1 行→Simon (1971)；P.28 倒數第 2 段→Granger (1967)。 (2) 未列於參考文獻 (P.50) 者，包括 Bai and Ng (2002)、Da, Engelberg and Gao (2011)、Sims (1980)、Stock and Waston (2002, 2005) 等。 8. 其他建議或疑問： (1) P.6 圖 1-2 資料來源為本研究，但本文似未就消費者網路行動模式進行相關研究，亦非本文主旨。 (2) P.12 表 2-1 的 21 個搜尋數列，宜以顏色區隔，以利閱讀。 (3) P.14 第 2 段，何謂美國消費者情緒指標與消費行為，有「很強的消費關係」？ (4) P.14 第 2 段最後一行，無須粗體。 (5) P.14 第 2 段提及 CCI 與 MCSI 具「共線性」，惟共線性 (collinearity) 一詞通常用於迴歸分析，若意指高度正相關，則敘明即	7.謝謝您的建議，期末報告內文圖示已做相關調整。 8. (1) 期末報告研究內容已針對消費者行動模式進行討論，相關內容請見第五章研究成果。 (2) 已參照修改。 (3) 已參照修改。 (4) 已參照修改。 (5) 已參照修改。
---	--

可。 (6) P.18 表 3-1 關於 Google Trends 關鍵字來源網中的人次占比加總超過 100%，建議說明人次占比如何計算。 (7) P.30 倒數第 2 段，VAR 模型的例子與本文的預測目的無關，易使讀者困惑。 (8) P.41 第 3 行提及模型的總體因子包括「國家生產力」，但其後變數卻用「人均 GDP」。兩者似有區別，若是用人均 GDP，則敘明即可，不建議提及「國家生產力」。 (9) 如圖 5-3、圖 5-4 與圖 5-5 等多圖合一的大圖，或可考慮單頁橫向表示，以利閱讀。	(6) 已參照修改。 (7) 已參照修改。 (8) 相關指標已重新調整。 (9) 相關圖示已做調整。
捌、金檢處 一、提供參考建議事項 (一) 為掌握網路大數據以及歸納網路爬蟲的訊息，本研究嘗試利用主成份法 (Dynamic Factor Model, DFM) 來建立預測民間消費的有用指標，在資料處理過程中須勞心勞力與任勞任怨，而且勞力密集度相當高，讓我們十分感謝，也非常敬佩!以下針對期中報告提出一些參考意見： 1. 模型遺漏重要解釋變數：就民間消費模型分析(p.41 及 p.43) 觀察，民間消費需求函數，在學理上一般採實質變數設定，表示當所得與物價同比例	1.本計畫會以主計總處公佈的實質消費支出當作被解釋變數，來避免移動變數的問題。

變化時消費需求不變（齊零次函數），至少在目前的變數中可另增CPI或民間消費平減價格來捕捉遺漏效果。 2. 解釋變數型態選取問題：消費函數中的利率，代表當前與未來消費支出的替代，通常採用短期利率，且就商品均衡（IS曲線）而言，民間消費函數中的利率項必須是實質利率（剔除通膨率）。 3. 方程式設定一致性：當解釋變數採用平均每人GDP時，被解釋變數須採用平均每人消費；參酌另種作法是直接以人口搭配GDP為解釋變數，以維持等號兩邊函數設定的一致性。 4. 民間消費預測：由於搜尋聲量的主成份很難推估未來值，目前研究只是樣本內的事後模擬，並不能從事預測；或可觀察搜尋聲量的前期項對當期消費是否有顯著影響，以提升未來預測。 5. 「Pokemon GO」與民間消費關聯請增加文字說明：股價指數在民間消費估計式並不具統計顯著性(p.42表5-7及p.44表5-9)，因此「Pokemon GO」聲量的主成份雖對中華電POK_S2412股票報酬率有影響(p.58附表2-2及p.59附表2-4)，但與本文的民間消費的關聯，可否請以文字完整補充說明。 6. 可否文字補充「Pokemon GO」主成份選取標準：在p.5表1-2	2.在解釋變數的選取上，本研究在期末報告的研究中，以總體領先經濟指標作為經濟面向指標。 3.本計畫目前以國發會之總體經濟領先指標為主要經濟面影響指標。 4.模型設定為動態的主成份分析，它會有落後期的設定，以落後期的方式來談預測力。 5.謝謝評審指導，「Pokemon GO」相關內涵已參考計量分析科之建議，相關內容已刪減。 6.本計畫在選取「Pokemon GO」關鍵字時，是以遊戲中常使用到的道具及
---	---

「Pokemon GO」的第一主成份組成項目中，左四列「超夢」並非本次開放的 145 種精靈，如果考量地區性可選亞洲限定版「大蔥鴨」，目前除了有違規以「寶貝蛋」來霸佔道館，台灣多數是以「暴鯉龍」來守道館，可用以觀察玩家的遊戲持久性（功力等級），可否請說明選取「超夢」的原因。 7. 因果關係檢定及 VAR 並未有實證結果：p.28 列示因果檢定及 p.29 列示 VAR 等理論說明，是否表示期末報告將會列出實證結果？ (二) 有關搜尋引擎大數據指標，本報告利用 Google Trends 資料萃取出四大主成份因子，其中第一主成份為一般消費，對民間消費有正面影響，第二主成份為進階消費，對民間消費則有負向影響，但觀察兩種主成份的細目，都包含日用消耗品及耐久財，似乎區隔不很明顯，尤其第二主成份「進階消費」雖以耐久財為主，但尚包含多項一般經常性消費如飲料、食品、書籍、文具等，似可補充說明一般消費及進階消費之差異。此外，若第一及第二主成份已解釋大部分民間消費，且第三主成份及第四主成份包含的細目很少（分別為 3 項及 1 項），將第三及第四主成份納入分析的意義似乎不大。 (三) P.42 水準值模型中，搜尋聲量第二主成份「進階消費」對民間消費總量為顯著負向影響，似可補充說明為何第二主成份「進階消費」對民	人物角色作為關鍵字來源依據，此次在台灣開放的角色人物是神奇寶貝的第一代為主，「超夢」則是第一代中，能力值是最高的，且擁有超能力的特殊技能，是玩家夢寐以求的精靈。 7.本計畫在期末報告的研究中已增加因果檢定的方法。 (二) 本計畫在做主成份時，已進行相關調整，並未再將主成份再做命名，因此不會有此問題再次發生。 (三) 本計畫在做主成份時，不會將主成份再做命名，因此不會有此問題再次發生。
---	---

間消費總量呈負向關係。 二、文字及圖表之修正建議 1. 部分圖表與本文未連結，例如第一章的表 1-1、表 1-2、圖 1-1 及圖 1-2。 2. P.7 第 3 段第 1 行「社群網站將所有使用者使用者的想法」，刪除贅字「使用者」。同段第 4 行「透過電腦的的運算能力」，刪除贅字「的」。P15 倒數第 2 段亦同。 3. P.11 第 2 段第 2 行「此指標可直接橫量民眾」，修正為「此指標可直接衡量民眾」，同段第 3 行「觀注力」，修正為「關注力」。 4. P.14 倒數第 2 行「不管是在樣本內、樣本的預測外」，修正為「不管是在樣本內、樣本外的預測」。 5. P.40 第 1 行「電商 Facebook 紛絲團」，修正為「電商 Facebook 粉絲團」，第 3 行「資訊的累績」，修正為「資訊的累積」。 6. P.75 的第一及第二步驟的內容說明有重複情形。	1. 謝謝評審的建議，本計畫將會在文字及圖表上做修正。 2. 感謝評審的訂正，已將內文進行調整。 3. 感謝評審的訂正，已將內文進行調整。 4. 感謝評審的訂正，已將內文進行調整。 5. 感謝評審的訂正，已將內文進行調整。 6. 感謝評審的訂正，已將內文進行調整。
--	---

附件三 期末審查意見及回應表

「應用大數據提升台灣民間消費預測」委託研究計畫 期末報告審查會

時 間：民國 105 年 12 月 15 日上午 10 時至 11 時 30 分

地 點：中央銀行第 2 大樓第 1102 會議室

主 席：林處長宗耀

報告人：郭副教授迺鋒（世新大學財務金融學系）

出 席：

評論人：林教授馨怡（政治大學經濟學系）

王教授淑美（台灣大學生物產業傳播暨發展學系）

經研處：林處長宗耀、林副處長淑華、吳副處長懿娟、廖研究

員俊男、張副研究員天惠、鄭副研究員漢亮、吳副研

究員俊毅、陳專員佩玗、林專員依伶、朱專員浩榜、

蕭宇翔、李宗憲

金檢處：潘稽核雅慧（代理侯研究員德潛出席）、盧專員月雲

外匯局：曹副研究員體仁

業務局：張專員國興

壹、評論人意見與報告人答覆：

評審意見	研究團隊回應
壹、林教授馨怡： 一、期中報告回應、文獻回顧： 1. 研究團隊回應期中報告之說明，本人不易了解。建議應有更明確之回應，例如修正後之文字段落位於報告中第幾頁。 2. 期中報告時，建議目前預測標的與零售資料較接近，而與總體民間消費有落差。但研究團隊之回應為“本研究目前透過更詳盡之模型設計來預測我國民間消費走勢”。但期末報告中，並未有相關說明，希望能補充說明。 3. 應加強文獻回顧，例如：報告中所提之 Schmidt and Vosen (2009) 應訂正為 Vosen and Schmidt (2011)。 二、模型與實證： 1. 期末報告附表 4-1 的單根檢定表所提供之訊息並不完整。再者，由民間消費之 P-value 觀之，並不拒絕單根之虛無假設，但研究團隊並未提出如何解決。若忽略此單根問題，	一、 1. 本計畫在文字敘述及格式上，期末報告評審回覆將會依照評審的建議做修正，謝謝評審的提醒。 2. 本研究之研究目的應用大數據提升民間消費預測，透過網路搜尋（檢索）指標，來比較傳統僅採用總體經濟變數來預測民間消費走勢之相關效果，故就民間消費總量為研究標的，仍有其必要性。受限於目前的時間及資源影響，仍以民間消費總量為主要預測目標。未來研究將建議可多方測試，以零售資料的進行評估。 3. Schmidt and Vosen 為兩作者在 2009 年於 Ruhr Economic Papers 刊登之作者序，後續其以 Vosen and Schmidt 的作者序於 2011 年在 Journal of Forecasting 刊登，內文以全數調整為 Vosen and Schmidt (2011)。感謝評審提醒。 二、 1. 本研究於期末報告修正中，已提供完整模型操作之訊息，詳細資訊請見 P.43 及 P.58，謝謝評審的建議。

評審意見	研究團隊回應
則總體實證之統計推論將不正確。 2. 期末報告並未有預測模型之建構、設定說明等，讀者無法了解本計劃的預測方法。建議研究團隊可以有預測模型建構等進一步說明。 3. 期末報告沒有預測模型之比較及相關檢定，讀者無法了解本計劃是否提出提升民間消費預測的模型或方法。研究團隊可參考使用如 RMSE, MAE 等預測誤差比較，或是 Diebold and Mariano (1995)、Clark and McCracken (2001) 和 Clark and West (2007) 等檢定。 4. 在期中報告建議中，本人建議高維度問題，可以使用其他變數挑選方法，但研究團隊回應說明並不容易理解。如：“採納總體經濟領先指標做為目標變數”？ 5. 於資料量龐大，雜訊也相當多，即使研究團隊不使用 LASSO 等變數挑選機制，也建議在預測前先進行因果關係檢定，排除雜訊，過濾有用訊息，但報告中的附件四所提供之因果關係檢定訊息太少，無法判斷。 三、文字及報告內文： 1. 第四章，研究方法部份應統一所有定義及符號，太瑣碎的部份建議直接刪去，以利讀者閱讀。 2. 第五章第四節關鍵社群網站分析並未說明其和民間消費預測之關聯，故未能提供讀者太多訊息，建議刪除。	2. 本研究於期末報告修正中，已提供完整模型操作之訊息，詳細資訊請見 P.44，謝謝評審的建議。 3. 研究於期末報告修正中，已提供完整模型操作之訊息，詳細資訊請見 P.48，表 5-15。謝謝評審的建議。 4. 謝謝評審的建議，目前本研究在預測模型設計為考量傳統指標如 GDP、通膨及由本研究萃取之總體經濟領先指標，進行多樣化的模型估計及預測績效檢測，詳請見第五章實證研究結果。 5. 在進行模型變數挑選時，本研究以相關分析等分析方式，將相關性較高之變數做交錯的篩選，避免影響模型預測效力。 1. 謝謝評審的建議。 2. 有關關鍵社群網站分析，乃新興指標，雖目前受限於資料期間，無法納入預測模型之中，但透過多變量分析，本研究希望發掘其中脈絡，為民間消費預測未來研

評審意見	研究團隊回應
3. 附件五並未能提供讀者太多訊息，建議刪除。 四、未來修正建議： 1. Vosen and Schmidt（2011）提出即時預測美國民間消費，應為最適切本計劃之參考文章，應增加說明該文章。也建議研究團隊可以參考該文章作法。 2. 研究團隊可以先說明如何萃取大數據資料中的有用訊息，並以此建構預測模型，接著比較預測模型之表現，選擇預測表現最佳模型。最後，說明及比較使用大數據訊息之預測模型是否能夠提升台灣民間消費預測。如此聚焦於台灣民間消費預測評估應更能彰顯本研究貢獻。	究提供更多新興資訊。 3.已參照評審意見刪除。 1.目前本研究在實證分析，在前ARIMA 模型設計 ARIMA、ARIMAX+ 傳統模型及 ARIMAX+GT、ARIMAX+PCA 總指標及 ARIMAX+DFM 等，透過如 RMSE、MAE、MAPE、以及 DM Test 等客觀指標，比較模型預測效果，據以提供政策制定者方向，與 VAR 模型亦有相關檢測，請見 P50、P.55、P.62、P.64。 2.謝謝評審的建議，於第五章實證研究，已做相關內容調整，透過比較預測效果，聚焦於網路大數據指標對於民間消費預測之貢獻，請見第五章第二節及第三節之資訊。
貳、王教授淑美 本人肯定本計畫中，央行希望利用大數據提升民間消費預測的想法，並肯定研究團隊的努力。因為本人從事電商之研究，民間電商亦常藉觀察情緒、關鍵字等網路輿情作為當季行銷的參考依據。由於林教授關於計量模型之建議已相當完整，故不再贅述，以下謹對其他部分補充相關建議： 一、站在政策制定者的立場，既有預測模式與新的預測模式相較，兩者關聯度為何？由於既有指標已有一定程度的預測能力，故新建的網路檢索指標有哪些新的益處？例如，可	一、謝謝您的建議，本研究目前 ARIMA 模型設計 ARIMA、ARIMAX+ 傳統模型及 ARIMAX+GT、ARIMAX+PCA

評審意見	研究團隊回應
提供即時性的訊息，作為當季、當月、當週等短期即時監測之用，有助政策制定者及時反應。 二、建議使讀者更易於了解研究報告結論或建議的意涵。例如，網路檢索的即時性指標有兩個，其中表現較好的指標包括哪些主成分的因素？ 三、目前模型或許有助掌握當期消費，但輿情等網路情資含有許多即時性的資訊，故應可進一步發展，進行下一個月、甚至下一週等更短期內的預測，以提供政策制定者即時性的回饋，且當今黑天鵝事件眾多，有助縮短預測及反應上的時間落後。	總指標及 ARIMAX+DFM 等，透過如 RMSE、MAE、MAPE、以及 DM Test 指標，比較模型預測效果，據以提供政策制定者方向。 二、有關網路檢索指標之組成，已於附件六納入說明。
參、侯研究員德潛 一、從學理上的消費函數來表現推估解釋變數：當期實質所得、實質利率的未來值，然後透過這個模型求出預測值，本文是直接利用領先指標的主成份來推估，給我們的感覺比較突兀，可以用傳統方式表現一個，添加大數據檢索是樣本內、樣本外的估計結果，都把它呈現出來，讓我們更容易理解。 二、通常台灣人可能會在淘寶買一些便宜的大陸產品，不會在淘寶網買的台灣消費者，所以現在主計總處的民間消費也可能不包含這些東西，那我不知道這裡列淘寶網的分析是不是那麼有必要性，如果沒必要就拿掉算了，因為淘寶跟我們的民間消費關聯沒有那麼密切。	一、謝謝您的建議本研究目前在 ARIMA 模型設計 ARIMA、ARIMAX+ 傳統模型及 ARIMAX+GT、ARIMAX+PCA 總指標及 ARIMAX+DFM 等，做預測模型比較，並可從中觀察網路檢索指標是否對民間消費預測有所助益，於 VAR 模型中亦有做相關處理，請見第五章實證研究，請見 P50、P.55、P.62、P.64。 二、本研究將淘寶網資訊建立網路檢索指標之考量，主要參考 EagleEye 所公佈之 2014 年第 4 季「台灣購物網站排行榜」中前 50 大電商網中所公布之資訊，台灣民眾於淘寶網消費具有一定的市場比例。

評審意見	研究團隊回應
三、我們現在民間消費可能遺漏的就是所謂 Uber，但是資料相當不容易取得的，是不是可以稍微談一下有這個東西，但是我們資料不太可能拿到，就可以了。 四、主成份分析是我看結論上說他的 VAR 分析是本文的重點，但是這個 VAR 分析我只看到估計值沒有看到他的衝擊反應變數跟他的變數，是不是可以一併列出來，因為就你 49 頁主成份分析 ARIMA 這邊，我看他的結果，ARIMA(0,1,1) 部分它的 MA 是很顯著的，但是當 AR(1) 擺下去以後，這個 MA 馬上就變得不顯著，但是就 2-2 你擺的網路檢索指標以後，AR(1) 也受到影響，顯示這個網路檢索指標相對來說是有用的，但是到 VAR 這邊，因為它本身的限制是只能擺 lag，沒有同期的結構關係，所以這裡擺的幾乎都是 lag，lag 以後我就發現，那事實上民間消費總和預測 2 那邊是比較有作用力，所以要詳細了解他是不是真正的衝擊影響有多大呢？可能要看一下衝擊反應跟變數，幫助我們提供一些表給我們做參考。 至於其他的細部意見，書面資料會提供郭副教授參考。	三、對民間消費進行預測，各式與消費有關之資訊，皆可能有其預測效果，如 Uber 付費資料，或 ATM 領取資訊，惟部分資料難以取得（如 Uber 付費資料），而本研究目前以網路檢索指標為目標。 四、謝謝您的建議，第五章研究結果於，本研究已納入 ARIMA、ARIMAX- 傳統模型及 ARIMAX-GT 模型，並透過 RMSE、MAE、MAPE、以及 DM Test 進行納入網路大數據對民間消費預測之模型比較，於 VAR 模型也納入衝擊反應及預測誤差拆解資訊，請見附件七及附見八。
肆、廖研究員俊男 一、本計畫是大數據不是大計量，不過我想它也不是小計量，他這個東西既然是題目，我們當初規劃	一、本研究已納入 ARIMA、ARIMAX- 傳統模型及 ARIMAX-GT 模型，並透過 RMSE、MAE、MAPE、以及

