

東海大學資訊管理研究所
碩士學位論文

以資料探勘分析台灣社會、經濟與環境的政府效率
之影響—以台灣地方政府開放資料為例

**An Analysis of the Production Indicators of Society,
Economy and Environment with Data Mining - Taking
Taiwan's Open Data as an Example**

指導教授：姜自強 博士

研究生：宋羿勳 撰

中華民國 106 年 7 月

東海大學資訊管理學系碩士學位
考試委員審定書

資訊管理學系研究所 宋昇勳 君所提之論文

以資料探勘分析台灣社會、經濟與環境的政府效率之影響—以台灣地方政府開放資料為例

經本考試委員會審查，符合碩士資格標準。

學位考試委員會 召集人：黃克敏 (簽章)
委員：許愷中
姜百強
楊朝棟
陳淑言

中華民國 106 年 7 月 6 日

謝誌

很高興終於能夠在兩年讀完碩士，雖然短短兩年但在我的求學生涯中卻是最刻苦銘心的一段時間，從大學剛畢業懵懵懂懂的進入東海資管所，開始看論文看期刊，做研究做實驗經過種種的考驗，曾經想就此放棄但因為同學師長的鼓勵，使我能夠繼續完成論文。

承蒙口試委員黃悅民博士、許懷中博士、姜自強博士、楊朝棟博士和陳牧言博士，在口試前細讀我的論文，並在口試相當有限的時間內給我許多寶貴的建議，能使本文更加完善，特此感謝。在研究所論文撰寫期間特別感謝研究室的同學，不厭其煩地幫助我完成論文，也非常感謝林灼榮老師的幫忙，使我在研討會之後能夠即時修正論文的的方向，並且使用正確的模型來導正論文，也非常感謝我的指導老師姜自強教授，讓我在研究所兩年不僅僅學習到研究的內容與知識，也學會了很多做人處事得道理，教導我怎麼樣自己尋求答案，能夠在未來創造自己應有的價值，也讓我學習到在挫折中如何在再次站起來，轉換心情之後重新接受挑戰，感謝學弟妹在口試當天的幫忙可以讓我在口試更能夠準備論文內容，真心感謝你們為我所做的一切。

當然也很感謝我的父母及女友，也因為有你們的支持，我才能夠走到這裡完成學業，我也相信這段日子過了，所有累積的磨練能讓我更茁壯更有抗壓性，在未來職場上能夠更得心應手勇於接受挑戰而不退縮。最後感謝祖師大自在王佛能夠幫助我有形無形的一切，使我能夠順利完成碩士學位。

宋昇勳 謹誌於

東海大學資訊管理學系

2017年07月

論文名稱：以資料探勘分析台灣社會、經濟與環境的政府效率之影響—以台灣地方政府開放資料為例

校所名稱：東海大學資訊管理學系研究所

畢業時間：民國 106 年月

研究生：宋羿勳

指導教授：姜自強

論文摘要：

在開放政府與資料開放的時代，政府開放式資料(Open Government Data)建立中央政府與地方政府的資料，透過政府資料庫方式讓民眾可取得資料並且建立資訊的平台與橋梁。本研究利用政府開放資料，將資料進行分類，分為社會、經濟與環境三大類別並取出 2013 年至 2015 年分為 22 個縣市的資料，並且進行包絡分析法(DEA-SBM)將資料分為良產出、不良產出以及投入，使社會、經濟與環境的經營效率可以量化為經營能力的指標，求出此 3 年間的目標變數。本研究運用包絡分析法(DEA-SBM)的方式做出目標變數運用決策樹模型(Decision Tree)產出重要指標。政府能夠提升效率在社會、經濟與環境的方面更有效的經營台灣的各方面發展以及兼具環境保護的永續經營。再將社會、經濟與環境中的 15 個變數進行決策樹模型(Decision tree)，可以得出 3 個重複重要的指標變數影響台灣的社會變化、經濟發展與經營能力和環境的汙染與變化。運用 k-means 分群將 22 個縣市分為首都、五都以及其他 16 個縣市，了解 2013 年至 2015 年台灣的經濟影響著社會的發展以及環境的變化。