評審意見	研究團隊回應
的題目就是要改善預測，需要提 供給我們提升的幅度有多少，並 詳細的說明整個研究成果。 二、網路檢索的結果只有 0.0008，這個 東西對預測是有很大的幫助嗎？ 至於在未來如果這個變數有很好的 實用性的話，未來如何去設定， 可能需要在文章中交代一下。 三、只用了 ARIMA、DFM 這兩種形 式，因為其他還有很多種形式，那 當初沒有考慮的原因是什麼，可能 也可以做個交代。 四、總體經濟指標，事實上它國發會就 已經編了領先指標，那個數列是很 完整的，那你們今天拿的那六個， 只是其中那六個抓出來，少了台經 院的營業氣候測驗點，只少了一個， 因為這個東西是跟實質 GDP 相對應的，所以我不知道說你們何 苦要再去弄一個主成份。 五、網路社群的資料如果要留下來，需 要跟我們交代清楚，它跟民間消費 預測的關係，未來你要怎麼應用， 計畫強調是大數據，在技術移轉方 面，我們希望包含前面 Google Trends 那些，還有後面的 Fb 或是 粉絲專頁那些，到時候給我們上 課，可以讓我們比較完整的了解。	DM Test 進行納入網路大數據對 民間消費預測之模型比較，於 VAR 模型也做相關模型比較設 計，謝謝您的建議，請見第五章 實證研究。 二、目前本研究之實證分析已進行 模型調整，相關結果請見第五章 實證分析，此外，有關網路檢索 指標及民消費綜合指標之編 制，已於第五章第一節進行相關 說明。 三、目前本研究已納入 ARIMA、 ARIMAX 進行模型比較，詳請 見第五章研究結果。 四、為編制民間消費預測綜合指 標，本研究考量總體經濟面、消 費者心理及實際消費檢索狀 況，故採納多總體經濟領先指 標、消費者信心指數、及網路檢 索指標，透過主成份分析，期望 以多樣化的指標建置，觀察網路 大數據對民間消費預測之成效。 五、未來將在技術轉移時，提供 Google Trends 及 FB 粉絲專業部 分，進行相關說明，以及資料提 供。
伍、鄭研究員漢亮（經研處） 一、 全文架構 本文主要目的是使用總體數據、搜 索引擎大數據及關鍵社群網站大 數據等，利用主成分析拆解成若干	

評審意見	研究團隊回應
主成分，再透過向量自我迴歸（VAR）及 ARIMA 模型預測民間消費水準值及成長率。在此，謹提出以下意見供參考： 二、報告內容 （一）架構 「第二節搜尋大數據預測應用研究」，諸多文獻談及大數據對於汽車銷售及旅遊消費應用，與「第三節網路大數據消費預測應用研究」多有重複，建議將其合併。 （二）模型假設及實證分析 1. 誠如第三節所言，文獻上研究大數據對消費的預測能力經常與消費者信心指數等總體變數進行比較，或是在預測模型納入金融面因子、所得、股價等總體變數，比較總體變數及 Google trends 變數之解釋能力。因此，建議將民間消費預測綜合指標及網路檢索指標的預測能力，與總體領先指標進行比較，以凸顯大數據消費預測應用相對傳統方式之重要性及必要性。 2. 頁 44 主成分分析中，網路檢索指標主成分平方累積變異量，累積至第二因子達 0.945%，然而主成分總指標，平方累積變異量累積至第二因子，只有 0.747%。主成分總指標解釋能力相對較低，是否有必要對綜合總體領先指標、消費者信心及網路檢索指標再進行主成分	（一）謝謝您的建議，兩節之規劃分別為「搜尋大數據預測應用研究」針對其他總體經濟變數預測，而「網路大數據消費預測應用研究」針對民間消費預測，後者為本文研究主旨，故透過兩節分別陳述。 （二） 1. 謝謝您的建議，本研究目前在 ARIMA 及 VAR 模型，分別就網路檢索指標、總體經濟領先指標及民間消費綜合指標等組成進行估計，並透過 RMSE、MAE、MAPE、以及 DM Test 等指標來進行預測效力比較，詳請見第五章之實證研究說明。請見 P50、P.55、P.62、P.64。 2. 本研究建置多樣化指標，網路檢索指標、民間消費綜合指標、希望透過多種變數組合與模型估計，找到能增進民間消費預測效果之成果，故再次進行主成份萃取出民間消費綜合指標亦為相關測試及努力方向之一。

評審意見	研究團隊回應
分析值得再斟酌。 3. 本文雖列出民間消費預測指標及網路檢索指標 ARMIA 樣本內配適能力，惟並沒有進一步比較不同模型之樣本外預測能力，建議補上相關討論。 4. VAR 模型分析除進行樣本外預測，建議仿照 Pavlicek (2014) 進行因果分析、衝擊反應及預測誤差拆解，以支持 Google trends data 對消費具有明顯影響力。 (三) 其他 1. 頁 58，「比較主成分模組之成果，於 VAR 模組，主成分模型配適度仍相對較高，而在樣本外預測表現，主成分模組亦較能展現變動趨勢」。為確保結論正確，建議可使用 Diebold and Mariano (1995) 檢驗不同模型的樣本外預測能力。 2. 頁 44，第五行「見附錄六」應改成「見附錄五」，第十行「相關分資」改成「相關分析」。 3. 頁 43，第五行陡坡值圖「(圖 5-3)」應為 (圖 5-5)。	3.謝謝您的建議，已增加 ARMIA 樣本外分析，相關內容請見第五章實證研究內容。 4.謝謝您的建議，目前於 VAR 分析中，已增加因果分析、衝擊反應及預測誤差拆解資訊，詳請見附件七級附件八。 1.謝謝您的建議，目前本研究已納入 Diebold and Mariano (1995) 來檢定模型預測效力，詳請見第五章實證研究說明。請見 P50、P.55、P.62、P.64。 2.謝謝您的指正，已做相關內文調整，故刪除相關文句。 3. 謝謝您的指正，已做相關內文調整，故刪除相關文句。
陸、業務局 一、 文字更正：P9 第一段第 4 行借與？；P10 圖 2-1 橫座標 HHI Duration，直座標 Fraction of Time in 10+ Min Session，內文未說明，宜加註解釋；P17 第二段第 2 行最為消費變項；P62-63 圖 5-18 盒鬚圖、圖 5-19 分配圖上文字字型太小且不夠清楚，宜加黑放大。 二、 期中報告曾建議要應用大數據提	一、 (1)以調整文字為借用，請見 P.5(2)已加入圖 2-1 之備註說明(3)已修正，請見 P.13(4)已進行調整，請見 P.75-76。 二、受限於研究時程，故目前本研

評審意見	研究團隊回應
升模型預測，所用數據（data）須通過強外生性（strong exogeneity）檢定，報告結論將此檢定列為未來研究方向。事實上，第五章研究成果，作者只是做了樣本內的模擬預測，而不是樣本外的推估預測，兩者在應用及處理方式，有所不同。 三、根據 P23 圖 3-2 之研究架構，社群網站爬文所擷取的大數據資訊可轉化為時間數列，進而執行因果分析，包括 1.向量自我迴歸模型；2.衝突反應；3.預測誤差拆解。然而，第五章實際研究成果中，作者只用主成份及大數據資料迴歸得到 ARIMA 及 VAR 模型用以預測，後續衝突反應及預測誤差拆解分析，皆被省略了，而且期中報告曾經有人建議，要先建立 Baseline 模型以作為各種模型預測之比較基準，似乎也被忽略了。	究於第五章針對 ARIMA 模型樣本內及樣本外進行分析，未能處理強外生性檢定，故納入未來研究方向之一。 三、謝謝您的建議，本研究目前在 VAR 模型中，建立 VAR 傳統模型、VAR-GT 模型及 VAR-PCA 及 VAR-DFM 模型，做預測模型比較，並可從中觀察網路檢索指標是否對民間消費預測有所助益，請見第五章實證研究。有關模型衝擊反應及預測誤差拆解資訊請見附件七及附件八。
柒、計量分析科 一、模型對民間消費的預測結果 相較於期中報告，本次期末報告增加納入大數據指標（如網路檢索指標），對民間消費的預測（或配適）結果，惟仍不易從內文判斷，加入大數據指標是否對民間消費的預測具提升效果，理由如下（以下僅就主成份分析提供意見供參，因 DFM 預測結果較主成份為差）： 1. 主成份之 ARIMA 模型分析（P.48） (1) 作者僅比較不同模型的樣本內配適度（以 Log Likelihood 衡量），包括加入網路檢索（模型 1-2、2-2）、民間消費預測綜合指標（模型 1-3、2-3），以及未	一、 1. (1) 謝謝您的指正，本研究已針對民間消費預測綜合指標（PCA 及 DFM）、網路檢索指標，分別就 ARIMA 模型之樣本內及樣本外，進行預測比較分析，並就

評審意見	研究團隊回應
加入大數據指標（模型 1-1、2-1），惟此並不保證樣本外也有同樣的結果，因此建議在 ARIMA 模型亦進行樣本外預測分析（如 Schmidt and Vosen（2009）的作法，且本文在 VAR 模型有樣本外預測結果，但在 ARIMA 模型沒有，造成前後文說明不一致）。 (2) 對於加入大數據指標之模型配適結果，是否有明顯優於未加入大數據指標部分，建議應採相關統計檢定（如 Diebold and Mariano 檢定）來驗證。否則僅從 Log Likelihood 數值大小、圖 5-9，並無法得知該改善是否具有統計顯著性。 (3) 一般文獻上較常使用的配適度為 $\text{adj. } R^2$，而較少見 Log Likelihood，建議可多呈現模型配適與預測表現評估的相關指標，如樣本內配適可列出 $\text{adj. } R^2$，預測表現可列出預測均方差（mean squared error, MSE）或均方差的平方根（root mean squared error, RMSE）。 2. 主成份之 VAR 模型分析（P.51）作者僅列出加入大數據指標的 VAR 估計結果，而未比較沒有加大數據指標的結果，此不僅與 ARIMA 模型之作法不一致，且不易比較有加入及未加入大數據指標的 VAR 預測結果。因此，讀者無法從圖 5-10 的 VAR 樣本外預測表現（且僅預測 2 季），得知預測結果是否	VAR 做樣本外預測分析，並分別針對傳統變數模型，施以如 RMSE、MAE、MAPE、以及 DM Test 進行納入網路大數據對民間消費預測之模型比較，詳請參考第五章之研究結果。謝謝您的指正，請見 P.50、P.55、P.62、P.64。。 (2) 謝謝建議，本研究已採用 RMSE、MAE、MAPE、以及 DM Test 進行預測模型效度比較。 (3) 謝謝您的建議，本研究在預測分析已採用 MSE（ARIMA 及 VAR）及 RMSE（ARIMA）進行效度比較。詳見第五章分析結果。 2. 謝謝您的指點，本研究在 VAR 實證研究已納入多種模型比較，包含 VAR 傳統模型、納入檢索指標之 VAR-GT 模型、以及民間消費綜合指標（PCA 及 DFM）之模型比較，並施以預測效度分析，詳見第五章實證結果，請見 P.50、P.55、P.62、P.64。

評審意見	研究團隊回應
有所改善 3. 為凸顯加入大數據資料可以提升民間消費預測特點，建議作者可先在模型內考慮總體經濟領先總指標與消費者信心指數，接著再加入網路檢索總指標，檢驗網路檢索總指標是否含有額外協助預測民間消費的訊息。 二、模型設定 1. 建議本文將模型的預測結果轉換成民間消費年增率，並計算民間消費年增率預測值的預測誤差。 本報告預測係針對民間消費的水準值(且為取 log 後的民間消費水準值，如 P.50) 進行預測表現的評估，惟對央行與政策執行者而言，更關心民間消費的成長率（或年增率）。 2. P.37 可見民間消費有其季節性，但本文皆未調整，且落後期數頂多 1 至 2 期，恐影響估計或預測結果。事實上，本文參考之 Matsumoto et al. (2013)，在預測旅遊支出時，亦經季節調整（P.13）。 3. P.43、44、45 皆指出本報告篩選因素的標準為特徵值大於 1，或是累積變異數百分比達 70%以上，然而 P.32 在介紹主成分分析時卻寫道「利用 Bai and Ng (2002) 所發展出的資訊判定準則（information criteria）判斷最適因素個數。」建議將有關 Bai and Ng (2002)	3.謝謝您的建議，本研究目前在 ARIMA 模型設計 ARIMA、ARIMAX+ 傳統模型及 ARIMAX+GT、ARIMAX+PCA 總指標及 ARIMAX+DFM 等，做預測模型比較，並可從中觀察網路檢索指標是否對民間消費預測有所助益，請見第五章實證結果。 1.謝謝建議，本研究在進行模型測試時，受限於時間人力，建議未來透過更多民間消費代表變數進行模型預測。 2. 謝謝您的建議，惟本計畫之研究重點是預測 再樣本點不足情況下,故進行迴歸季節調整，以避免喪失過多資訊，難以進行後續的模型預測。 3.謝謝您的建議，已將相關敘述刪除。

評審意見	研究團隊回應
的敘述刪除，以免讀者誤解。 4. P.48 ARIMA 的設定僅包括 ARIMA (0,1,1) 及 ARIMA (1,0,1)，惟尚有 ARIMA (1,1,1)、ARIMA (1,1,0) 等組合可考慮，作者宜解釋為何模型僅有上述 2 種（是否依照哪些準則選取?）。 5. 本文擬運用大數據提升民間消費預測，惟本文採用的模型 ARIMA，模型設定可寫成： $y_t = c + \sum_{j=1}^p \alpha_j y_{t-j} + \beta x_t + \sum_{k=1}^p \gamma_j \varepsilon_{t-k}$ 其中，x_t 為本文萃取出之民間消費預測綜合指標或網路檢索指標，x_t 與 y_t 時點皆為 t 期，則此模型設定係對當期民間消費做預測（即 nowcasting），似無法對未來 h 期的民間消費做預測，$h \geq 1$（除非先計算出 x_{t+h}）。若需對未來 h 期做預測似乎僅能利用本文的 VAR 預測模型。作者應針對如何實際運用本報告的模型做長短期預測加以說明。 6. P.51 通常 VAR 模型會用來進行衝擊反應分析與預測誤差變異數分解等，很少以 VAR 模型來分析解釋變數之顯著與否，或進行預測。因此，若僅要預測民間消費，似無必要使用 VAR 模型，使用自變數、應變數皆落後兩期之 OLS 估計結果即可。 7. P.54 DFM 模型中的方法 B 僅考慮總體經濟指標 1 與網路檢索	4. 本研究目前採用多種指標進行預測模型效度觀察及比較，如 RMSE、MAE、MAPE、以及 DM Test 進行預測模型效度比較，請見第五章實證結果說明，請見 P.50、P.55、P.62、P.64 DM Test 之分析結果。 5. 謝謝您的建議，本研究於第五章實證研究已做相關調整，據以進行相關比較，請見 P.41、P.57。 6. 謝謝您的建議，在第五章實證研究 VAR 分析處，估計結果說明已做相關調整，請見 P.52。 7. 本研究在設計方法 B 之成份挑選十，以其因子負荷量為依據，較

評審意見	研究團隊回應
指標 1 及 CCI，但是從前面的主成份分析 (P.45) 可知，總體經濟指標 2 與網路檢索指標 2 與民間消費的相關性分別達 0.94、0.55，而總體經濟指標 1 與網路檢索指標 1 與與民間消費的相關性卻為-0.06、-0.77(見 P.45)，因此作者宜說明為何方法 B 是挑選總體經濟指標 1 與網路檢索指標 1，且 DFM 之因子挑選應有其依據 (詳 Bai and Ng (2002)) 而非主觀認定，以符合標準研究程序。 8. 期中報告曾設定民間消費需求函數，並說明初步研究成果，不知為何捨棄不用，建議仍應納入本文研究，並力求周延與嚴謹。 9. 民間消費指標為何僅侷限在國發會編製之領先指標下之 6 項指標 (僅差營業氣候測驗點乙項)，若是如此，何必再大費周章編製此總體指標，直接拿國發會領先指標即可。正確做法應是考慮所有可能影響民間消費的所有指標，再萃取出單一指標。 10. 本文 P.49 實證結果與 P.111 第六章結論說明相當混亂不清，部分說明亦有誤，請再檢視釐清。 三、文章內容 1. P.9 建議增列表格方式整理文獻回顧，以利讀者了解每篇文獻的研究主題、方法、所用資料等，目前安排過於凌亂，難	可代表該主成分之元素做挑選，故皆以第一主成分為主。 8. 目前本研究對預測模組施以如 RMSE、MAE、MAPE、以及 DM Test 進行納入網路大數據對民間消費預測之模型比較。詳見第五章實證研究內容，謝謝，請見 P50、P.55、P.62、P.64DM Test 之分析結果。 9. 為編制民間消費預測綜合指標，本研究考量總體經濟面、消費者心理及實際消費檢索狀況，故採納多總體經濟領先指標、消費者信心指數、及網路檢索指標，透過主成份分析，期望以多樣化的指標建置，觀察網路大數據對民間消費預測之成效。 10. 謝謝評審的建議，已針對 ARIMA 模型做進一步的分析，請見 P51 後的詳細分析；結論的部分請見 P.86M。 1. 謝謝您的建議，礙於目前研究時程問題，未能即時彙整，建議未來延伸研究可以更清楚之方式整理之。

評審意見	研究團隊回應
了解文獻發展脈絡。 2. P.36 指出對粉專的敘述統計分析上，會採用文字雲的方式呈現，惟後續的實證研究，並無相關的分析。 3. P.43 所謂總體經濟指標之建置未說明清楚，是否係萃取 P.39 表 5-1 除「民間消費支出合計項」外，其餘 7 項指標之主成分？ 4. P.45 總體領先指標 1 與民間消費相關係數為-0.06，指標 2 則達 0.94，但圖 5-8 卻呈現相反結果，請再檢視兩者一致性。 5. P.49 消費者信心指標並未用於主成分分析，原因為何？ 6. P.49、P.51 估計結果表格應陳列出估計係數的標準差或其他統計量（如 t 值），而非只告訴讀者其統計顯著。 7. P.57 第一段敘述應是指與 P.49 表 5-10 比較，建議說明應嚴謹明白，究竟係與 ARIMA 何者模型比較。 8. P.60 關鍵社群網站指標部分 (1) 本報告選擇 8 個網購粉專的資料，惟此與 P.41 表 5-2 所列 10 個關鍵社群網站大數據觀測站不符。此外，日本樂天市場-台灣，因關鍵字資料不足，無法進行盒鬚圖、分配圖等分析，是否仍有保留該網購粉專之必要？	2.受限於研究時程，以及民間消費預測之研究內容，目前則不進一步對發文文字進行探勘，建議後續研究時可朝文字探勘進行相關研究，謝謝您。 3.總體經濟指標之萃取，為以領先指標為主，由外銷訂單指數、股價指數、工業及服務業受僱員工淨進入率、核發建照面積、SEMI 半導體接單出貨比，經由主成份萃取而得。 4.謝謝您的指正，已做相關內文調整。 5.在以 PCA 及 DFM 萃取民間消費綜合指標時皆有納入消費者信心指標，再請查核。 6.受限於版面限制，故目前本研究以估計係數搭配由 P-value 所判定之顯著水準標記呈現，謝謝您的建議。 7.謝謝您的建議，考量模型差異，本研究目前依據 ARIMA 及 VAR 個別進行預測效度比較，請見第五章實證研究之相關說明。 8. (1)謝謝您的建議，研究團隊以 8 個網購粉專的資料作為分析，發現樂天資料相對明顯少，故不針對日本樂天市場-台灣做分析。

評審意見	研究團隊回應
(2) 作者指出，「情緒指標圖，可以看出對該商品的正面或負面評價與喜好程度；盒鬚圖與分配圖，可以看出對該商品的正面或負面評價與喜好程度。」，惟情緒指標圖、盒鬚圖與分配圖有何差異？建議作者補充說明。 (3) 有關粉專關鍵字詢問度、正負面情緒、喜好度最高時點是否與民間消費成長率變化有關？又粉專的各種指標多為右尾分配，其意義為何？建議作者多作說明。 (4) 不同相關性與連結性分析結果可能彼此衝突，如 P.109 指出，相關性圖結果顯示，Yahoo 奇摩購物中心「家電」、「視聽」、「電玩」搜尋相關性高，惟樹狀圖的結果顯示，「戶外」、「票券」為主軸的關鍵字，兩種分析結果似無明顯關聯，又如何詮釋此結果？	(2) 謝謝您的建議，分配圖的用途在於聚資微觀，8 個網購粉專經過 Fb 爬蟲出來的資料數目相多當，透過分配圖的展現，可以快速透過視覺了解該資料的資料型態；而盒鬚圖涵義為五數會總，針對第一位數、中位數、第三位數、最小值和最大值及超過三個標準差的極端值，有助於進一步深入分析，請見第五章第四節之說明。 (3) 研究團隊尚未將網購粉專資料與民間消費成長率作結合，目前只探討 Google Trends 與 Fb 的連結度，且因為網購粉專資料量不足無法探討民間消費成長率，因此採 Google Trends 的資料做詳細探討。網購粉專的資料型態呈現出來的多為右尾分配，表示多數不為常態分配，且資料顯示為右尾分配，可以從分配圖的結果知道該筆資料是否為常態的集群分配，有助於研究團隊在深度上的理解，故放入內文做為參考資料，請見各網購粉專的分配圖。 (4) Yahoo 奇摩購物中心的相關性圖，將「家電」、「視聽」、「電玩」歸類為同一群組，彼此的相關性強烈；樹狀圖是以「家電」、「視聽」、「電玩」做關鍵字樹狀圖的主軸，向外做其他關鍵字的衍伸，可以知道 Yahoo 奇摩購物中心是以該三個關鍵字為起點，向外做進一步的延伸，其中延伸到樹狀圖最底層的關鍵字為「戶外」、「票券」，進一步瞭解到這兩

評審意見	研究團隊回應
(5) 本報告囿於網購粉專資料多由 2014 年起，故尚無法納入消費預測分析，並指出未來待資料量累積足夠可用於預測消費，作者或可先提出指引方針，說明如何運用粉專資料進行民間消費預測，以供本行未來擴充運用。 9. P.61 第一段第三行提及「第三大聲量」為「清潔」，不知當月係發生何事？ 10. P.154 附件四，請說明附表 4-1 之 ADF 設定，如是否存在截距、趨勢、落後期數等。 11. 建議作者增加全文摘要，並扼要彙整研究發現，並務必著重於闡明使用網路檢索與社群媒體民間消費輿情時間序列資料對於民間消費預測能力提升之有效性，並據此提出相關的政策建議，以利本行參考。 四、建議部分內容可刪減 1. 本文(不含附件)多達 111 頁，建議文內部分章節可做適度刪減，如： (1) P.26 圖 3-與圖 3-3 重疊的圖文方塊可刪除。 (2) P.26~P.29 第三節社群網路料擷取平台建立，與本文主旨較為無關，建議放	項商品是與其他關鍵字最具相關性。 (5)研究團隊尚未將網購粉專資料與民間消費成長率作結合，目前只探討 Google Trends 與 Fb 的連結度，且因為網購粉專資料量不足無法探討民間消費成長率，因此採 Google Trends 的資料做詳細探討。 9.目前本研究針對各電商網粉專進行敘述統計分析，受限於研究時程，進一步的事件文本分析並非為本研究目前範疇。 10.謝謝您的建議，目前本研究之單根分析，分別就存在截距、趨勢截距、無截距與趨勢進行檢定，相關內容。 11.已增加摘要說明。 四、 1. (1) 謝謝您的建議，但為求脈絡清楚說明，故不在此進行調整。 (2) 謝謝您的建議，已將相關內文調整至附件，請見附件四。

評審意見	研究團隊回應
在附件。 (3) P.30~ P.36 研究方法，作者詳列 ADF 單根檢定、因果關係檢定、ARIMA、VAR 模型，惟此皆為很常使用的基本計量方法，建議可縮短篇幅，且 VAR 部分，由於後續實證中並未進行衝擊反應及預測誤差拆解，故可刪除相關說明。另 P.35 第一段、第三段關於 VAR 之敘述多有重複，建議合併。此外，對於社群網頁多變量分析，如聲量圖、情緒指標圖、盒鬚圖、分配圖、相關性圖、樹狀圖、集群網絡圖等反而可多做介紹。 (4) P.37、P.38 既然本文估計或建置指標時，並無區分細項，似無須進一步觀察各細項之序列走勢，且表 5-2 與民間消費各細項無關。 (5) P.51、P.58 一般在說明 VAR 模型估計結果時，不會針對落後期係數解釋，因為若落後一期為負，落後二期為正，去解釋這結果並無意義，建議將相關文字刪除。 (6) P.61~P.106 關鍵社群網站指標中，詳列八家網購的個別指標分析，惟文內僅呈現、整理程式計算的「結果」，對結果的「分析」、「詮	(3) 謝謝您的建議，有關 VAR 模型衝擊反應及預測誤差拆解資訊已於第五章實證研究加以呈現及說明。有關社群網頁多變量分析，受限於研究時程，故未能增加相關說明，各分析圖表，於實證分析皆有特別說明。 (4) 民間消費細項資訊已調整刪除。 (5) 謝謝您的建議，內文說明已做相關調整，刪除相關部分。 (6) 謝謝您的建議，惟目前在關鍵社群網站指標分析，為透過各敘述統計圖呈現之，為求能解讀，在版面配置上較未能調整，而為求解讀上的脈絡，相

評審意見	研究團隊回應
釋」著墨甚少，使讀者不易從中瞭解各項結果的意義，且頁數長達 50 頁恐過於冗長。反而在 p.107 小結部分，作者對各家網購的結果說明較為清楚，且有進行比較，故建議可刪除 p.61~p.106 的結果。 2. P.157~ P.268 附件五之數值所能提供之資訊有限，建議可刪除此部分，或以電子檔方式提供本行參考即可。 五、其他 1. 數學式編碼請依順序編排。如第 34 頁第 (19) 式出現的位置早於第 35 頁的第 (17) 式。此外，文中數學式排版之大小、位置、格式與在內文中之表示符號皆十分不一致，建議作者進一步修改，以利讀者閱讀。 2. P.34 上半部 Granger 因果關係檢定之數學式排版過於混亂，應修正。 3. P.35 第四段，VAR 應為 Sims (1980) 提出。此錯誤期中報告已提出。 4. P.46 KMO 值大小代表之意涵未交代。 5. P.47 「觀察其相關性分析（請見表 5-7）」應為表 5-8。第二行表「5-7」漏掉。 6. p.54 內文註解 15，註解部分的標示卻為 16。 7. p.156 附表 4-3 主成分因素相關係數表細目，橫排應少一欄 FAC_COS_1（博客來），且表	關資訊較不建議刪除。 2.謝謝您的建議，已刪除相關資訊。 1.謝謝您的指正，已做相關編號調整，並做相關調整，請見 P34。 2.謝謝建議，後續於 VAR 模型因果分析已做相關調整，請見附件七及附件八。 3.謝謝指正，已進行調整。 4.已於註腳 11 進行相關說明，請見 P44。 5.謝謝指正。 6. 謝謝指正。 7.謝謝您的建議，相關內容已進行調整。

評審意見	研究團隊回應
中相關係數排列有誤。 8. p.157 附件五 15 大電商主成分分析成果，表中所列之主成分數值是否為 factor loading？請補充說明。 9. 部分名稱不統一： (1) P.7 將 Big Data 譯作巨量資料，與本文其他處之大數據不符。 (2) P.15 消費者情緒指數、消費者信心指數之英文已於 P.14 出現。 (3) P.14 使用原文 the Conference Board，P.15 卻使用美國經濟諮商會之譯名。 (4) 動態因子模型（如 P.25）、動態因素模型（如 P.30），而英文或簡稱建議於第一次出現時註明即可。 10. 格式： (1) P.7 建議修正第三段【粉絲頁面】之標點符號。 (2) P.33 註 5，逕引「陳旭昇（2014）」即可。 (3) 文中部分引文格式如「Bernanke, Boivin, and Eliaz (2002)」，應改成：「Bernanke, Boivin, and Eliaz (2002)」。 (4) 文中括號之全半形之使用必須統一，例如：P.25 全半形括號混用，P.31 都是使用全形括號，P.35 卻是使用半形括號。 11. 參考文獻有漏列且表達方式請	8.參照評審意見，附件五資訊已進行調整刪除。 9. (1) 該處 Big Data 為參照文獻內容摘錄，已加註大數據之說明，請見 P.3。 (2) 謝謝指正，已進行刪除。 (3) 謝謝指正，已將其統一使用 the Conference Board，請見 P.11。 (4) 謝謝您的建議。 10. (1) 已進行調整。 (2) 已進行調整。 (3) 謝謝您的建議，已進行調整。 (4) 謝謝您的建議。 11.謝謝您的建議，請見 P.92。

評審意見	研究團隊回應
統一。如，P.117 中線上購物四字的字體應改用標楷體。 12. 在圖目錄與表目錄中，似有缺漏部分圖表，如圖 1-1 與表 5-11 並未出現該目錄中，建議作者重新檢視。此外，圖目錄中出現多次不同電商網的聲量圖、情緒指標圖與金鬚圖等，每家電商約有 7 種圖，惟分開標示將造成圖目錄異常雜亂，建議作者在命名時，或可考慮改成針對個別電商彙整表示，如「博客來—7 大分析圖」，並於文中說明 7 大分析圖的組成為何，如此整體排版上或可更精簡易讀。	12. 謝謝您的建議，惟各圖示為求解讀觀察，故有一定的大小規模，若用組合圖示命名恐將不利閱讀，本文在第五章第四節說明前，已有強調各電商網會搭配之七種圖形，應可協助閱讀者判斷。
捌、金檢處 一、提供參考建議事項 電子通訊網路正在翻轉現實世界，改變消費行為與金融服務型態，挑戰傳統金融交易與支付應用以及監管制度的有效性（Kroszner, 2015），網路大數據應用是未來的趨勢，無論如何央行都必須面對，這是挑戰也是機會，很高興研究處對大數據的重視，也感謝研究團隊的任勞任怨。 本報告透過主成分分析，將總體經濟領先指標及大數據網路檢索指標各淬鍊出兩個主成分，以 ARIMA 與 VAR 時間序列模型及 DFM 動態因素模型來從事民間消費預測的提升，研究發現經考量大數據資訊的模型相較自迴歸模型有更好的配適度，雖然 DFM 民間消費綜合指標的模型配適度較低，但主成分模組在 VAR 模型的預測上已能捕捉民間消費的變動趨勢。茲將報告中可再思考的一些	一、謝謝評審的建議

評審意見	研究團隊回應
想法列出，以供作者及主辦科參考： 1. 在部分均衡分析下，傳統民間消費預測是利用國民所得帳的結構化資料，以學理上的消費函數方程式模型來表現，先推估解釋變數（當期實質所得，實質利率等）的未來值後，再透過模型求解出預測值；本文直接運用總體經濟領先指標的主成分來推估，如果時間允許的話，可否將傳統方式及添加大數據網路檢索指標的樣本內（或樣本外）估計結果一併呈現，以幫助閱讀者的瞭解。 2. 台灣人可能會上淘寶網買一些便宜的大陸產品，不會在淘寶網買台灣的消費品，現行統計的民間消費有可能並不含這些淘寶網資料，大數據的淘寶網分析不知是否有其必要？其次，民間消費統計可能遺漏的 Uber 資料不易取得，可否稍加文字說明。 3. 主成分 VAR 分析是本文的重點，可否將衝擊反應函數圖及變異數分解表一併列出，以供參考。 二、文字之修正建議 1. 部分方程式未照順序排列，例如 P34 方程式 (14)、(15) 接續為方程式 (19)、(20)。 2. P44 第 4 行「相關內容分析請	1. 本研究已納入 ARIMA、ARIMAX- 傳統模型及 ARIMAX-GT 模型，並透過 RMSE、MAE、MAPE、以及 DM Test 進行納入網路大數據對民間消費預測之模型比較，詳請參考第五章之研究結果。 2. (1) 本研究將淘寶網資訊建立網路檢索指標之考量，主要參考 EagleEye 所公佈之 2014 年第 4 季「台灣購物網站排行榜」中前 50 大電商網中所公布之資訊，台灣民眾於淘寶網消費具有一定的市場比例。(2) 對民間消費進行預測，各式與消費有關之資訊，皆可能有其預測效果，如 Uber 付費資料,或 ATM 領取資訊，惟部分資料難以取得（如 Uber 付費資料），而本研究目前以網路檢索指標為目標。 3. 已納入衝擊反應函數圖及變異分解表，謝謝建議，請見附件七及附件八。 1. 感謝指正，已進行修改。 2. 謝謝指正，附件六資訊已進行調

評審意見	研究團隊回應
建附件六」，修正為「相關內容分析請見附件六」。	整，並於正文在文字上做改正。
玖、 經研處(2017/1/9 報告模型) 一、全文架構 本文著要目的是使用總體數據、搜尋引擎大數據及關鍵社群網站大數據等，利用主成份拆解成若干主成份，再透過 ARIMA 模型及向量自我迴歸(VAR)預測民間消費差分值。 二、報告內容 (一) 模型假設及實證分析 1. 頁 2，表 5-1 呈列民間消費預測總指標（靜態萃取）及民間消費預測總指標（動態萃取），惟後續分析並未使用此兩變數，應說明未使用之原因或理由。 2. 頁 3，提及研究模型均以 $I(0)$ 及 $I(1)$ 進行模型設定，惟後續僅用 $I(0)$ 變數進行討論，前述論述應有誤。同頁，作者使用有截距、有截距與趨勢及無截距趨勢進行單根檢定，其中，台灣民間消費連鎖實質值在有截距、有截距與趨勢單根檢定中呈現 $I(2)$，理論上應差分兩次。因此，建議增加變數差分後之單根檢定，以確定序列階次判斷的準確性。 3. 頁 6，說明民間消費差分一次後已為定態變數，因此設定 $d=0$，但下一段卻說明經挑選	1. 謝謝評審意見，已於第五章第三節增補相關分析。 2. 謝謝評審意見，已將報告內容作進一步說明。本文依照 Dolado et al.(1990)的單根檢定步驟來決定是否放入截距及趨勢項：先放入時間趨勢項及截距項檢定，若無法拒絕虛無假設，則考慮去除時間趨勢項，單純考慮截距項來進行檢定，若仍無法拒絕，則考慮不含截距項及時間趨勢項的 ADF 進行檢定。由於針對民間消費進行有截距、有截距與趨勢兩組合的單根檢定時，截距、截距與趨勢項在 $\alpha=0.1$ 水準下係數檢定均為不顯著，因此本研究採用台灣實質民間消費(取對數)一階差分進入模型。 3. 謝謝評審意見，已將報告內容重

評審意見	研究團隊回應
後 ARIMA(p,d,q) 階次為 ARIMA(3,1,3)，即 $d=1$，相互矛盾 4. 頁 11，說明本文以 2014Q1 至 2015Q1 為估計區間，惟 2014Q1 至 2015Q1 僅 4 個樣本點不可能進行估計，估計區間說明是否為誤植？頁 20 的 VAR 預測也有相同的問題。 5. 頁 13，建議可將大數據 ARIMAX-GT 模型及 ARIMAX-傳統模型，再與 ARIMA 模型進行 Diebold and Mariano(1995)檢驗，比較不同模型的樣本外預測能力，以了解總體變數及大數據對預測民間消費的幫助 6. 頁 15，VAR_TRADE 民間消費衝擊反應函數，民間消費衝擊 GDP 衝擊對民間消費毫無影響，似不合理，有可能是資料誤植或圖形有誤。頁 18 的 VAR_GTF 衝擊反應圖形也有相同的問題。 7. 附件二，VAR 模型因果分析並在內文未加以說明，建議刪除或是移至內文說明。 (二) 其他 1. 頁 1，第三段說明調整聲量搜尋大數據資料成月資料，為內文使用的是季資料。同頁，第三段第一行 Vosen and Schmidt(2011) 與 Vosen and Schmidt(2009)之間少頓號。 2. 頁 15，第二段「總體經濟領先因子對民間變異的解釋程度來	新修正，請見 P51。 4. 頁 11 所呈現的兩張圖，為樣本外預測的比較圖形，由於樣本外期間設定 2015Q2-2016Q1，因此實證分析軟體預設僅往前呈現 4 季，以突顯樣本外預測之比較。 5. 謝謝評審意見，本研究已增補至報告中。 6. 進行調整，已進行調整，並在附件七及附件八加註說明。 7. 已進行調整，並在附件七及附件八加註說明。 1. 謝謝指正，已在文字上修正。 2. 謝謝指正，已在文字上修正。

評審意見	研究團隊回應
得高」，遺漏了消費兩字，應改為民間消費變異。 3. 頁 16「表油表 5-14」應改為「由表 5-14」。	3.謝謝指正，已在文字上修正。
壹拾、 計量分析科(2017/1/9 報告模型) 一、變數處理與相關分析 1. 因子萃取的標準作法，所有變數應為 $I(0)$ 序列，進行標準化後再萃取因子，故因子應為 $I(0)$ 序列。請團隊說明為何在表 5-2 之民間消費預測總指標為 $I(1)$ 序列。 2. P.1 表 5-1 列舉的變數資料頻率不一(有月、有季)，建議說明如何將月資料轉換成季資料。 3. P.2、P.3 建議列出本文用到的變數即可。例如，本文並無直接將 CCI、利率等變數用於預測，而是將其用於主成分分析，故似無必要列出其單根檢定結果。 4. P.3「本研究模型均以 $I(0)$ 及 $I(1)$ 進行模型設定、模型估計後檢視殘差均為定態。」此句不易理解，是否係指本文係依據 ARIMA 的階次決定是否要對變數作差分。 5. 本文為說明表 5-2 的意義，建議可採用 PP 或 KPSS 檢定。另列出差分後之單給檢定結果。 6. 作者表 5-1 與表 5-2 中，有民間消費預測總指標 1、2、3、dfm，惟後面表 5-4 相關係數分析與 P.5 之後皆沒有使用這幾個變數，請作者說明理由。	1.謝謝評審建議，在單根檢定表格下，本研究亦有列出變數處理準則。(請見表 5-11 及表 5-25)。 2.轉換資料的方式，是採取 EViews 中的轉頻方法，指數的數值採用平均法，量的數值採用加總法。 3.謝謝評審建議，已進行修改，請見 P49。 4.謝謝評審提問，本研究建構的 ARIMA 模型階次以 AIC 最小值作為挑選模型原則，經挑選最適 ARIMA(p,d,q) 階次為 ARIMA(3,1,3)。 5.謝謝評審建議，已進行修改，請見 P49。 6.本研究將其陳述在第五章第三節。

評審意見	研究團隊回應
7. 表 5-4 相關係數分析中，為何 DGT_FAC_1 與 DGT_FAC_2 對其他變數相關係數皆相同？且兩者相關係數為 1？請確認。 二、ARIMA 模型 1. P.6 說明民間消費取一次差分後為定態序列，惟觀察表 5-2 發現，民間消費連鎖實質值可能為 I(2)及 I(1)，請團隊說明為何 ADF 檢定選擇無截距與趨勢之設定。 2. CPIYOY 放入模型作為解釋變數的意義何在？實質民間消費等變數是否未取對數，而直接取差分？ 3. 若 ARIMA(3,1,1)模型為基準模型，請團隊在進行 DM 檢定時，應予基準模型進行比較，而非以兩個擴充模型(傳統模型與大數據模型)進行討論。 4. 圖 5-1 與 5-2，以 DPC 變數來看，PC 已取差分，為何左軸單位這麼大，請確認。 5. 本文應係以實質民間消費的一階差分進行迴歸分析(由表 5-6 本文的估計結果可發現，截距	7.謝謝評審勘誤，已將該部分修正。 1.謝謝評審意見，已將報告內容作進一步說明。本文依照 Dolado et al.(1990)的單根檢定步驟來決定是否放入截距及趨勢項：先放入時間趨勢項及截距項檢定，若無法拒絕虛無假設，則考慮去除時間趨勢項，單純考慮截距項來進行檢定，若仍無法拒絕，則考慮不含截距項及時間趨勢項的 ADF 進行檢定。由於針對民間消費進行有截距、有截距與趨勢兩組合的單根檢定時，截距、截距與趨勢項在 $\alpha=0.1$ 水準下係數檢定均為不顯著，因此本研究採用台灣實質民間消費(取對數)一階差分進入模型。 2.謝謝評審意見，已將模型修正。 3.謝謝評審意見，本研究已增補至報告中。 4.謝謝評審意見，已將模型及圖表修正，請見 P68。 5.謝謝評審意見，請見 P53。