關鍵字：政府開式資料(Open Government Data)、分群(k-means)、包絡分析法(DEA-SBM)、決策樹(Decision tree)、資料探勘(Data Mining)

Title of Thesis: An Analysis of the Production Indicators of Society, Economy and Environment with Data Mining - Taking Taiwan's Open Data as an Example

Name of Institute: Tunghai University, Institute of Information Management

Graduation Time: 2017/07

Student Name: Yi-Hsun Sung

Advisor Name: Tzu-Chiang Chiang

Abstract:

In an age when the concepts of Open Government and Open data are prevalent, the Open Government Data (OGD) provides a platform of information sharing, which enables the public to access to the governmental data. An Analysis on Economic, Social and Environmental OGD This paper, by Data Envelopment Analysis (DEA-SBM), classifies the data available on the platform of OGD in the period of 2013-2015 into 3 categories, finds out the input and output factors, and formulates pointers to measure the efficiency of government's economic, social and environmental policies. This paper finds out the target variable and output pointer by DEA-SBM and Decision-tree Model, respectively. The government can make a balance among economic growth, social development and environmental sustainability, so as to enhance governance efficiency. The 13 pointer variables representing the interaction between economic, social and environment factors help us to look at the social and environmental implication of Taiwan's economic development from 2013 to 2015.

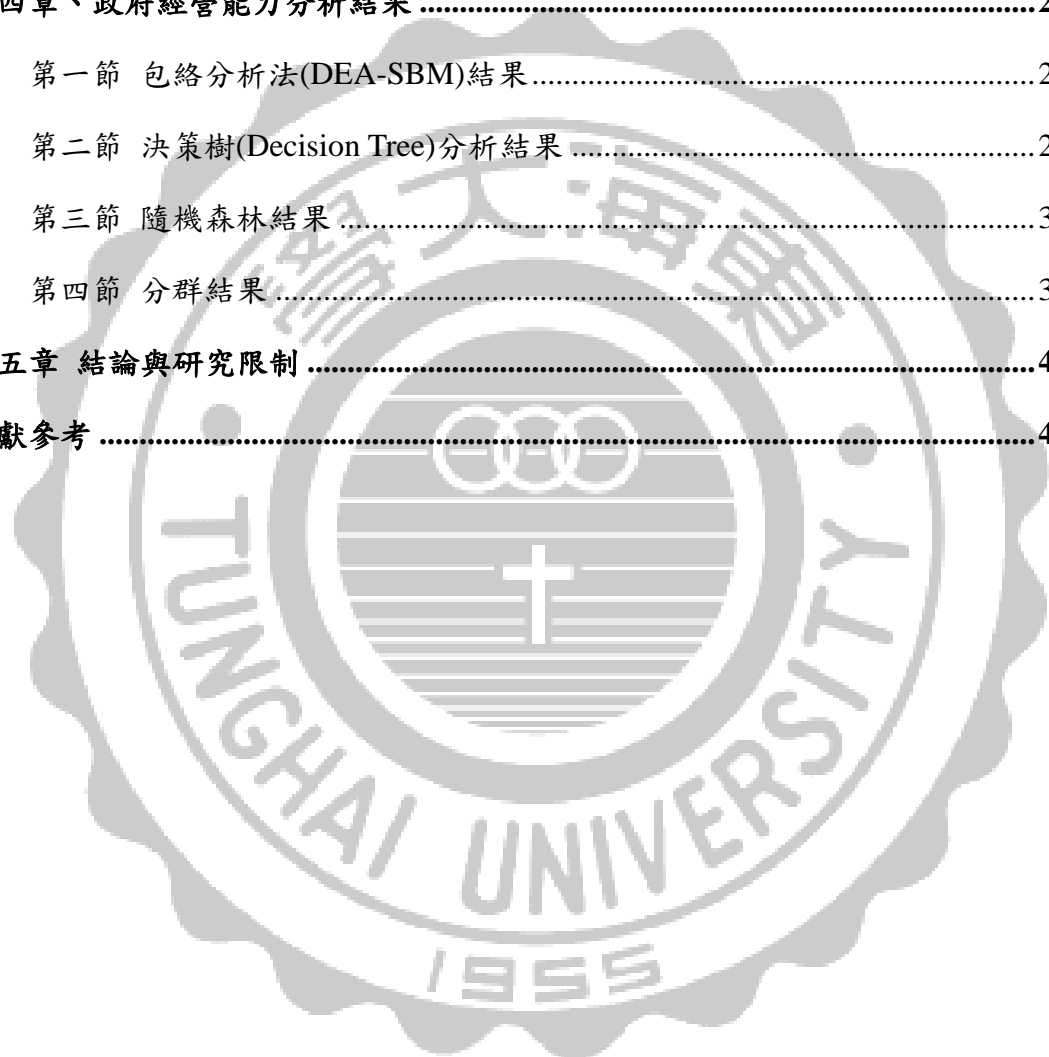
Keyword: Open Government, Scale, Data Envelopment Analysis, Decision tree, Data Mining

目錄

頁次

目錄	III
圖目錄	V
表目錄	VI
第一章 緒論	1
第一節 研究背景與動機	1
第二節 研究目的	2
第三節 論文架構	4
第二章 文獻探討	5
第一節 政府資料與全觀型政府	5
壹、政府開放資料	5
貳、永續發展	6
參、全觀型政府	7
第二節 資料目標—資料包絡分析法	8
壹、包絡分析法(DEA)之定義	8
貳、資料包絡分析法(DIDEA)之基本假設	8
參、資料包絡分析法(DEA)之定義	9
肆、各相關研究對於資料包絡分析法(DEA)的看法	10
第三節 決策樹(Decision tree)	11
壹、使用決策樹(Decision tree)所做之相關研究	11
貳、公式	11
參、決策樹(Decision tree)的修剪	11
伍、決策樹(Decision tree)的分類方法	11
第四節 隨機森林	12

第三章 分析政府經營能力方法	14
第一節 研究流程	14
第二節 研究主要變數說明	15
第三節 數據分析目標	17
第四節 決策樹分析	19
第四章、政府經營能力分析結果	21
第一節 包絡分析法(DEA-SBM)結果.....	21
第二節 決策樹(Decision Tree)分析結果	28
第三節 隨機森林結果	33
第四節 分群結果	34
第五章 結論與研究限制	42
文獻參考	43



圖目錄

	頁次
圖 1-1 本研究架構.....	4
圖 2-1 政府開放資料.....	5
圖 2-2 DEA 相對無效率的決策單位圖.....	8
圖 2-3 隨機森林.....	13
圖 3-1 研究流程.....	14
圖 4-1 2013 年至 2015 年平均 DEA-SBM 模型分數.....	23
圖 4-2 2013 年至 2015 年 DEA-SBM 為 1 的縣市個數.....	25
圖 4-3 2001 年至 2016 年垃圾清運量生產量及資源回收量折線圖.....	29
圖 4-4 2013 年至 2015 年決策樹圖.....	30
圖 4-5 2013 年決策樹圖.....	30
圖 4-6 2014 年決策樹圖.....	31
圖 4-7 2015 年決策樹圖.....	32
圖 4-8 隨機森林圖.....	34
圖 4-9 2013 年各縣市分群.....	35
圖 4-10 2014 年各縣市分群.....	37
圖 4-11 2015 年各縣市分群.....	37
圖 4-12 六都垃圾政策時間關係圖.....	38
圖 4-13 六都事業廢水罰鍰折線圖.....	40
圖 4-14 六都生育補助政策下新生兒個數.....	40
圖 4-15 六都電力售電量與政策折線圖.....	41