評審意見	研究團隊回應
項數值很大，推測可能係以實質民間消費的一階差分(未取log)為被解釋變數)。惟對央行與政策執行者而言，更關心民間消費的成長率(或年增率)，建議本文可將模型的預測結果轉換成民間消費年增率，計算民間消費年增率預測值的預測誤差。 6. 表 5-6，ARIMA 分析結果中之 ARIMAZ GT-Model，DGT_FAC_1 與 DGT_FAC_2 已取差分，為何其估計係數卻非常大，請確認結果。 7. P.6 第一段指出，民間消費經 1 階差分後為定態序列，因此設定 $d=0$，為何不是設定 $d=1$？ARIMA 格式既為 ARIMA(p,d,q)，為何表 5-5 的 model 形式為(x,x)(0,0)，請問(0,0)的意思為何？ 8. P.8 ARIMA 的傳統模型及 GT-MODEL 的解釋變數加入 GDP，此在實務上進行預測時會有困擾，即 T 期 GDP 數值公布時，T 期民間消費也實現，故不須對 T 期的民間消費預測，而若要對 T+1 期的民間消費也要對 T+1 期的 GDP 進行預測，增加預測上的複雜性，此問題於期中報告即已提出。 9. P.8 請說明樣本內「靜態預測」與「動態預測」差異何在，是否僅差在帶入不同之 AR 值，而其他解釋變數仍以實際值代	6.謝謝評審意見，已將模型及圖表修正，請見 P53。 7.謝謝評審意見，已將報告內容重新修正，請見 P51。 8.謝謝評審意見，請見 P53。 9.謝謝評審意見，請見 P55。

評審意見	研究團隊回應
入。不同模型的樣本內估計結果，建議以靜態的「配適值」評估即可。 10. P.9 觀察圖 5-1 及圖 5-2，模型殘差仍可能有序列相關，建議進一步對殘差進行檢定。若殘差有序列相關，有可能係因本文並未處理季節性因素所致，建議加以處理。另亦可嘗試將民間消費改為對上年同期差分。 11. P.11 本文說明調整後判定係數，惟在本文並無相對應之探討內容，請團隊加以補充。 12. 樣本外預測部分： (1) P.11 為何樣本期間自 2004 年起，卻僅以 2014Q1 至 2015Q1 為估計區間，對 2015Q2 及 2016Q1 進行預測？是否為誤植？(VAR 分析的部分亦同。) (2) 樣本外預測結果之動態預測績效比較部分，大部分 ARIMA(3,1,1)較佳(表 5-9)，似乎顯示以大數據建構之民間消費預測模型預測能力不佳，請作者針對此點補充說明。 (3) 請作者進行 ARIMA(3,1,1)、ARIMAX(2,1,0)傳統模型與 ARIMA GT-model(3,1,0)三個模型之兩兩模型互相比較之 DM test，而非僅針對 ARIMAX(2,1,0)傳統模型與 ARIMA GT-model(3,1,0)進行 DM test。 (4) P.13 作者說明「納入網路大數據...績效表現不俗」，惟從樣本	10.謝謝評審意見，請見附件七及附件八。 11.謝謝評審意見，請見 P.47。 12. (1)謝謝評審意見，樣本外預測軟體操作及期間設定，均參照陳旭昇(2009)、楊奕農(2007)《時間序列分析》進行。 (2)謝謝評審意見，已於第六章結論進行相關可能說明，請見 P.86。 (3)謝謝評審意見，本研究已增補至報告中。 (4) 謝謝評審意見，已調整做修正。

評審意見	研究團隊回應
外預測結果來看，此結論似乎不合宜，請作者重新檢視模型、樣本設定與結果等細節再做結論。 三、VAR 模型 1. VAR 模型之變數排序有其意義存在，對衝擊反應的結果亦有影響，若在時間有限下，團隊可參考一般化的衝擊反應 (generalized impulse response) 分析，避免模型錯誤設定問題。 2. 請補充對圖 5-5 與圖 5-6 之結果說明。此外，圖 5-5 與圖 5-6 中下方之衝擊反應函數為何為 0？或許係因變數大小差異過大，無法觀察部分變數之反應，建議團隊可以標準化各變數，以利讀者了解各衝擊反應關係。 3. P.18 指出，網路檢索指標 1 對民間消費影響呈平穩負向，但從圖來看，對民間消費影響似乎落在零軸附近為顯著異於 0。 4. 樣本外預測部分： (1) 樣本外預測結果，大部分以 VAR_TRADF 較佳(表 5-18)，作者可否能須針對此點補充說明。 (2) 請補充有關 DM test 結果。 5. P.23 附件二列出 VAR 模型因果分析的資訊，建議可說明表中各估計式的模型設定為何？以利讀者了解，如 Equation 為 dpc 時是否係指模型的被解釋	1. 謝謝評審意見。 2. 謝謝評審意見，已調整做修正。 3. 謝謝評審意見，已調整做修正。 4. (1) 謝謝評審意見，已進行模型修改，請見第五章第二節及第三節之說明。 (2) 謝謝評審意見，已進行修改，請見 P62。 5. 謝謝評審意見，已調整做修正。

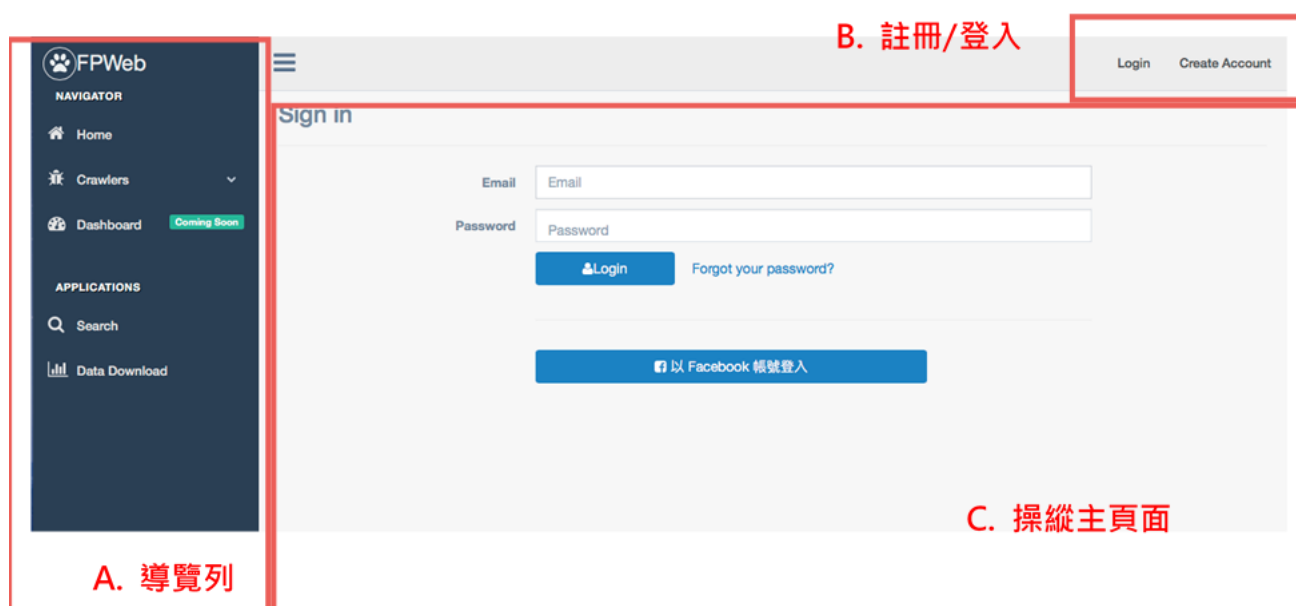
評審意見	研究團隊回應
變數為 dpc？而 Excluded 為 drgdp 時，是表示模型刪除了變數 drgdp，但其他變數仍留在迴歸式的等式右方？ 6. 表 5-16 之 LM 檢定，由於本文僅用 VAR(3)進行討論，似不用呈列 VAR(1)及 VAR(2)之結果。 7. P.19，因果分析應為附件 2，非附件 1，請團隊重新檢視並增加相關研究發現之討論。 四、文字修改及其他意見 1. P.1，第二段消費者信心指數 CCI 應為「消費者信心指數(CCI)」；第三段則應修正為 Vosen and Schmidt(2011)、Vosen and Schmidt(2009)、Matsumoto, Matsumura and Shriaki(2013)。 2. 表 5-1，資料來源部分，主計處應為「主計總處」。 3. 請團隊統一各模型之英文簡稱，以利讀者閱讀。 4. 請團隊說明如何取差分，是 YoY 或是 QoQ。 5. 表 5-14，請團隊重新檢視，勿將負號及數字分隔兩行。 6. 請團隊說明為何 VAR 模型與 ARIMA 模型最大落後期設定不同之原因。 7. P.22 附表右下欄所列之資訊(如 R^2 等)應有誤，宜確認本文中資料之正確性。 8. 不論是從 VAR 衝擊反應、預測誤差拆解、ARIMA 樣本外預測結果，加入網路檢索的結	6.謝謝評審意見，已調整做修正。 7.謝謝指正，已調整做修正。 1.謝謝指正，已進行修改。 2.謝謝指正，已進行修改，請見 P.42。 3.謝謝指正，已進行修改。請見第五章第二節及第三節說明。 4.差分為以 QoQ 處理。 5.謝謝指正，已進行修改。 6.在進行落後期篩選時，本研究考量有二，第一，樣本數過短問題，第二，以 AIC 指標做模型最適篩選。兩者模型設定不同，但皆是以找出最適估計模型為目標。 7. 謝謝指正，已進行修改。 8. 謝謝指正，已進行修改，請見第五章第二節及第三節小結說明。

評審意見	研究團隊回應
果，並沒有明顯比沒有加入的好，甚至有時較差。而且也難以在實務上應用(因為解釋變數有 GDP)。建議作者應加入實證結果總結，究竟加入網路檢索的結果，會不會帶來預測上的好處？	

附件四 社群網路資料擷取平台

為蒐集關鍵社群網站民眾行為資料，做關鍵社群網站大數據資訊蒐集工作，本研究已開發資料擷取平台，相關資訊如下說明：本網站平台之使用流程可分為兩部份：「爬蟲資料收集」及「搜尋彙整下載」。爬蟲資料收集部份，可分別設定 PTT 及 FB 兩隻爬蟲搜集來源（如：臉書中某些粉專頁面），為能針對電商網，對民眾消費行為進行觀察，本研究目前主要以臉書為觀察重點，因為臉書為依據各電商的粉絲團經營，並有固定項目商品資訊可供觀察，相較 PTT 則屬於民眾自發性的討論樣態。

而等待爬蟲搜集足夠資料量後，則進入「搜尋彙整下載」部份，此部分依使用者所需之關鍵字、日期相關資訊彙整，而從前一階段之所爬資料產生簡報供後續使用。



附圖 4-1 社群網站資料擷取平台主畫面示意圖

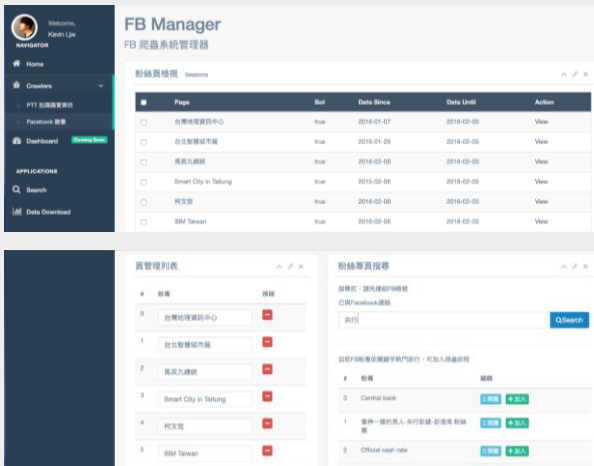
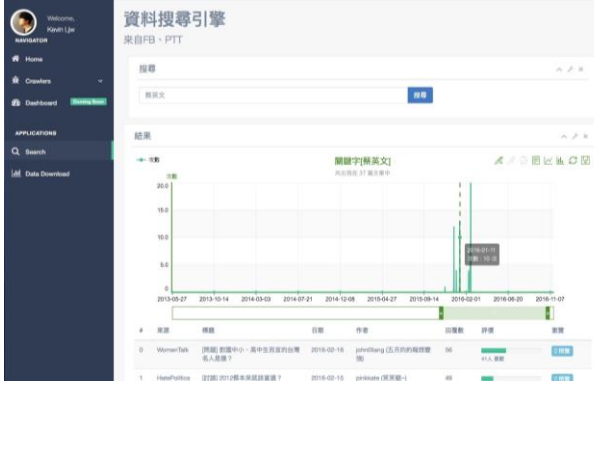
A. 導覽列：瀏覽網站平台並前往各項功能之導覽列（所有功能齊聚於導覽列內，使用者可以按上至下順序依序設定完畢，並產出所需之文件），從上至下依序為：Home（首頁，查看目前平台收集資料統計量），Crawler（爬蟲爬文設定，子項目分 PTT 版面及臉書粉專設定），Dashboard（統計數據視覺化儀表板，提供使用者鎖定之粉專、版面之數據統計視

覺化圖表)，Search（網站內搜尋引擎，提供所收集資料之關鍵字即時收集，並可以查看貼文），Data Download（簡報資料統整下載，設定好關鍵字、日期區間、資料來源後，便可以輸出下載xlsx 檔案）

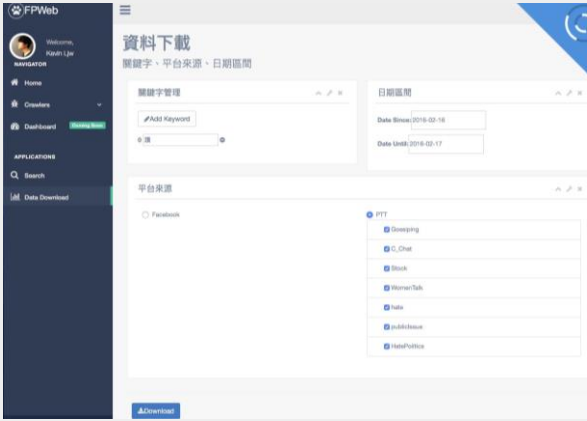
B. 註冊/登入：本網站平台之設計具有個人化帳號，除了可以使個人擁有其粉專、關鍵字相關設定外，亦可保全網站平台內使用者之隱私性。使用者僅需在第一次使用時填入個人郵件、帳密相關資訊，而在下次使用時以個人帳戶登入即可使用。

C. 操縱主頁面：其內容根據導覽列選擇不同而產生，使用者主要在這個框架內操縱。接下來的內容，便會逐一介紹根據不同導覽列之點選項目而相應出現之操縱主頁面。說明表請見附表 4-1。

附表 4-1 操縱主頁使用說明表

	臉書爬蟲管理介面 1. 在導覽列中，點選「Crawler」選擇「Facebook」進入此管理頁面。 2. 此頁面亦分三欄位，與 Ptt 管理頁面相似，最上方為「粉專頁檢視」欄位，此欄位提供檢視此平台之爬蟲正收集之版面。而左下欄位則是「粉專管理列表」，提供版面刪除作業，而右下方欄位則為「粉專新增」欄位，使用者可以透過此欄位搜尋關鍵字，找出最相關之粉專（最相關粉專會列於頂端），並且新增至左側欄位中，在按下「update」爬蟲便會開始抓新增版面的資料。
	資料搜尋引擎 1.在導覽列中，點選「Search」進入此搜尋頁面。此頁面目前提供平台內所得之 Ptt 資料庫搜尋，能夠依據使用者輸入之關鍵字搜尋（搜尋規則如同 Google 之符號搜尋規則），而結果會以關鍵字關聯性排序，並以動態圖表依日期顯示其聲量。 2. 請參考左圖之範例，搜尋結果將會出現於搜尋輸入列下方。首先是依日期排序之聲量圖表，而接著是依據搜尋相關性排序之文章，可以從中觀察群眾輿論並且透

	過連結開啟此頁檢視。
--	------------

	資料下載 1.在導覽列中，點選「Data Download」進入此下載頁面。 2.本頁面提供 xlsx 資料格式之簡報下載，在使用者設定此次所要下載之「關鍵字」（不限多寡）、抓取「時間區間」以及資料來源（ptt 或 fb 可分別分次下載）後，按下「Download」按鈕。略待幾秒，就會彈跳出現下載檔案選擇介面。
--	---

附件五 網路大數據—搜尋指標篩選過程

壹、15 大電商網站中的分類項目「中類」做整理

將 15 大網購網站中的分類項目「中類」做整理，並依照國民所得消費分類家計物價與網購商品清單分類對照。

第一步驟：

整理出各類別中的分類項目，分為大類、中類、小類、品牌列表、注目商品。

第二步驟：

將中類的分類項目，放入 Google Trends 搜尋趨勢，搜尋出的關鍵字分別為：購物中心 74 個、拍賣 397 個、服務+119 個，數據資料為：2004 年 1 月 4 日至 2016 年 5 月 22 日，關鍵字每週被人們搜尋的頻率。

第三步驟：

將中文的關鍵字翻譯成英文，並依照各類別作編碼。

第四步驟：

將各類別搜尋出的關鍵字做綜合整理，整理出商品的唯一清單，共有 529 項，其中購物中心佔了 42 項，拍賣佔了 370 項，服務+佔了 117 項。

第五步驟：

將「中類」的關鍵字，依照國民所得消費分類家計物價，做分類對照。

					國民所得的解消費												家庭月支出								
單次編號	測碼	國碼	代碼	中檔	1.1食品 及非酒 精飲料	1.2酒精 類產品	1.3衣著 鞋履及 雜項用 品	1.4住宅 服務、 水電氣 熱及其 他燃料	1.5家具 器具及 家庭裝 潢	1.6醫療 保健費	1.7交通 通訊	1.8通信 與文化	1.9教育	1.11餐 館及旅 遊業	1.12其他各 對等項目 之開列(不 另繳納消 費稅)		食品飲 料及其 基	衣着裝 飾用品	住宅服 務、水 電氣及 其他 燃料	家具器 具及家 庭裝飾 材料	醫藥保 健	運輸交 通及郵 訊	休閒、 文化及 教育消 費	餐飲及 旅館	(什)其 餘消費
M01	Yaloo全部商品分類		1 Yali A001	送禮進配											1										
M02	Yaloo全部商品分類		2 Yali A002	電腦								1	1											1	
M03	Yaloo全部商品分類		3 Yali A003	廣播								1	1											1	
M04	Yaloo全部商品分類		4 Yali A004	軟體								1	1											1	
M05	Yaloo全部商品分類		5 Yali A005	家電					1													1			
M06	Yaloo全部商品分類		6 Yali A006	視聽																	1				
M07	Yaloo全部商品分類		7 Yali A007	電話								1												1	
M08	Yaloo全部商品分類		8 Yali A008	相機								1												1	
M09	Yaloo全部商品分類		9 Yali A009	音響機								1											1	1	
M10	Yaloo全部商品分類		10 Yali A010	媒體								1												1	
M11	Yaloo全部商品分類		11 Yali A011	汽車							1												1		
M12	Yaloo全部商品分類		12 Yali A012	機車																				1	
M13	Yaloo全部商品分類		13 Yali A013	GPS							1													1	
M14	Yaloo全部商品分類		14 Yali A014	服裝			1																		
M15	Yaloo全部商品分類		15 Yali A015	小褲內衣																					
M16	Yaloo全部商品分類		16 Yali A016	鞋子			1																		
M17	Yaloo全部商品分類		17 Yali A017	包包			1																		
M18	Yaloo全部商品分類		18 Yali A018	精品																					
M19	Yaloo全部商品分類		19 Yali A019	手錶																					
M20	Yaloo全部商品分類		20 Yali A020	珠寶																					
M21	Yaloo全部商品分類		21 Yali A021	配飾																					
M22	Yaloo全部商品分類		22 Yali A022	表飾																					
M23	Yaloo全部商品分類		23 Yali A023	鐘錶							1														
M24	Yaloo全部商品分類		24 Yali A024	洗沐沐浴						1												1			
M25	Yaloo全部商品分類		25 Yali A025	家具					1																
M26	Yaloo全部商品分類		26 Yali A026	傢具					1																
M27	Yaloo全部商品分類		27 Yali A027	廚具					1																
M28	Yaloo全部商品分類		28 Yali A028	衛浴					1																
M29	Yaloo全部商品分類		29 Yali A029	清潔																					
M30	Yaloo全部商品分類		30 Yali A030	食品	1																				
M31	Yaloo全部商品分類		31 Yali A031	飲料	1																				
M32	Yaloo全部商品分類		32 Yali A032	咖啡																					
M33	Yaloo全部商品分類		33 Yali A033	鹹甜											1										1
M34	Yaloo全部商品分類		34 Yali A034	玩具								1	1											1	
M35	Yaloo全部商品分類		35 Yali A035	童裝																					
M36	Yaloo全部商品分類		36 Yali A036	童鞋			1																		
M37	Yaloo全部商品分類		37 Yali A037	運動								1													
M38	Yaloo全部商品分類		38 Yali A038	器材								1													
M39	Yaloo全部商品分類		39 Yali A039	票券							1				1										
M40	Yaloo全部商品分類		40 Yali A040	書籍																					
M41	Yaloo全部商品分類		41 Yali A041	文具									1												

附圖 5-1 國民所得消費分類家計物價與網購商品清單分類對照-1

附圖 5-2 國民所得消費分類家計物價與網購商品清單分類對照-2

附圖 5-3 國民所得消費分類家計物價與網購商品清單分類對照-3

附圖 5-4 國民所得消費分類家計物價與網購商品清單分類對照-4

貳、將 Google Trends 資料轉換頻率

➤ 第一步驟：

將整理好的 Google Trends 數據資料，移除中文、英文、時間數列，僅留下代碼、數據資料，儲存成 EViews 專屬使用的 Excel 檔，檔名、Excel 中的 Sheet 不可有中文字、使用『_』符號來做文字間隔，否則 EViews 無法順利讀取檔案。使用代碼轉換頻率，是為了保有數值在 EViews 中的順序，如果以英文文字輸入 EViews 後，EViews 會按照 A~Z 的順序排列，轉換完頻率後，需再將英文各別對照中文文字，才能進到 SPSS 作後續分析

➤ 第二步驟：

在 EViews 中開啟新的工作表，資料頻率為：Weekly、開始時間：2004 01 04、結束時間：2016 05 22，輸入 EViews 專屬使用的 Excel 檔，選取「下一步」，選取「完成」，再輸入開始時間：2004 01 04，即可完成資料的輸入

➤ 第三步驟：

在 EViews 中開啟新的工作表，資料頻率為：Weekly、開始時間：2004 01 04、結束時間：2016 05 22，輸入 EViews 專屬使用的 Excel 檔，選取「下一步」，選取「完成」，再輸入開始時間：2004 01 04，即可完成資料的輸入

➤ 第四步驟：

在 EViews 中點選「New Page」，可以開啟一個新的工作頁面，頻率分別選擇「Monthly」、「Quarterly」，建立好新的頁面後，回到原始輸入的 Weekly 資料，將所有代碼選取並複製，貼至新的頁面 Monthly、Quarterly，即可將週資料轉成月資料、季資料

Group: GROUP01_WEEKLY

Workfile: UNTITLED

ViewProcObject

PrintNameFreezeDefault

SortEdit+/Smpl+/Compare+/

YAH1_X001

YAH1_X002

YAH1_X003

YAH1_X004

YAH1_X005

1/04/2004	0	81	67	95
1/11/2004	0	69	63	84
1/18/2004	0	46	37	90
1/25/2004	0	76	72	78
2/01/2004	0	87	57	90
2/08/2004	0	78	44	85
2/15/2004	0	82	46	94
2/22/2004	0	78	40	93
2/28/2004	0	75	43	86
3/07/2004	0	79	96	91
3/14/2004	0	73	56	81
3/21/2004	0	75	64	84
3/28/2004	0	100	33	94
4/04/2004	0	84	30	100
4/11/2004	0	73	58	86
4/18/2004	0	72	79	84
4/25/2004	0	76	69	91
5/02/2004	0	79	81	88
5/09/2004	0	68	62	89
5/16/2004	0	75	64	94
5/23/2004	0	74	52	90
5/30/2004	0	88	60	92
6/06/2004	0	78	42	81
6/13/2004	0	75	58	81
6/20/2004	0			

Group: GROUP01_MONTHLY

Workfile: UNTITLED

ViewProcObject

PrintNameFreezeDefault

SortEdit+/Smpl+/Compare+/

YAH1_X001

YAH1_X002

YAH1_X003

YAH1_X004

YAH1_X005

2004M01	0	68	59.75	79.25
2004M02	0	80	46	90
2004M03	0	81.75	59.75	87.5
2004M04	0	76.25	59	90.75
2004M05	0	76.8	63.8	90.6
2004M06	0	73.5	45.25	90.25
2004M07	0	82.5	66.5	83
2004M08	0	74	44.2	80.4
2004M09	0	70.5	48.75	81.5
2004M10	0	64.4	37.4	79.4
2004M11	0	60.75	43	77
2004M12	0	70.5	55.25	82.25
2005M01	0	52.8	40	69.6
2005M02	0	51.25	35.25	63.75
2005M03	0	59.25	67.5	72.75
2005M04	0	52.75	48.25	65
2005M05	0	54.4	36.8	66.6
2005M06	0	51	36	57
2005M07	0	53.4	31	56.6
2005M08	0	59.25	30.25	55
2005M09	0	44.75	23	48.75
2005M10	0	41.6	34.4	48.2
2005M11	0	38.75	33	45
2005M12	0	39.5	27.75	45.25
2006M01	0	31.8	21	36.8
2006M02	0			

Group: GROUP01_QUARTERLY

Workfile: UNTITLED

ViewProcObject

PrintNameFreezeDefault

SortEdit+/Smpl+/Compare+/

YAH1_X001

YAH1_X002

YAH1_X003

YAH1_X004

YAH1_X005

2004Q1	0	76.8461538	54.4615384	85.9230769
2004Q2	0	75.6153846	56.6153846	87.4615384
2004Q3	0	75.5384615	52.4615384	81.5384615
2004Q4	0	65.1538461	44.6153846	79.5384615
2005Q1	0	54.3076923	43.9230769	68.7692307
2005Q2	0	52.8461538	40.0769230	62.7692307
2005Q3	0	46.7692307	28.3076923	53.6923076
2005Q4	0	40.0769230	31.9230769	46.3076923
2006Q1	0	34.4615384	25.5384615	40.5384615
2006Q2	0	34.7692307	23.6923076	40.4615384
2006Q3	0	34.4615384	21.0769230	35.6923076
2006Q4	0	29	22.3571428	34.4285714
2007Q1	0	26.1666666	20.3333333	28.3333333
2007Q2	0	26.6923076	20.2307692	34.5384615
2007Q3	0	31.7142857	18.7857142	33.5
2007Q4	0	27.9230769	21.5384615	33.9230769
2008Q1	0	24.3846153	18.5384615	29.1538461
2008Q2	0	25.3076923	18.2307692	34.5384615
2008Q3	0	27.3076923	18.9230769	30.2307692
2008Q4	0	23.3846153	17.8461538	28.8461538
2009Q1	0	21.5384615	14.0769230	26.8461538
2009Q2	0	23.9230769	16.3130769	26.2307692
2009Q3	0	24.7692307	17.3076923	27.6923076
2009Q4	0	21.8461538	16.6923076	27.6923076
2010Q1	0	22.0769230	16.3076923	26.2307692
2010Q2	0			

附圖 5-5Yahoo 購物中心，以代碼方式轉換頻率

➤ 第五步驟：

將季資料的數值，全部選取且複製，貼到新的 Excel 檔，將整理好的 Google Trends 數據資料開啟，利用代碼對應數值的方式，將中文文字貼上，另存成 SPSS 專屬使用的 Excel 檔

EViews 轉換頻率 季頻率資料-代碼							Excel 原始檔-Google Trends 數據資料整理 對應EViews代碼，將中文文字貼上							Excel 新檔案 數據資料：季頻率，中文文字，進入SPSS作分析						
View	Proc	Object	Print	Name	Freeze	Default	Sort	Edit	Y	Simple	Compare	Y	Y							
Group: GROUP01_QUARTERLY Workfile: UN...							8H急運配 電腦 週邊 軟體 家電													
YAH1_X001 YAH1_X002 YAH1_X003 YAH1_X004 YAH1_X005							haste with 8H computer periphery software appliances							8H急運配 電腦 週邊 軟體 家電						
YAH1_X001 YAH1_X002 YAH1_X003 YAH1_X004 YAH1_X005							Yah1 x001 Yah1 x002 Yah1 x003 Yah1 x004 Yah1 x005													
2004Q1		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-01-04 - 2004-01-10		0	81	67	95	74	2004 Q1	0	75.6153846	56.6153846	85.9230769	64.9230769	
2004Q2		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-01-11 - 2004-01-17		0	69	63	84	72	2004 Q2	0	75.6153846	56.6153846	85.9230769	64.9230769	
2004Q3		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-01-18 - 2004-01-24		0	46	37	60	68	2004 Q3	0	75.6153846	56.6153846	85.9230769	64.9230769	
2004Q4		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-01-25 - 2004-01-31		0	76	72	78	70	2004 Q4	0	75.6153846	56.6153846	85.9230769	64.9230769	
2005Q1		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-02-01 - 2004-02-07		0	87	57	90	60	2005 Q1	0	75.6153846	56.6153846	85.9230769	64.9230769	
2005Q2		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-02-08 - 2004-02-14		0	78	44	65	68	2005 Q2	0	75.6153846	56.6153846	85.9230769	64.9230769	
2005Q3		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-02-15 - 2004-02-21		0	82	46	94	63	2005 Q3	0	75.6153846	56.6153846	85.9230769	64.9230769	
2005Q4		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-02-22 - 2004-02-28		0	78	40	93	64	2005 Q4	0	75.6153846	56.6153846	85.9230769	64.9230769	
2006Q1		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-02-29 - 2004-03-06		0	75	43	88	62	2006 Q1	0	75.6153846	56.6153846	85.9230769	64.9230769	
2006Q2		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-03-07 - 2004-03-13		0	79	96	91	58	2006 Q2	0	75.6153846	56.6153846	85.9230769	64.9230769	
2006Q3		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-03-14 - 2004-03-20		0	73	56	81	68	2006 Q3	0	75.6153846	56.6153846	85.9230769	64.9230769	
2006Q4		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-03-21 - 2004-03-27		0	75	54	84	54	2006 Q4	0	75.6153846	56.6153846	85.9230769	64.9230769	
2007Q1		0	75.6153846	56.6153846	85.9230769	64.9230769	2004-03-28 - 2004-04-03		0	100	33	94	63	2007 Q1	0	75.6153846	56.6153846	85.9230769	64.9230769	

附圖 5-6 Yahoo 購物中心，EViews 代碼的季數據，對應中文文字，以便後續作 SPSS 分析

附件六 15 大電商網主成份分析成果

壹、十五電商網第一主成份資訊

有關電商網萃取步驟，可分為表 5-1 四步驟。

附表 6-1 電商主成份萃取操作步驟

步驟	操作內容
Step1	15 大購物網站，原始的中類關鍵字共有 1838 個，考慮到各網站使用的關鍵字重複，將重複的詞彙挑選出來並去除，整理出唯一清單的中類關鍵字：1148 個。
Step2	將各購物網站的中類關鍵字，做第一次的主成份分析，相關內容請見如下陳列。
Step3	將第一次主成份分析無法顯示出成份因子的三的購物網站：Costco、露天、Yahoo 拍賣，將中類關鍵字放回大類做分類，用大類做出主成份分析，知道大類的平方差累積變異量後，再對應到中類的關鍵字。
Step4	在主成份分析中，僅採用 15 大購物網站第一大主成份，去進行後續的網路檢索指標萃取及民間消費綜合指標淬鍊。

而在經由主成份分析後，各電商網的因子萃取概況如表 5-2 所示。

附表 6-2 電商主成份萃取因子概況

15 大購物網站	商品分類數值			中類關鍵字 作第一次主 成份及主成 份分析個數		中類關鍵字 作第二次主成份 主成份分析個數
	大類	中類	小類			
Boo-博客來	-	29	290	4		-
Cos-Costco	15	118	279	9	未顯示成份因子	3
Etm-東森購物	14	198	137	18		-
Fam-FamiPort	9	57	65	7		-
Goh-GoHappy	15	135	493	13		-
Ibo-ibon	8	62	85	6		-
Mom-momo	-	19	98	3		-
Myf-myfone 購物	-	38	246	4		-
Pch-PChome 24h 購物	-	14	355	3		-
Pch-PChome 購物中心	-	8	141	(因 PChome 中類數目較少，兩個一起做分析)		-
Rak-樂天市場	12	76	436	6		-
Rut-露天拍賣	24	475		39	未顯示成份因子	10
Tao 淘寶網	-	26	378	3		-

15 大購物網站	商品分類數值			中類關鍵字 作第一次主 成份及主成 份分析個數		中類關鍵字 作第二次主成份 主成份分析個數
	大類	中類	小類			
Yah1-Yah 全部商 品分類	-	42	171	4		-
Yah2-Yah 品牌旗 艦分類	-	28	172	(因 Yahoo 「品牌旗艦」 的關鍵字已 經全部包含 在「全部商 品」中，只做 一次分析)		-
Yah b-Yahoo 拍賣	22	393	-	38	未顯示成 份因子	9
Yah s-Yahoo 服務	4	120	-	12		-

以下則分別呈現 15 大購物網站之因子成分，黃色標註為第一大主成份之項目，並包含其平方差累積變異量。

附表 6-3 Boo-博客來主成份解釋變異量資訊

商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量
CD	.948	日用	.816	餐廚	-.356
家居	.947	簡體書	.811	無印良品	-.525
保健	.946	外文書	.802	手錶	-.615
DVD	.929	美妝	-.771	清潔	-.678
STARBUCKS	-.922	MOOK	.721	3C	.546
包包	.898	icash	-.661	中文書	.249
文具	.898	家電	.649	美食	-.545
生活	.892	服飾	-.076	MUJI	-.498
設計	.835	休閒	.257	配飾	.321
雜誌	.819	婦幼	-.274		

註：已顏色標註之商品項目，為第一主成分組成資訊。

附表 6-4 Cos-Costco 主成份解釋變異量資訊

商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量
FAC_COS_ 書籍文具	.986	FAC_COS_ 電子產品 1	.882	FAC_COS_ 汽百五金 1	.388
FAC_COS_ 家俱收納	.974	FAC_COS_ 生活家電 1	.868	FAC_COS_ 生活家電 2	-.212
FAC_COS_ 婦幼玩具	.954	FAC_COS_ 禮品特展	.860	FAC_COS_ 服裝鞋包 1	-.500

商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量
FAC_COS_運動休閒 2	.951	FAC_COS_服裝鞋包 2	.762	FAC_COS_食品飲料 3	-.034
FAC_COS_食品飲料 2	.946	FAC_COS_食品飲料 1	-.163		
FAC_COS_電腦週邊 1	.920	FAC_COS_運動休閒 1	.152		
FAC_COS_居家生活 2	.918	FAC_COS_美體美髮 1	-.287		
FAC_COS_美體美髮 2	.915	FAC_COS_居家生活 1	-.297		
FAC_COS_健康保健	.915	FAC_COS_珠寶手錶 1	-.141		
FAC_COS_汽百五金 2	.883	FAC_COS_電子產品 2	-.405		

註：已顏色標註之商品項目，為第一主成分組成資訊。

附表 6-5 Cos-Costco 主成份項目資訊

類別成份矩陣	關鍵字									
FAC_COS_書籍文具	書籍	音樂	影片	文具用品	文具					
FAC_COS_傢俱收納	家具	辦公	辦公家具	戶外家具	兒童家具					
FAC_COS_婦幼玩具	兒童用品	兒童服飾	童書	玩具						
FAC_COS_運動休閒 2	花卉	戶外	園藝用品							
FAC_COS_食品飲料 2	飲料	茶	有機食品	可可亞	有機	穀物				
FAC_COS_電腦週邊 1	電腦	平板	印表機							
FAC_COS_居家生活 2	生活雜貨	家飾	耗品	清潔用品						
FAC_COS_美體美髮 2	化妝品	男性	香水	個人						
FAC_COS_健康保健	保健食品	保健	食品	保健用品	用品					
FAC_COS_汽百五金 2	家庭	五金工具								
FAC_COS_電子產品 1	相機	樂器	攝影機	3C						
FAC_COS_生活家電 1	配件	生活	居家	廚房	居家生活					
FAC_COS_禮品特展	節慶	商品	禮品							
FAC_COS_服裝鞋包 2	服飾	包包	鞋襪							

附表 6-6 Etm-東森購物主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
醫美	.989	男裝	.957	舒壓	.927	開架	.894	平輪	.850
推薦	.988	客廳	.950	鞋墊	.922	美髮	.885	最新	-.844
孕婦	.972	水產	.945	鑽石	.916	飲品	.881	香氛	.841
收納	.972	時尚	.943	眼鏡	.914	手錶	.880	衛生棉	.831
行李箱	.967	洗衣機	.942	茶葉	.912	品牌	-.880	美肌	.830
沙發	.966	床墊	.939	中大尺碼	.911	元氣	.879	尿布	.829
甜點	.962	床包	.935	機車	.908	款式	-.861	蛋糕	.827
專櫃	.959	穿搭	.934	藍寶	.905	玉石	-.857	行車紀錄器	.827
女裝	.959	童裝	.932	筆電	.904	導航	.856	墨鏡	.825
嚴選	.958	露營	.927	港澳	-.898	休閒鞋	.851	彩妝	.821
蔬菜	-.816	飾品	.759	器材	-.719	紅寶	.665	液晶電視	-.600
配件	-.792	流行	-.759	情趣	-.714	露營用品	.665	健身器材	-.597
影音	-.785	保健	-.758	保健用品	-.711	床組	.663	水果	-.595
運動	.784	內衣	.757	品味	-.706	美容家電	.659	APPLE	.583
健食	-.782	健身	-.756	明星	-.697	食材	.650	窈窕	-.574
小物	.778	單眼	.756	婦幼	.689	北台灣	.641	草本	-.527
美顏	.774	翡翠	-.748	餐廚	.678	咖啡	.638	零食	.395
茶具	-.772	精品	-.729	玩樂	.678	果乾	.626	硬碟	-.109
珍珠	-.765	鍋具	.728	冰箱	-.676	智慧型手機	.617	藝術	.014
萬用	.760	珠寶	.723	衣物	-.676	按摩舒壓	.602	肉類	.206
東森	-.126	微單	.078	數位相機	-.438	纖體	-.356	美材	-.514
桌上型電腦	.115	機能	-.330	冷氣空調	-.027	飲料	.290	女內褲	.255
醫療	.116	紙	.405	週邊	.442	金飾	.155	名筆	-.440
家飾	-.173	運動休閒	.174	電腦週邊	-.426	婦幼用品	-.075	保養	.641
中國大陸	-.212	美髮造型	-.213	男內褲	.379	東南亞	.453	材質	.586
汽車	.385	長程	-.325	宗教	-.421	LV	.271	中台灣	.622
料理	.254	鉑金	.309	字號	-.511	書房	-.601	水	-.533
寢飾	-.230	瑞士錶	-.219	設備	-.497	東台灣	.512	卡拉OK	-.631
床罩	-.027	網路設備	-.430	百貨專櫃	.049	東北亞	-.334	男仕	-.482
南臺灣	.321	男內衣	-.081	汽車百貨	.382	專櫃品牌	.064	碗盤	.465
沖泡	.609	居家清潔	-.033	劇院	.014	自行車	.084	離島	.336
翡翠	-.576	銀髮	.605	休閒	.042	面膜	-.456	COACH	-.282