表目錄

	頁次
表 3-1 變數說明	15
表 3-2 產出敘述性統計	17
表 4-1 2013 年 DEA-SBM 分數	21
表 4-2 台中市投入產出分數	22
表 4-3 2014 年 DEA-SBM 分數	23
表 4-4 2015 年 DEA-SBM 分數	25
表 4-5 2013 投入產出相關分析	26
表 4-6 2014 投入產出相關分析	27
表 4-7 2015 投入產出相關分析	27
表 4-8 決策樹分支點	28
表 4-9 預測垃圾清運量生產誤差表	29
表 4-10 2014 年垃圾清運量敘述統計表	32
表 4-11 隨機森林	33
表 4-12 新北市需改善之產出表	36
表 4-13 工業園區水支援回收政策	39

第一章 緒論

第一節 研究背景與動機

台灣是一個民主國家，一個民主國家分為水平和垂直的規劃分布，垂直的反映在中央政府的分層上(陳師孟，2006)。而水平則反映在地方政府，中央政府將公共財產分配給地方政府，但卻有分配的優劣之處，2009 年行政院將台灣分為六大都會區推動縣市合併升格目前台灣為六個直轄市台北市、新北市、桃園市、台中市、台南市以及高雄市，呈現了 6 個直轄市 16 個縣市的新格局。六都的架構在於減少行政分層提升市府效率，為了要將公共的經濟支援和其他福利措施能有效的運用在各地方政的權政上，因此本研究期望將各縣市政府的效能提升，並且將資源分配平衡可以使各縣市的社會福利分配、經濟利益以及環境保護相互平衡的新市政。

近年來，台灣政府開放式資料庫，受到國各國政府的影響台灣政府逐漸重視開放式資料，我國政府推動各縣市鄉鎮政府釋出的開放式資料，提供民眾使用提倡政府政策透明提升公民參與公共政策議題，所以經由政府的開放式資料庫，許多民眾藉此可以改善和監督政府施政，政府開放資料分為中央機關和地方政府，依照不同的部門單位給予相關的開放式資料，本文是藉由探討中華民國政府 22 個縣市的開放式資料其中分為社會、經濟與環境三個方面進行數據分析，藉此評估我國社會、經濟與環境的發展以及相互的關係進行研究。

一般而言在營利事業的機構當中，可以明確的算出經營績效，以及投入產出的成本效益。而對政府而言要評估效益就需要透過某些函數進行分析，因為政府有多元化的性質，所以往往在談到政府的效益時都有一定的困難。因此本研究使用資料包絡分析法(Data Envelopment Analysis,DEA)衡量政府的效率，DEA 的運用非常廣泛，包含教育、醫療、農業、銀行、郵局、自來水公司、電力公司及政府機構，在對於政府單一的部門已經有需多的研究，但針對政府的整體效能卻比較

少。因此本文藉由 DEA 評估台灣政府 22 個縣市的政府效率分別以社會、經濟與環境的角度觀察。

社會部分分為：人口數、弱勢族群福利、刑事案件以及就業人口等資料。此部分是為了反映社會的福利和社會成長的指標，各縣市因人數的不同對於社會的變化有相當重要的影響，而弱勢族群的福利包括了老人安養、兒童及少年教養、特殊境遇家庭緊急生活扶助和身心障礙人福利等反映了社會資源的運用和給弱勢族群的保障，刑事案件表現了該縣市的治安狀況，也同時在教育方面以及警政機關的效率上有所反應。就業人口代表著該縣市的勞動人口以及生產力。