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
玉石				包					
健康	-.684	日系	-.531	生活家電	.004	睡衣	-.360	季節家電	-.266
東森嚴選	-.213	泳裝	.291	即食	.152	女鞋	-.357	專櫃女鞋	-.049
用品	-.621	廚衛	-.326	廚房家電	-.121	比基尼	-.076	居家服	.032
流行女鞋	-.533	美日	.405	票券	.310	塑身	-.320	米麵	.119
百貨	.604	歐系	-.242	日韓	.018	塑身衣	.264	鞋材	-.182
臥室	.668	配飾	-.272	熱門	.408	平板電腦	.486	特惠	-.134
NB	-.001	成分	-.563	造型	-.178	男鞋	-.007		
居家	-.534	人身部品	-.210	按摩	-.438	男包	.109		

註：已顏色標註之商品項目，為第一主成分組成資訊

附 6-7 Fam-FamiPort 主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
預付卡	.973	運動	.897	娛樂	.827	健康生活	-.574	電信	.459	好康	-.382
點數卡	.964	雲端列印	.895	atm	.802	服務	.560	休閒	.555	信用卡	-.541
儲值	.956	台鐵	.894	famiport	.791	自動儲值	.539	旅遊	.570	展演	.079
會員	.948	折價券	.892	快遞	.786	票券	.473	命理	-.508	網路	.125
壽星	.941	統聯	.882	91 全家館	.779	帳號管理	-.070	店到店	.469	職棒購票	.217
順豐	.938	查詢	.874	彈性	.742	電子地圖	-.367	照護	.487	網路會員	-.045
客運	.937	行動	.867	點數兌換	.698	公家機關	.250	列印服務	-.201	HAPPY GO 卡	.246
繳費	.936	電影票	.864	高鐵	.689	兩岸	-.047	獨享	.277		
點數	.924	潔衣家	.845	住宿	.621	遊戲	.393	宅配通	.619		
集點	.914	代碼繳費	.844	溫泉	.585	銀行保險	.403	限時優惠	.332		

註：已顏色標註之商品項目，為第一主成分組成資訊

附 6-8 Goh-GoHappy 主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
周邊	.985	洗衣機	.957	開架	.943	紙本	.879	保養	.765
耳機	.981	機車	.957	美髮	.941	運動	.865	宅配	.764
客廳	.979	女裝	.953	戶外	.940	品牌	-.849	太陽眼鏡	.741
沙發	.978	儲存	.953	眼鏡	.935	內衣	.845	鍋具	.726

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
清潔	.974	行李箱	.952	童裝	.926	好禮	.843	餐廚	.717
收納	.970	滷味	.952	手錶	.917	書包	.832	咖啡	.717
藥妝	.965	嚴選	.952	香氛	.910	蛋糕	.829	婦嬰	.715
專櫃	.964	點數	.951	生鮮	.887	珠寶	.824	美容家電	.707
甜點	.963	餐券	.949	寵物	.885	美食	.815	影音	-.693
床墊	.959	雲端	.948	彩妝	.881	蔬果	.798	寢具	.667
體育	.652	專櫃包	.545	傢俱	.101	保健食品	-.427	旅遊	.468
其他	-.646	手機周邊	.527	瑞士錶	-.052	精選	-.356	植栽	-.024
鞋	-.640	獨立筒枕	.524	家飾	-.021	茶	-.148	用品	-.492
牛仔	-.626	穿戴	.518	機能	-.173	辦公用品	-.278	教具	-.488
日用	-.622	窈窕	-.501	電腦	.348	書房	-.470	週邊	.565
男士	-.621	電子票券	.459	總覽	-.046	內著	.300	運動機能	.218
冰箱	-.615	印表機	-.029	數位相機	-.289	休閒	.433	金飾	.282
特賣	.565	臥房	.032	兒童用品	-.210	瑜珈	-.386	汽車百貨	.539
智慧型手機	.552	電視	.177	國際精品	.079	生活百貨	.353	longchamp	.427
健身器材	-.545	織品	-.053	童書	.210	專櫃品牌	.165	美體	.558
紳士	.405	熟食	-.291	票券	.411	廚房用具	.119		
流行	-.647	生活家電	.072	單眼相機	.490	童襪	.015		
香水	-.359	電玩	-.301	網路	.087	零嘴	.026		
人身部品	-.130	衛浴用品	-.342	下著	.372	收納用品	-.201		
玩具	-.572	廚房家電	-.048	人氣	.155	特搜	-.260		
名牌	-.567	運動休閒鞋	-.095	溫泉	.558				
熟女	.223	日韓	.081	coach	-.326				
apple	.477	按摩	-.449	罐裝飲料	-.030				
家電	-.310	小吃	-.529	季節家電	-.289				

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-9 Ibo-ibon 主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
數位	.979	政府	.899	叫車服務	.815	展覽	.022	套版	.250
講座	.960	商品	.882	銀行	.810	電影	.213	海報	.471
圖像	.942	戲劇	.876	線上遊戲	.790	運動	.086	專區	.394
圖片	.929	外籍	.867	加油站	.739	親子	-.294	舞蹈	.138
證券	.923	分割列印	.861	銀行保險	.735	會員	-.361	交貨便	-.578
廣告刊登	.923	有線電視	.860	掃描	.729	勞動部	-.096	保險	.510
生活情報	.919	文件	.851	電信	.664	電子發票	-.375	旅遊	.662
網路遊戲	.915	活動	.841	第三方支付	-.547	學校	.385	演唱會	-.521
交通	.906	金融	.840	便民服務	.489	兌換	-.339	紅利點數	.109
音樂	.902	授權	.822	列印	-.026	悠遊卡	-.413	一卡通	.560
ibon mobile	-.458	服務	.469	找工作	.160	icash	-.408	代收	.437
好康	-.390	慈善	.470	捐款	.037	慈善捐款	.341	統一超商電信	-.101

限時 搶購	.193	電子票證	.114					
----------	------	------	------	--	--	--	--	--

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-10 Mom-momo 主成份解釋變異量資訊

商品	平方差累積 變異量	商品	平方差累積 變異量
精品	.982	家電	.664
保健食品	.965	保養	.041
生活用品	.886	服裝	.110
包包	.869	休閒	.442
生活	.858	寢具	-.173
資訊	.855	美妝	-.637
團購	-.826	美食	-.393
團購美食	-.826	內著	.470
傢俱	.705	鞋子	-.456
3C	.677		

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-11 Myf-myfone 主成份解釋變異量資訊

商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量
配件	.965	電子	.824	家電	.597	3C	.467
筆電	-.946	日用	.807	台灣	.053	香氛	-.617
食尚	-.941	OUTLET	-.783	保養	-.266	消費	.403
精品	.940	居家	.763	服飾	-.201	戶外	-.658
保健	.938	數位	.754	汽機車	.022	寢具	-.402
食品	.934	電腦週邊	.741	婦幼	-.382	票券	-.086
影音	.902	智慧手機	-.730	相機	.257	鞋子	-.440
包包	.899	手機	-.701	傢俱	.474	平板	-.520
手機平板	-.874	DIY	-.675	內著	.220		
美妝	-.830	手錶	-.667	列印	-.599		

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-12 Pch-PChome 主成份解釋變異量資訊

商品	平方差累積 變異量	商品	平方差累積 變異量
包包	.930	家電	.607
生活	.928	通訊	.506
保健	.912	3C	.393
時尚	-.911	NB	-.200
美妝	-.881	數位	.677
週邊	-.864	運動	-.577
食品	.856	鞋子	-.304
日用	.856	衣服	.218
手錶	-.731		
戶外	-.710		

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-13 Rak-樂天市場主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
安全	.942	耗材	.878	手機	-.770	生鮮	-.679	衛浴用品	.557
配件	.940	甜點	-.877	居家用品	.766	香氛	-.672	男性	-.547
廚房	-.937	周邊	-.868	軟體	.762	電腦周邊	.662	女士	.532
家具	.924	人社	.848	居家	.742	嗜好	.657	外出	-.517
包包	.920	文具用品	.846	手錶	-.739	彩妝	-.655	兒童服飾	.500
生活	.913	收藏	.838	情趣商品	.734	美食	-.650	休閒	.028
精品	.906	玩具	.823	戶外	-.710	商科	.634	電腦	.142
保健	.904	用品	.821	花	.692	茶品	.627	園藝	-.062
女裝	-.896	清潔	-.817	潮流	.686	家電	.572	孕婦裝	-.173
影音	.887	調味	-.788	話機	.685	電玩	.558	保養	-.379
美體	-.148	內衣	-.542	睡衣	.269				
汽機車	-.043	運動	-.592	包款	-.024				
飲料	-.044	嬰幼兒用品	.549	男裝	-.110				
相機	.179	寢飾	.492	零食	-.441				
服裝	-.307	美髮	-.689	生活家電	.208				
音樂	.432	影片	.109	流行男裝	.347				
百貨	-.359	良品	-.619						
藝術品	.371	生活雜貨	.496						
寵物用品	-.185	修繕	.129						
飾品	-.536	女鞋	.265						

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-14 Rut-露天拍賣主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
FAC_RUT_書籍文創科學 1	.990	FAC_RUT_女包精品與女鞋 1	.930	FAC_RUT_不動產百業服物 1	.735	FAC_RUT_音樂電影 2	.163	FAC_RUT_手機通訊 1	-.537
FAC_RUT_生活居家 1	.988	FAC_RUT_保養彩妝 2	.903	FAC_RUT_流行女裝 1	-.725	FAC_RUT_玩具公仔 2	-.254	FAC_RUT_男裝配件男鞋 1	-.521
FAC_RUT_電腦電子周邊 1	.987	FAC_RUT_名產食品 2	.864	FAC_RUT_女性飾品配件 2	.725	FAC_RUT_女性內衣 2	-.070	FAC_RUT_流行女裝 2	.509
FAC_RUT_音樂電影 1	.978	FAC_RUT_鐘錶眼鏡 2	.848	FAC_RUT_休閒旅遊 2	.714	FAC_RUT_明星偶像 1	-.272	FAC_RUT_不動產百業服物 2	.632
FAC_RUT_古董收藏 1	.967	FAC_RUT_手機通訊 2	.832	FAC_RUT_書籍文創科學 2	-.080	FAC_RUT_相關攝影機 2	-.190	FAC_RUT_電玩遊戲 2	-.553
FAC_RUT_相關攝影機 1	.959	FAC_RUT_明星偶像 2	.824	FAC_RUT_電腦電子周邊 2	.079	FAC_RUT_名產食品 1	-.357	FAC_RUT_家電影音周邊 2	.517
FAC_RUT_嬰幼童母親	.949	FAC_RUT_男裝配件男鞋 2	.810	FAC_RUT_運動健身 2	-.166	FAC_RUT_交通工具 1	-.354	FAC_RUT_女包精品與女鞋 2	-.004

FAC_RUT _玩具公仔 1	.948	FAC_RUT _電玩遊戲 1	.808	FAC_RUT _生活居家 2	.038	FAC_RUT _鐘錶眼鏡 1	-.459	FAC_RUT _名產食品 3	.326
FAC_RUT _運動健身 1	.947	FAC_RUT _家電影音周邊 1	.800	FAC_RUT _保養彩妝 1	-.190	FAC_RUT _休閒旅遊 1	-.367	FAC_RUT _玩具公仔 3	.134
FAC_RUT _女性內衣 1	.939	FAC_RUT _交通工具 2	.751	FAC_RUT _古董收藏 2	-.121	FAC_RUT _女性飾品配件 1	-.494	FAC_RUT _男裝配件男鞋 3	-.080
FAC_RUT _流行女裝 3	-.353	FAC_RUT _女性內衣 3	.059	FAC_RUT _鐘錶眼鏡 3	.037				
FAC_RUT _明星偶像 3	-.386	FAC_RUT _古董收藏 3	.188	FAC_RUT _保養彩妝 4	.004				
FAC_RUT _女性飾品配件 3	.068	FAC_RUT _相關攝影機 3	.046	FAC_RUT _電腦電子周邊 4	.085				
FAC_RUT _不動產百業服物 3	.201	FAC_RUT _家電影音周邊 5	-.070						
FAC_RUT _相關攝影機 4	-.093	FAC_RUT _名產食品 5	.098						
FAC_RUT _家電影音周邊 4	-.207	FAC_RUT _男裝配件男鞋 5	.034						
FAC_RUT _保養彩妝 3	.229	FAC_RUT _家電影音周邊 3	-.131						
FAC_RUT _電腦電子周邊 3	.038	FAC_RUT _男裝配件男鞋 4	-.192						
FAC_RUT _玩具公仔 4	.058	FAC_RUT _女性內衣 4	.169						
FAC_RUT _名產食品 4	-.015	FAC_RUT _電玩遊戲 3	-.039						

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-15 Rut-露天拍賣主成份項目資訊

類別成份矩陣	關鍵字									
FAC_RUT_書籍文創科學 1	雜誌	文學	語言用書	設計	科學	工具書	考試用書	音樂	辦公用品	社會科學
	青少年	外文書	兒童讀物	簡體書	英文書	教育	小說	日文書	心理	書籍
	善本	休閒娛樂	藝術	觀光旅遊	電子書	字典	攝影	保健	漫畫書	禮品
	行銷	古書								
FAC_RUT_生活居家 1	衛生	居家安全	花草	衛浴設備	家具	五金工具	家庭	寢飾	生活雜貨	廚房用品
	庭院	清潔用品								

FAC_RUT_電腦電子周邊 1	電子	製造	貿易	筆記型電腦	電子字典	設備	耗材	高科技	線材	測量
	儀器	電子零件	零組件	化工	網路設備	容器	配件	電腦周邊	材料	電腦軟體
	出租	租賃	導航	桌上型電腦						
FAC_RUT_音樂電影 1	樂譜	vcd	錄音帶	錄影帶	刊物	立牌	cd	黑膠唱片	dvd	影片
FAC_RUT_古董收藏 1	企業	竹製品	民俗	古董家具	木製品	貝類	宗教	獎章	藏書票	作品
	標本	書畫	戲偶	文房四寶	玻璃	化石	文獻	鼻煙壺	器皿	紙鈔
	紀念票	古董相機	券	礦石	收藏品					
FAC_RUT_相關攝影機 1	數位單眼相機	零件	dslr	dc	鏡頭	類單眼	微單眼	自拍	數位相機	周邊
	運動攝影機	視訊	翻轉相機	evil						
FAC_RUT_嬰幼兒母親	文具	哺育	童裝	嬰兒服						
FAC_RUT_玩具公仔 1	連鎖店	18 禁	日本	國際	益智遊戲	GK 模型	鐵道模型	整人玩具	遙控模型	紙牌遊戲
	魔術道具	塑膠	卡漫	玩具	lego	玩偶	組裝模型			
FAC_RUT_運動健身 1	按摩器材	健身	運動服飾	運動護具						
FAC_RUT_女性內衣 1	成套	角色扮演	調整型內衣	束衣	胸罩	肚兜	家居服			
FAC_RUT_女包精品與女鞋 1	名牌	精品包								
FAC_RUT_保養彩妝 2	薰香	身體保養	香水	足部保養	臉部保養	男士清潔	眼部保養			
FAC_RUT_名產食品 2	肉類	小家電	調理包	茶類	製品	有機食品	外燴	飲料	香料	罐頭食品
	營養品	特產	餐飲	食品	餐具	穀物				
FAC_RUT_鐘錶眼鏡 2	懷錶	對錶	古董錶							
FAC_RUT_手機通訊 2	無線電	手機門號	手機配件	模型	答錄機	專案	iphone	樣品機		
FAC_RUT_明星偶像 2	團體	男藝人	寫真集	運動明星	日本女藝人					
FAC_RUT_男裝配件男鞋 2	公事包	首飾	道具服	男錶	隨身	男士保養	內衣褲	背心	男鞋	手帕
FAC_RUT_電玩遊戲 1	攻略	軟體	大型電玩	nintendo 3ds	電競	ps vita	商品	playstation portable	playstation 4	任天堂 Wii U
	xbox one	電玩								
FAC_RUT_家電影音周	充電	健康	影音	耳機	品牌	吸塵	洗衣	冷藏	冰箱	喇叭

邊 1	器					器	機			
	mp3	智慧型	空氣清淨	空調	環控	除濕	播放機	淨水設備		
FAC_RUT_交通工具 2	物聯網	機車百貨	自行車							
FAC_RUT_不動產百貨服物 1	土地	租售	測試	工商	房屋出租	創業加盟	ruten	房屋出售	外包	居家服務
	接案									
FAC_RUT_流行女裝 1	套裝	洋裝	禮服	小可愛	專區	中國風				
FAC_RUT_女性飾品配件 2	別針	眼鏡盒								
FAC_RUT_休閒旅遊 2	國內旅遊									

附表 6-16 Tao 淘寶網主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
珠寶	.924	服務	.744	箱包	-.346
汽車	.920	男裝	.719	建材	-.135
百貨	.917	家裝	.707	家電	.084
運動	.896	採購	.702	女裝	.677
辦公	.887	美食	.678	配飾	.190
手機	.864	市場	.667	花鳥	-.215
娛樂	.861	家居	-.237		
戶外	.854	用品	.068		
母嬰	.841	農業	.217		
數碼	.746	鞋類	.309		

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-17 Yah1-Yah 全部商品主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
耳機	.980	運動	.830	8H 急速配	.565	飲料	.369	票券	.365
美妝	.975	廚具	.828	電腦	.253	食品	-.636	內睡衣	-.132
清潔	.966	珠寶	.784	相機	.196	精品	-.671		
餐券	.959	包包	-.778	視聽	-.420	服裝	.639		
手機	.948	婦幼	.746	軟體	-.432	洗沐-沐浴	-.200		
機車	.934	家具	-.713	汽車	.453	家電	-.372		
童裝	.932	保健	-.701	衛浴	.370	電玩	-.362		
戶外	.920	DVD	-.690	週邊	.481	配飾	-.163		
手錶	.902	寢具	.637	文具	-.529	鞋子	.268		
GPS	-.865	玩具	-.633	書籍	-.304	童鞋	-.356		

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-18 Yah b-Yahoo 拍賣主成份解釋變異量資訊

商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量	商品	平方差累積變異量
FAC_YA HB_圖書 影音	.983	FAC_YA HB_相機 攝影 2	.946	FAC_YA HB_居家 家具 3	.748	FAC_YA HB_美食 與地方 1	.084	FAC_YA HB_古董 藝術 1	.547
FAC_YA HB_原創 設計 2	.976	FAC_YA HB_女裝 與服飾 5	.937	FAC_YA HB_嬰幼 兒 1	.702	FAC_YA HB_汽機 車精品 1	.298	FAC_YA HB_偶像 球卡 1	.583
FAC_YA HB_電腦 平板 1	.975	FAC_YA HB_偶像 球卡 2	.922	FAC_YA HB_居家 家具 2	.700	FAC_YA HB_手錶 與飾品	.396	FAC_YA HB_運動 戶外 1	.636
FAC_YA HB_古董 藝術 2	.970	FAC_YA HB_美容 保養 2	.918	FAC_YA HB_相機 攝影 3	.686	FAC_YA HB_家電 影音 1	.362	FAC_YA HB_相機 攝影 1	-.132
FAC_YA HB_家電 影音 2	.966	FAC_YA HB_運動 戶外 3	.895	FAC_YA HB_電玩 遊戲 3	-.641	FAC_YA HB_女裝 與服飾 1	.314	FAC_YA HB_女裝 與服飾 2	-.309
FAC_YA HB_汽機 車精品 2	.961	FAC_YA HB_女包 精品 3	.891	FAC_YA HB_家電 影音 3	.476	FAC_YA HB_電玩 遊戲 1	.040	FAC_YA HB_古董 藝術 3	.110
FAC_YA HB_美食 與地方 2	.953	FAC_YA HB_嬰幼 兒 2	.861	FAC_YA HB_汽機 車精品 3	.400	FAC_YA HB_男性 精品	.265	FAC_YA HB_偶像 球卡 3	-.327
FAC_YA HB_手機 配件 2	.952	FAC_YA HB_寵物 用品 2	.859	FAC_YA HB_手機 配件 1	-.026	FAC_YA HB_女包 精品 1	.358	FAC_YA HB_原創 設計 3	.023
FAC_YA HB_玩具 模型 1	.950	FAC_YA HB_寵物 用品 1	.837	FAC_YA HB_電腦 平板 2	.239	FAC_YA HB_玩具 模型 2	.515	FAC_YA HB_女包 精品 2	-.313
FAC_YA HB_電玩 遊戲 2	.949	FAC_YA HB_居家 家具 1	.775	FAC_YA HB_原創 設計 1	.154	FAC_YA HB_美容 保養 1	.571	FAC_YA HB_美食 與地方 3	-.054
FAC_YA HB_運動 戶外 4	-.219	FAC_YA HB_女包 精品 5	.107	FAC_YA HB_嬰幼 兒 3	-.007	FAC_YA HB_美食 與地方 4	-.075	FAC_YA HB_女包 精品 4	.291
FAC_YA HB_玩具 模型 3	.394	FAC_YA HB_女裝 與服飾 4	.236	FAC_YA HB_手機 配件 4	.246	FAC_YA HB_電玩 遊戲 4	.046	FAC_YA HB_原創 設計 4	-.256
FAC_YA HB_美容 保養 3	-.058	FAC_YA HB_女裝 與服飾 3	-.321	FAC_YA HB_電腦 平板 3	.057	FAC_YA HB_美容 保養 4	.293	FAC_YA HB_嬰幼 兒 4	-.164
FAC_YA HB_古董 藝術 4	-.249	FAC_YA HB_運動 戶外 2	-.244	FAC_YA HB_手機 配件 3	-.478	FAC_YA HB_玩具 模型 4	.142		

註：已顏色標註之商品項目，為第一主成分組成資訊

附表 6-19 Yah b-Yahoo 拍賣主成份項目資訊

類別成份矩陣	關鍵字									
FAC_YAHB_圖書影音	圖書	雜誌	音樂	文具	影片					
FAC_YAHB_原創設計 2	設計	生活雜貨	嬰幼	居家	手工					
FAC_YAHB_電腦平板 1	筆記型電腦	pda	cpu	零組件	耗材	網路設備	主機板	線材	記憶體	印表機
	蘋果電									

	腦									
FAC_YAHB_古董藝術 2	金屬 商品	紫砂壺 陶器	蜜蠟	藝術品	貝類	書畫	風水	文獻	銅	開運
FAC_YAHB_家電影音 2	電視	投影機	布幕	冷氣空調						
FAC_YAHB_汽機車精品 2	汽車配件	機車配件	機車零件	汽車零件	衛星導航					
FAC_YAHB_美食與地方 2	生鮮食品	乾貨	保健食品	有機食品	南北	零嘴				
FAC_YAHB_手機配件 2	模型機	手機吊飾	無線電	iphone	雙卡雙待手機	展示機	行動網卡			
FAC_YAHB_玩具模型 1	GK 模型	紙牌遊戲	布袋戲偶	遙控模型	鐵道模型	模型	整人玩具	益智遊戲	魔術道具	絨毛玩偶
	金屬模型	玩具	速食店							
FAC_YAHB_電玩遊戲 2	多媒體	ps3	主機	攻略	電玩遊戲					
FAC_YAHB_相機攝影 2	相機	攝影機	記憶卡	底片相機	拍立得相機	讀卡機				
FAC_YAHB_女裝與服飾 5	婚紗									
FAC_YAHB_偶像球卡 2	紙鈔	紀念票								
FAC_YAHB_美容保養 2	衛生用品	薰香	保健衛生	香水	足部保養	眼部保養	體香劑	精油		
FAC_YAHB_運動戶外 3	提包									
FAC_YAHB_女包精品 3	精品包	跟鞋	訂製鞋							
FAC_YAHB_嬰幼兒 2	用具	用品	哺育	安全用品						
FAC_YAHB_寵物用品 2	鼠	兔								
FAC_YAHB_寵物用品 1	貓用品	水族用品	鳥用品	昆蟲	兩棲	爬蟲	狗用品			
FAC_YAHB_居家家具 1	照明	日用品	文具用品	五金	家庭	事務用品	廚房	浴室用品	庭院	床墊
	雜貨	餐具	居家安全	鍋具						
FAC_YAHB_居家家具 3	收納用品									
FAC_YAHB_嬰幼兒 1	幼兒	嬰幼兒	兒童	孕婦裝	兒童玩具	外出	寢具	包鞋		
FAC_YAHB_居家家具 2	手工藝	傢俱	園藝	家飾	衛浴設備	清潔用品				
FAC_YAHB_相機攝影 3	LOMO 相機									
FAC_YAHB_電玩遊戲 3	虛擬實境									
FAC_YAHB_家電影音 3	生活家電	廚房家電								
FAC_YAHB_汽機車精品 3	賽車									

附表 6-20 Yah s-Yahoo 服務主成份解釋變異量資訊

商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量	商品	平方差累積 變異量
專業	.971	國內機票	.888	程式設計	.853	國外旅遊	.801	團體	.740
裝潢	.961	接案	.888	攝影	.845	水上摩托車	.797	禮儀	.736
沙灘車	.947	輕型機車	.887	國內訂房	.837	進修	.795	服務	.711
試用	.939	活動	.887	創業加盟	.836	占卜	.795	敞篷車	.708
車輛	.936	餐飲服務	.886	預售屋	.831	融資	.780	家事管理	.701
休閒	.909	室內設計	.884	飯店	.822	新成屋	.779	四門	.694
吉普車	.909	海外	.879	商用車	.821	空調	.765	居家清潔	.690
旅遊	.906	大型重型機車	.876	代標	.809	殯葬	.757	代書	.667
重型機車	.898	國內	.871	航空假期	.808	平面	.757	當舖	.662
房車	.893	國際機票	.862	電動自行車	.805	網站	.744	留遊學	.634
雙門	.611	名店	.358	開鎖	.331	寵物	.468	搬家	-.114
修繕	.589	代購	.004	工商	.353	國外	.525	修改	.634
好康	-.557	翻譯	.049	跑車	.339	算命	.601	代辦	.594
國外訂房	.541	車位	.062	護照	.405	租借	.514	土地	.672
維修服務	.531	結婚	.036	自由行	.123	租售	.508	工作職缺	.382
motel	.518	訂做	.189	出租	.371	搬家公司	.023	快遞	.330
行李箱出租	-.482	租車	.226	清洗	.264	特色	.416	簽證	.570
spa	-.472	育兒	.067	電動機車	.268	休旅車	.586	考照	.484
掀背車	.447	客製化	.260	外包	.207	旅行車	.413	售屋	-.015
轎跑	-.414	婚禮	.054	攤位	.421	房屋	.619	租屋	.623
宅配	.563	冷氣	.296	美甲	-.006	配鎖	.139	機場接送	-.009
網拍	.281	民宿	.520	票券	.433	貸款	-.009	經紀	.024
店面	.433	徵信社	.100	美容	.168	國內旅遊	.417	共乘	.336
高鐵假期	-.333	ATV	-.344	水電維修	.496	貨卡車	-.175	遊輪假期	-.140

註：已顏色標註之商品項目，為第一主成分組成資訊

貳、十五電商主成份分析效度

KMO 為取樣適當性衡量量數，值介於 0 至 1 之間，越大表示主成分效度越高。附表 5-21 呈現出十五大電商成分因子之 KMO 資訊，以及 Bartlett 檢定資訊，Bartlett 值顯著表示因子間相關係數越高，適宜進行主成份分析。

附表 6-21 十五電商主成份分析效度

編號	類組特色	變數個數	KMO	Bartlett 球形檢定
1	Boo	29	0.918	0 (顯著)
2_1	Cos	118	0.849	0 (顯著)
2_2	Cos	25	0.923	0 (顯著)
3	Etm	198	NA	0 (顯著)
4	Fam	57	0.931	0 (顯著)
5	Goh	135	0.737	0 (顯著)

編號	類組特色	變數個數	KMO	Bartlett 球形檢定
6	Ibo	62	0.926	0 (顯著)
7	Mom	19	0.894	0 (顯著)
8	Myf	38	0.927	0 (顯著)
9	Pch1	14	NA	0 (顯著)
10	Rak	76	0.926	0 (顯著)
11_1	Rut	475	NA	0 (顯著)
11_2	Rut	73	0.881	0 (顯著)
12	Tao	26	0.904	0 (顯著)
13	Yah1	42	0.931	0 (顯著)
14_1	Yah b	393	0.973	0 (顯著)
14_2	Yah b	69	0.865	0 (顯著)
15	Yah s	120	0.875	0 (顯著)

附件七 應用大數據提升台灣民間消費預測-VAR 實證分析

VAR_TRADF 及 VAR_GTF 兩模型皆經由 Lagrange-multiplier test，兩者在落後三期時無序列相關，配合本研究之 VAR 模組分析。並由 Jarque-Bera test，符合常態分配，請見表 5-16 及表 5-17。

附表 7-1 Lagrange-multiplier 檢定

模型	VAR-GT 模型		VAR-傳統模型	
變數	chi2	Prob>chi2	chi2	Prob>chi2
落後 1 期	51.980	0.041	44.615	0.000
落後 2 期	36.664	0.438	35.316	0.004
落後 3 期*	34.356	0.547	21.093	0.175

附表 7-2 Jarque-Bera 檢定結果

模型	VAR-GT 模型		VAR-傳統模型	
變數	chi2	Prob>chi2	chi2	Prob>chi2
dlog_pc	1.965	0.374	0.857	0.652
dlog_rgdp	1.749	0.417	1.017	0.602
cpiyoy	0.447	0.800	1.216	0.544
buseyc_fac_1	4.726	0.094	1.911	0.385
dgt_fac_1	0.040	0.980		
dgt_fac_2	1.405	0.495		
ALL	10.332	0.587	5.001	0.758

以下則就各模型之因果檢定，以及衝擊反應與預測誤差拆解個別呈現。

一、因果檢定

Granger causality 因果檢定表示在各估計模型中，移除各別變數的模型效力，據此驗證該變數對應變數的影響，若表中估計之 chi2 值為顯著，則該變數對應變數有一定影響力。附表 7-3 及 7-4 為 VAR-傳統模型、VAR-GT 模型之 Granger causality 因果檢定結果。其中，可發現 dgt_fac_1 顯著影響我國

民間消費之變化。

附表 7-3 VAR-傳統模型 Granger causality 因果檢定

Excluded	dlog_pc	dlog_rgdg	cpiyoy	buscyc_fac_1
dlog_pc		9.3678**	6.252	2.4303
dlog_rgdg	35.812***		4.374	5.4141
cpiyoy	23.786***	1.6579		16.247***
buscyc_fac_1	2.3427	31.485***	8.4661**	
ALL	95.926***	174.59***	16.392**	31.98***

*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

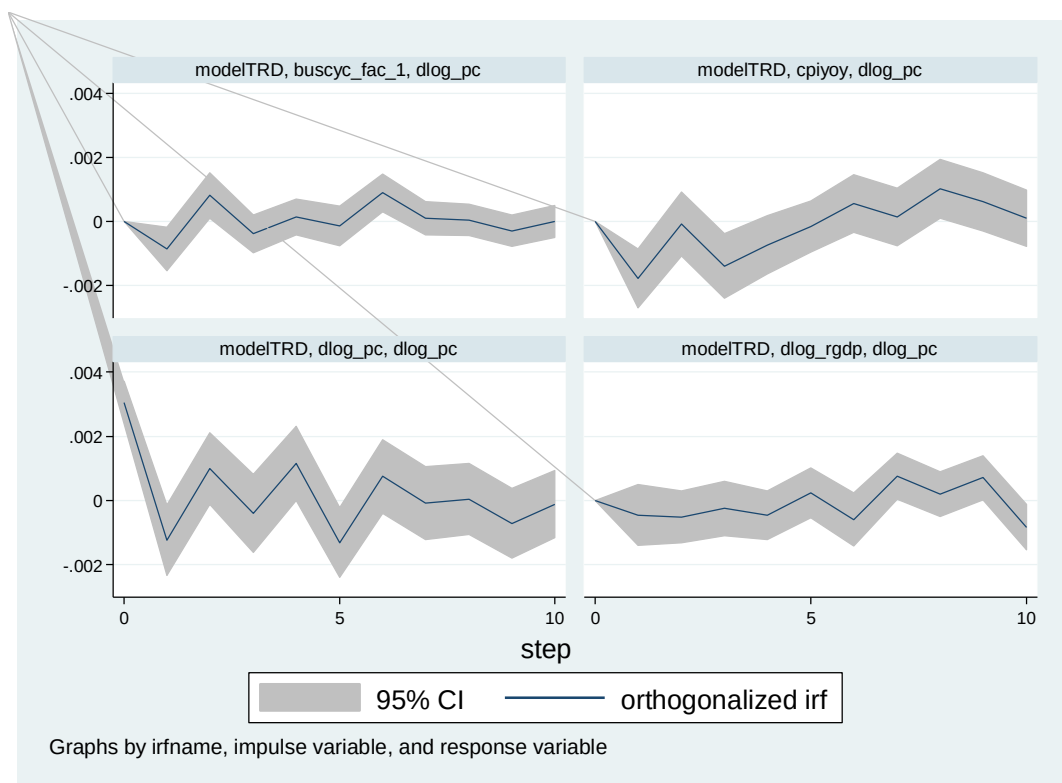
附表 7-4 VAR-GT 模型 Granger causality 因果檢定

Excluded	dlog_pc	dlog_rgdg	cpiyoy	buscyc_fac_1	dgt_fac_1	dgt_fac_2
dlog_pc		5.4221	5.6948	2.011	4.385	7.6367*
dlog_rgdg	11.727***		4.0697	4.2614	18.759***	41.071***
cpiyoy	34.313***	10.069**		15.458***	6.7027*	5.8568
buscyc_fac_1	3.0018	29.607***	7.3455*		8.1801**	11.881***
dgt_fac_1	8.9327**	5.9376	3.29	1.4529		5.8579
dgt_fac_2	6.1032	4.8547	4.4931	1.2209	11.018**	
ALL	128.32***	429.83***	26.655**	36.811***	137.22***	342.33***

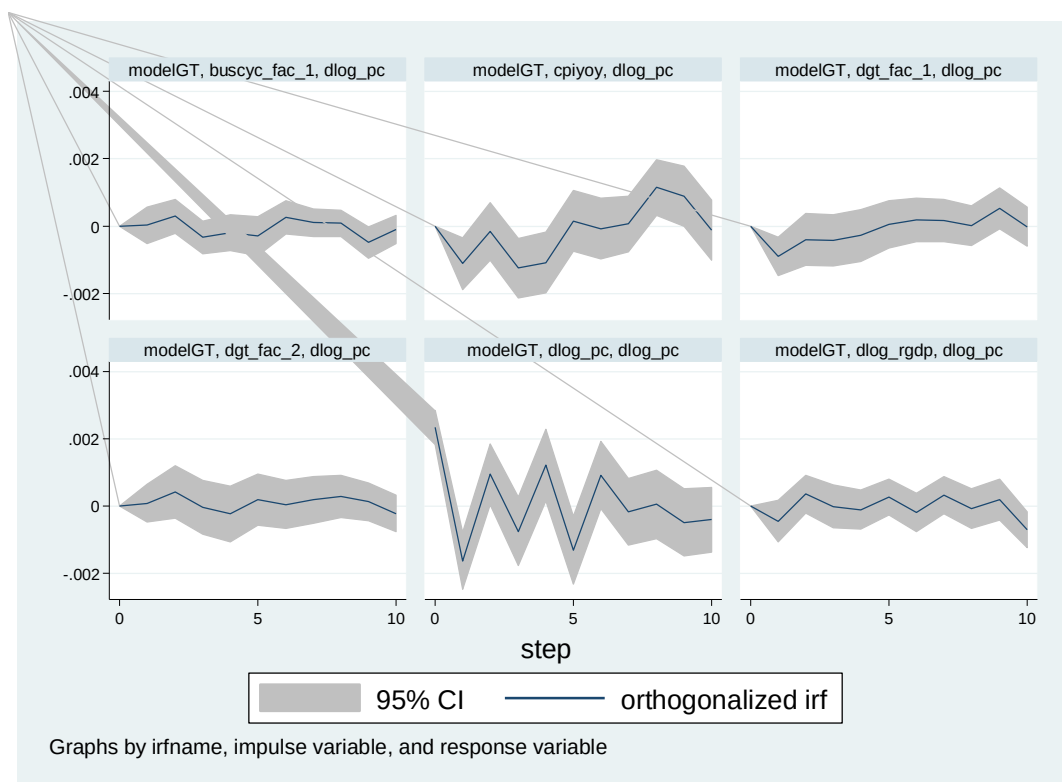
*表示 $\alpha=0.1$ ，**表示 $\alpha=0.05$ ，***表示在 $\alpha=0.01$ 水準下係數檢定為顯著

二、衝擊反應分析

衝擊反應函數是指其他情況不變之下，特定外生衝擊對模型內其他經濟變數所產生的動態衝擊過程。藉由衝擊反應分析可瞭解經濟變數隨時間經過產生自發性衝擊時，對其他經濟變數所造成的跨期動態影響。附圖 7-1 及 7-2 各為兩模組，民間消費在各落後期受到其他變項所影響之變化。由 VAR-傳統模型之結果可發現，其受到 GDP 的影響較為顯著，且並未有明顯收斂情況出現，而領先指標在前五期多為負向影響。在附圖 7-2 中可發現，網路檢索指標第二因子，在前兩期對民間消費的影響相對後期為高。



附圖 7-1 VAR-傳統模型衝擊反應分析



附圖 7-2 VAR-GT 衝擊反應分析

三、預測誤差拆解

預測誤差變異拆解是指將預測誤差的變異拆解成不同衝擊所造成的比率，亦即衡量內生變數波動被特定衝擊所解釋的比例（或貢獻度有多少）。由附表 7-5 可發現，民間消費在各期的變化主要來自自身變異，而消費者物價指數年增率也相較於實質 GDP（drgdp）、總體經濟領先指標因子對民間變異的解釋程度來得高。

附表 7-6 為呈現納入網路檢索指標之 VAR 模型，結果可發現網路檢索指標第一主成分對民間消費的變異解釋在各期皆相較第二因子為高，且與總體經濟領先指標相比，也相對較高。

附表 7-5 VAR-傳統模型預測誤差拆解

落後期	dlog_pc	dlog_rgdg	cpiyoy	buscyc_fac_1
1	1.000	0.000	0.000	0.000
2	0.770	0.003	0.220	0.007
3	0.697	0.106	0.191	0.006
4	0.611	0.092	0.284	0.013
5	0.603	0.090	0.280	0.027
6	0.613	0.085	0.275	0.027
7	0.582	0.120	0.271	0.027
8	0.573	0.137	0.264	0.026
9	0.556	0.130	0.288	0.025
10	0.557	0.134	0.283	0.025