經濟部分分為：平均每戶可資配所得、營利事業銷售額、歲出以及平均每人稅賦等等資料。此部分代表著各縣市的經濟因素可支配所得視為消費能力，經濟是透過多次的交易而產生的，如果有消費表示經濟成長也表示生產力提升，而營利事業銷售額也是表示該縣市的工商業發展總銷售額。歲出則表示該地方的預算代表該地方政府被分配到多少中央的財政資源，平均每人稅賦則表示該地方政府的稅務財政收入。

環境部分分為：空氣中懸浮粒子濃度、垃圾清運量、事業廢水罰鍰金額、平均每人環保經費、電力售電量、土地面積以及都市計畫區面積等資料。此環境部分代表該縣市的環境保護，其中空氣中懸浮粒子濃度表示該地區的空气品質，而垃圾清運量和事業廢水罰鍰金額皆表示該地區土地和水的污染，平均每人環保經費表示政府對於該縣市每人所需支付的環境保護經費，電力售電量則代表該地區用電的情形，也可以表示因用電所產生的汙染，土地面積和都市計畫面積代表該地區都市化的情形。

第二節 研究目的

本研究將此三大類政府開放式資料進行資料清理，本研究選取 2013 年至 2015 年的原始資料進行資料清理，為了了解台灣社會的動向各個數值不同的交互關係

影響台灣經濟的成長或衰退，以及環境的變化。本文使用數據包絡分析(DEA)將社會、經濟與環境的投入與產出進行分析，將整體效率做為目標變數，從包絡分析法的數值可以看出 2013 年之後政府經營的指標，推算出最小投入得最大成效，提升市府經濟效益，以及減少環境污染。

將包絡分析法(DEA)找出目標變數之後，使用決策樹對目標進行分析，找出關鍵的資料欄位，並且了解其資料之間的關係與時間和政府政策的關係。視覺化的圖表顯示資料的成果，未來預期的成果包括政府政策實施的改革是否有效，以及工商業投資的經濟效益和造成環境破壞的影響，希望透過數據包絡分析將環境、社會以及經濟的三大構面的產出及投入有更加良好的運用。

本次研究中的資料來源為中華民國統計資訊資料庫，以社會、經濟和環境為指標，於 2013 年至 2015 年的資料為研究對象，並以包絡分析法(DEA-SBM)分析出經營效率佳的年份，為了要進一步了解哪一些因素的影響所以進行決策樹分析，將三年的資料有效的運用，以中華民國主計處的年原始資料最為本自研究的研究對象。

第三節 論文架構

本研究架構先收集政府開放事資料，從中篩選本研究符合社會、經濟與環境的關鍵性指標變數，將這些開放式資料進行彙整，選出 15 個指標變數，在將其變數標準化以便於做包絡分析法(DEA-SBM)分析，DEA-SBM 分析需要將資料分為投入以及產出，在將其變數計算出新的經營指標，做為決策樹的目標變數。接著將 15 個變數以 DEA-SBM 模型產出的目標變數為目標進行決策樹以及隨機森林分析，分類出有效的指標變數，最將分類出來的指標變數，進行視覺化成現觀察時間和各個變數間的關係與變化。在將 22 個縣市做分群探討之間的相關性。