附表 7-6 VAR-GT 模型預測誤差拆解

落後期	dlog_pc	dlog_rgdg	cpiyoy	buscyc_fac_1	dgt_fac_1	dgt_fac_2
1	1.000	0.000	0.000	0.000	0.000	0.000
2	0.819	0.006	0.123	0.009	0.043	0.000
3	0.743	0.076	0.106	0.010	0.064	0.001
4	0.662	0.069	0.190	0.016	0.058	0.005
5	0.646	0.072	0.176	0.022	0.065	0.018
6	0.668	0.066	0.160	0.023	0.063	0.020
7	0.675	0.070	0.154	0.024	0.059	0.018
8	0.673	0.079	0.149	0.023	0.059	0.018
9	0.662	0.076	0.167	0.022	0.056	0.017
10	0.657	0.075	0.159	0.021	0.065	0.024

附件八 民間消費綜合指標預測效力-VAR 實證分析

VAR-PCA、VAR-PCATR、VAR-DFM 及 VAR-DFMTR 模型皆經由 Lagrange-multiplier test。由 Jarque-Bera test，各模組之殘差項皆符合常態分配，請見表附表 8-1 及附表 8-2。以下則就各模型之因果檢定，以及衝擊反應與預測誤差拆解個別呈現。

附表 8-1 Lagrange-multiplier 檢定

模型	VAR-PCA 模型		VAR-PCATR 模型		VAR-DFM 模型		VAR-DFMTR 模型	
變數	chi2	Prob>chi2	chi2	Prob>chi2	chi2	Prob>chi2	chi2	Prob>chi2
落後 1 期	16.846	0.051	8.409	0.078	39.200	0.035	29.362	0.022
落後 2 期	14.722	0.099	4.011	0.404	17.201	0.874	14.866	0.534
落後 3 期	12.483	0.187	7.174	0.127	27.329	0.340	21.202	0.171

附表 8-2 Jarque-Bera 檢定結果

模型	VAR-PCA 模型		VAR-PCATR 模型		VAR-DFM 模型		VAR-DFMTR 模型	
變數	chi2	Prob>chi2	chi2	Prob>chi2	chi2	Prob>chi2	chi2	Prob>chi2
dlog_pc	1.307	0.520	0.632	0.729	0.173	0.917	0.493	0.781
dlog_rgdg					1.497	0.473	0.383	0.826
cpiyoy					1.157	0.561	0.194	0.908
dpca_factor1	0.453	0.797			1.384	0.500		
pca_factor2	0.423	0.809			1.799	0.407		
dfm_factor1			28.353	0.000			0.802	0.670
ALL	2.183	0.902	28.985	0.000	6.011	0.814	1.872	0.985

一、因果檢定

Granger causality 因果檢定表示在各估計模型中，移除各別變數的模型效力，據此驗證該變數對應變數的影響，若表中估計之 chi2 值為顯著，則該變數對應變數有一定影響力。附表 8-3 至 8-6 為各模組之 Granger causality 因

果檢定結果。其中，在未納入總體經濟變數之 VAR-PCA 模型與 VAR-DFM 模型中，民間消費綜合指標對民間消費有顯著影響力。

附表 8-3 VAR-PCA 模型 Granger causality 因果檢定

Excluded	dlog_pc	dpca_factor1	pca_factor2
dlog_pc		9.6443**	14.968**
dpca_factor1	5.9753		3.0517
pca_factor2	16.627***	20.76***	
ALL	31.747***	27.923***	17.142***

附表 8-4 VAR-PCATR 模型 Granger causality 因果檢定

Excluded	dlog_pc	dlog_rgdp	cpiyoy	dpca_factor1	pca_factor2
dlog_pc		10.302**	12.764***	15.248***	20.081***
dlog_rgdp	12.715***		11.084**	13.916***	28.248***
cpiyoy	24.032***	10.592**		6.2851*	23.516***
dpca_factor1	1.2418	0.51786	4.6603		15.071***
pca_factor2	5.2103	31.337***	15.093***	9.7146**	
ALL	115.68***	234.23***	25.261**	69.363***	120.97***

附表 8-5 VAR-DFM 模型 Granger causality 因果檢定

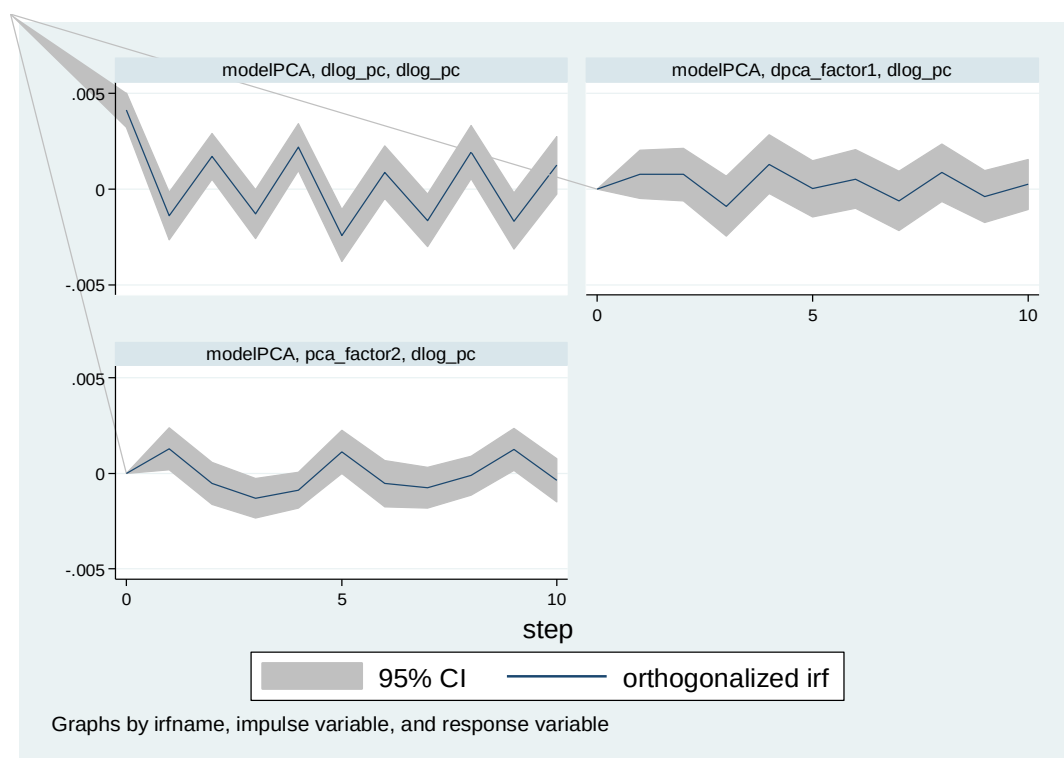
Excluded	dlog_pc	dfm_factor1
dlog_pc		20.992***
dfm_factor1	10.196**	
ALL	10.196**	20.992***

附表 8-6 VAR-DFMTR 模型 Granger causality 因果檢定

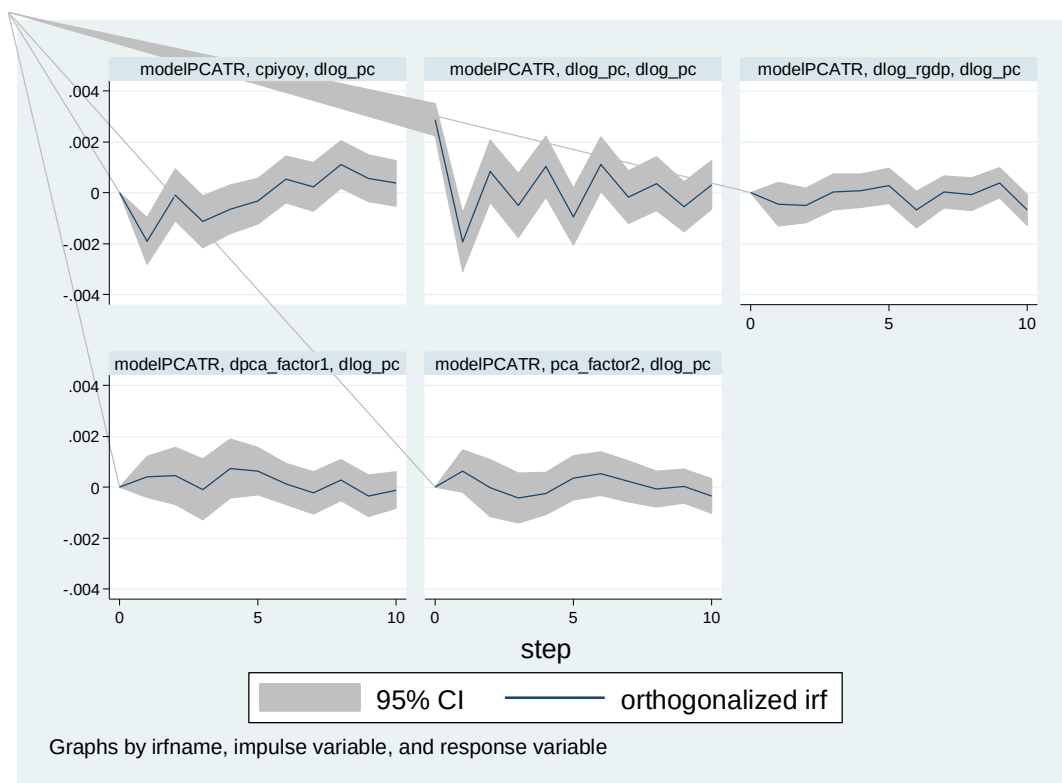
Excluded	dlog_pc	dlog_rgdg	cpiyoy	dfm_factor1
dlog_pc		12.779***	6.1038	20.274***
dlog_rgdg	31.427***		4.398	12.959***
cpiyoy	21.02***	8.7783**		14.811***
dfm_factor1	3.4687	11.903***	3.142	
ALL	99.483***	116.35***	10.215	71.436***

二、 衝擊反應分析

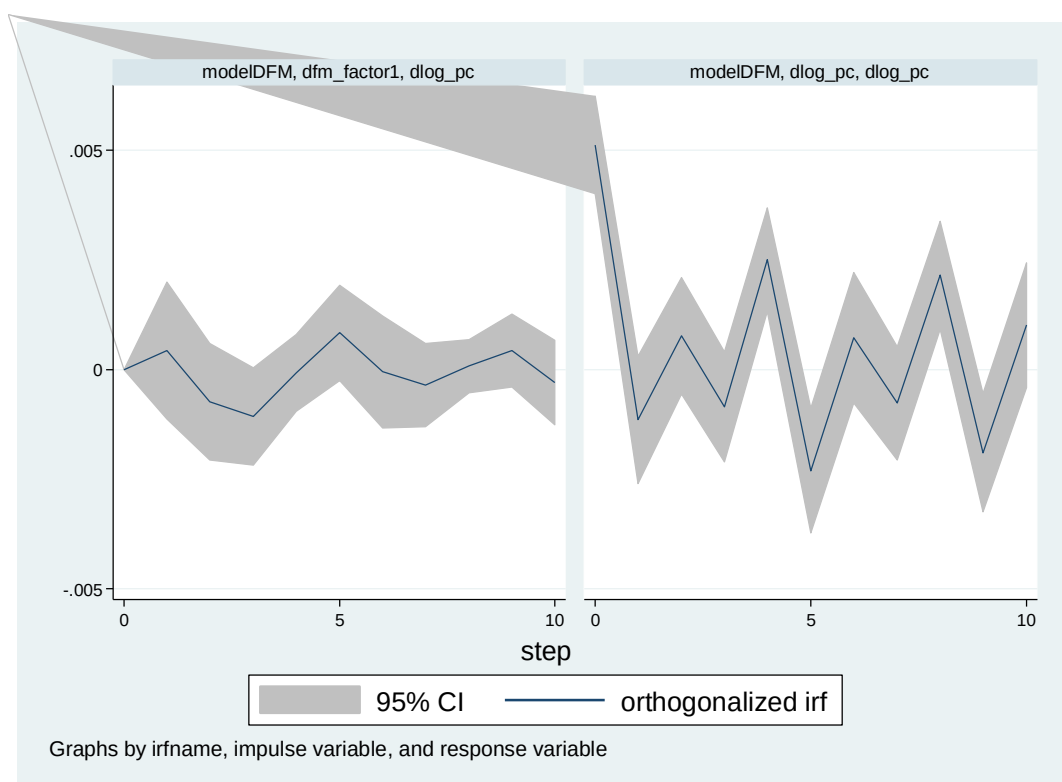
衝擊反應函數是指其他情況不變之下，特定外生衝擊對模型內其他經濟變數所產生的動態衝擊過程。藉由衝擊反應分析可瞭解經濟變數隨時間經過產生自發性衝擊時，對其他經濟變數所造成的跨期動態影響。附圖 8-1 至 8-4，為各模組中，民間消費在各落後期受到其他變項所影響之變化。由 VAR-傳統模型之結果可發現，其受到 GDP 的影響較為顯著，且並未有明顯收斂情況出現，而領先指標在前五期多為負向影響。



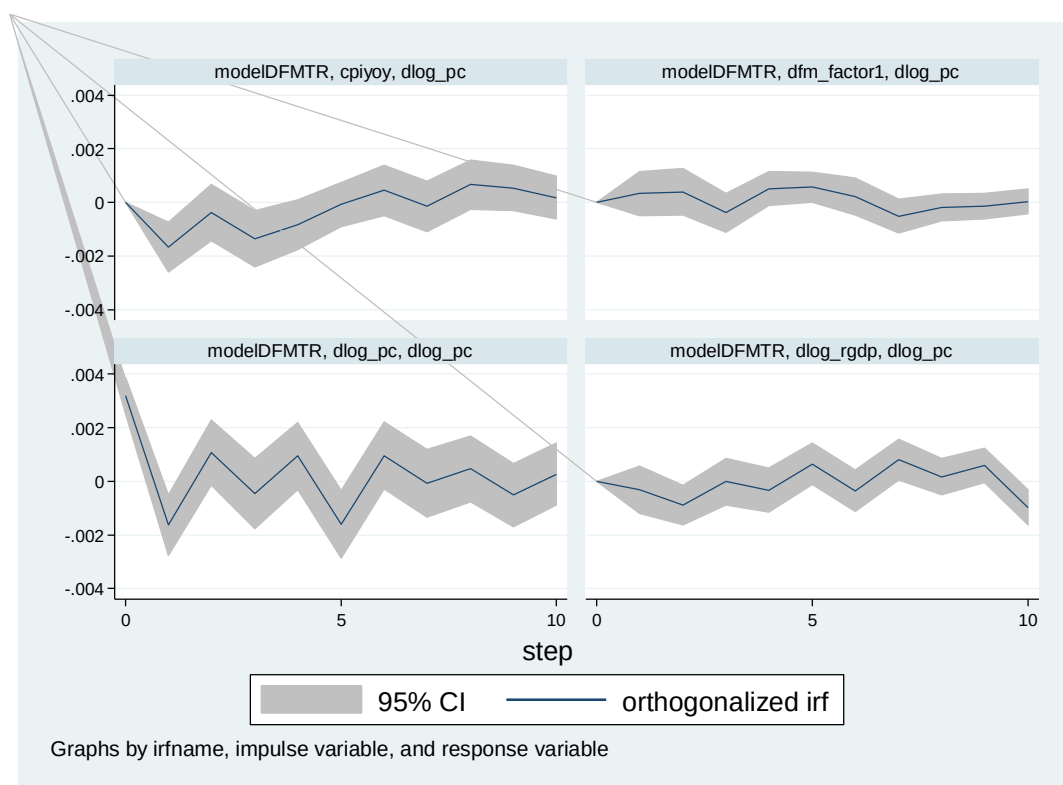
附圖 8-1 VAR-PCA 模型衝擊反應分析



附圖 8-2 VAR-PCATR 模型衝擊反應分析



附圖 8-3 VAR-DFM 模型衝擊反應分析



附圖 8-4 VAR-DFMTR 模型衝擊反應分析

三、預測誤差拆解

預測誤差變異拆解是指將預測誤差的變異拆解成不同衝擊所造成的比率，亦即衡量內生變數波動被特定衝擊所解釋的比例（或貢獻度有多少）。由附表 8-11 VAR-PCA 模組結果可發現，由靜態主成分萃取之民間消費綜合指標中，因子二的對民間消費的變異解釋比例皆較因子一來得高。而納入總體經濟變數之 VAR-PCATR 模型則顯示 cpiyoy 對民間消費的變異解釋較其他變數為高。附表 8-7 及 8-10 為呈現納入動態主成分萃取之民間消費綜合指標之模組。

附表 8-7 VAR-PCA 模型預測誤差拆解

落後期	dlog_pc	dpca_factor1	pca_factor2
1	1.000	0.000	0.000
2	0.752	0.094	0.153
3	0.742	0.090	0.168
4	0.748	0.094	0.158
5	0.763	0.079	0.157
6	0.759	0.079	0.162
7	0.751	0.077	0.172
8	0.755	0.075	0.170
9	0.767	0.069	0.164
10	0.763	0.067	0.169

附表 8-8 VAR-PCATR 模型預測誤差拆解

落後期	dlog_pc	dlog_rgdp	cpiyoy	dpca_factor1	pca_factor2
1	1.000	0.000	0.000	0.000	0.000
2	0.770	0.001	0.217	0.001	0.010
3	0.692	0.065	0.185	0.013	0.044
4	0.567	0.060	0.294	0.039	0.040
5	0.568	0.053	0.261	0.042	0.075
6	0.582	0.053	0.252	0.040	0.073
7	0.560	0.095	0.242	0.039	0.065
8	0.545	0.105	0.235	0.052	0.063
9	0.526	0.097	0.263	0.052	0.062
10	0.540	0.098	0.246	0.049	0.068

附表 8-9 VAR-DFM 模型預測誤差拆解

落後期	dlog_pc	dfm_factor1
1	1.000	0.000
2	0.895	0.105
3	0.880	0.120
4	0.889	0.111
5	0.891	0.109
6	0.878	0.122
7	0.881	0.119
8	0.883	0.117
9	0.886	0.114
10	0.886	0.114

附表 8-10 VAR-DFMTR 模型預測誤差拆解

落後期	dlog_pc	dlog_rgdg	cpiyoy	dfm_factor1
1	1.000	0.000	0.000	0.000
2	0.756	0.002	0.223	0.019
3	0.682	0.108	0.192	0.018
4	0.559	0.088	0.330	0.023
5	0.565	0.086	0.316	0.032
6	0.597	0.083	0.288	0.032
7	0.560	0.111	0.299	0.030
8	0.548	0.117	0.306	0.029
9	0.547	0.115	0.307	0.031
10	0.559	0.113	0.296	0.032

附件九 ARIMA 模型分析(納入 CCI)

附表 9-1 ARIMA 模型分析歷程

Variable	Coefficient	Variable	Coefficient
C	13772.270 ***	C	-0.9356 ***
DRGDP	0.107 ***	DRGDP	0.0966 ***
CPIYOY	-5064.710 **	CPIYOY	-6143.7100 ***
DCCI	3.693	DCCI	-582.2716
BUSCYC_FAC_1	2007.444	BUSCYC_FAC_1	3128.4690
DGT_FAC_1		DGT_FAC_1	-27736.5200 *
DGT_FAC_2		DGT_FAC_2	19065.6700
AR(1)	-1.337 ***	AR(1)	-0.9356 ***
AR(2)	-0.337 ***		
MA(1)	1.000 ***		
SIGMASQ	1.94E+08 **	SIGMASQ	2.53E+08 ***
R-squared	0.900941	R-squared	10094.98
Adjusted R-squared	0.878299	Adjusted R-squared	44807.83
S.E. of regression	15631.48	S.E. of regression	22.64126
Sum squared resid	8.55E+09	Sum squared resid	23.00621
Log likelihood	-485.071	Log likelihood	22.7766
F-statistic	39.79079	F-statistic	1.655633
Akaike info criterion	22.45777	Akaike info criterion	22.64126