如下圖 1-1 所示：

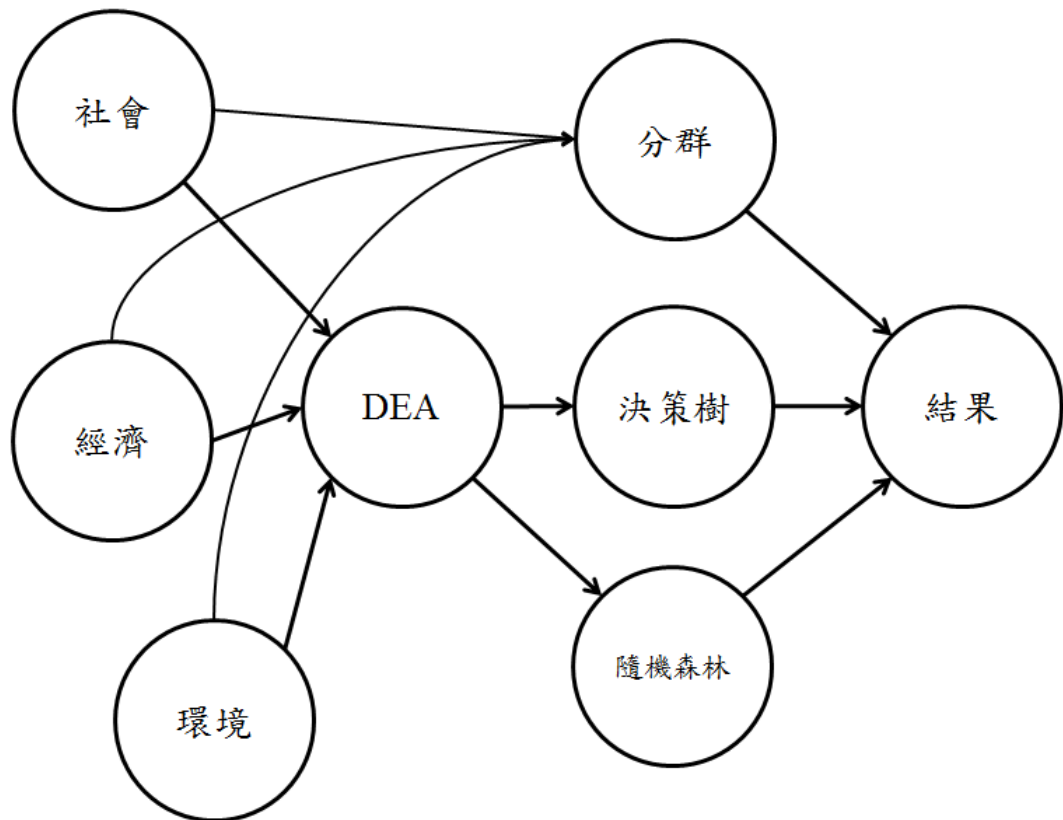


圖 1-1 本研究架構

第二章 文獻探討

第一節 政府資料與全觀型政府

壹、政府開放資料

開放資料(Open Data)指個人、私人機構或是政府組織透過事前挑選或可公開的文建或是資料庫，在不受智慧財產權以及其他管理機制限制的方式，運用資料管理和公眾授權的觀念，將資料公開給社會大眾使用，讓民間、公司或是機構可以使用公共的資料並將資料運用在不同的範疇，進行分析、創業做成資料庫之決策以及解決複雜問題(Gurin,2014)。

政府開放資料(Open Government Data)與開放資料(Open Data)有非常相同的地方，但是發展脈絡與主體則有些不同，學者觀察「政府開放資料」演進過程，發現「開放政府」與「政府開放資料」在網際網路與數位科技的進步的環境下，在加上民間許多組織和團體建意之下使得兩個結不同的該念逐漸匯聚，行成「政府開放資料」的風貌(Yu & Robinson,2012)如圖 2-1 所示。

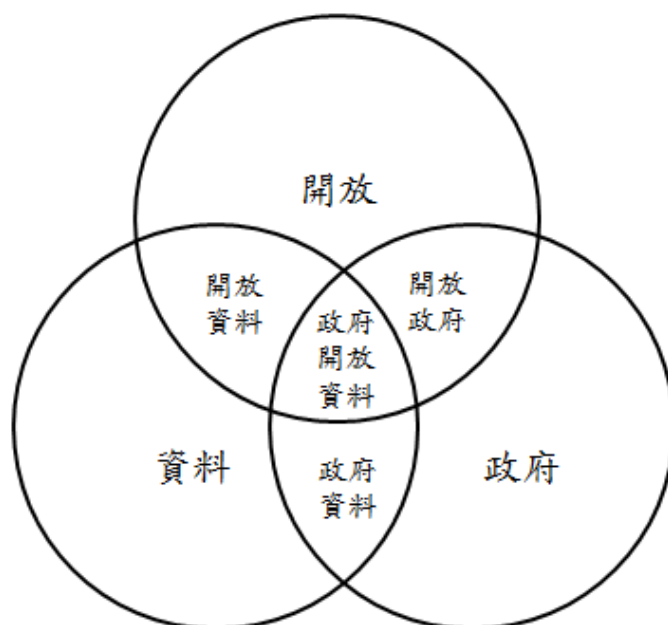


圖 2-1 政府開放資料

資料來源：本研究整理

都會發展對於不同範圍有不同的效果，像是人口、地區、環境、甚至到國家政治、經濟都有高度的密切關係。因此，我們在探討都會發展之時，除了必須追求都會的永續發展外，同時必須加強不同政府機關之間的，創造出所謂的全觀型政府（holistic government）的理念(Perri 6,1980)。

從人口聚居到都市化發展再到都會區，其所代表的是都會發展的歷史過程。由於，都市化發展快速擴張的結果，使得大都市中心區域產生「人口擁擠、生活環境惡化、交通頻繁、活動空間減少、生活品質下降、治安惡化」等若干公共議題；相對於此，在都會區附近的次級都市或衛星城鎮，亦為了提供都會區域發展而無限制的開發地方環境與資源，除了造成資源分配與使用嚴重的不均勻外，更嚴重的破壞了我們所居住的環境。隨著全球化科技與經濟的發展，也使得地球環境資源失衡的問題越顯嚴重。「永續發展」(sustainable development)一詞最早是聯合國環境規劃署（United Nations Environment Programme ,UNEP）、世界環境及發展委員會（The World Commission on Environment and Development,WCEF）與國際自然和保育聯盟（International Union for Conservation of Nature Resources ,IUCN）等三個國際保育組織於1980年出版的「世界自然保育方案」報告提出「滿足當代需求，同時不損及後代子孫，並滿足本身需求的發展」，此為永續發展第一個普遍被接受的定義，也提供了永續發展的一個概念範圍。

貳、永續發展

永續發展的已經運用在其他領域當中，像是永續農業、永續經濟、永續生態、永續社區發展等概念，因為領域之間的專業有差別，所以對於永續發展的意義有不同面向的看法，有些專家學者覺得永續發展需要同時包括社會、經濟與環境等三個構面。Munasinghe（1993）分析各國經濟發展的經驗與過程，發展出一個永續發展的整合性架構（廖俊松，民93），Munasinghe 學者認為一個國家的發展，應該包括三個主要目標：經濟成長、社會公平及生態的穩定與協調等三種不同架構。經濟成長與社會公平目標之間的關聯，在於世代內公平性的問題，即在同一

個時間內，所得分配與社會階級的問題，需要透過消除貧窮與增加就業等方法，來促進世代內所得分配的平均；另一方面，經濟成長與生態穩定協調目標之間的關聯，在於經濟政策對於環境影響評估與外部成本內部化的問題，即必須減少資源利用對於生態環境的不當負荷；最後，社會公平與生態穩定協調二個目標之間的關聯，在於世代之間的公平，意即要留給後人享用和使用資源與環境的權利。簡單的說，永續發展就是建立在經濟、社會以及生態下，來改善我們的生活環境和生活品質。

參、全觀型政府

全觀型政府 (holistic government) 的概念，是由英國學者 Perri 6(1980) 所提出，主張下一世代的政府應該為全觀型政府、預防性政府、改變文化以及結果取向的政府(彭錦鵬，民 94)。全觀型政府體制係以層級節制、功能分化的基本原則，這種功能性組織所產生的缺失稱為「功能性裂解治理」(functionally ,fragmented & governance,6 et al, 2002)，已經無法有效應付許多跨功能邊界的問題，以及提供民眾迫切且實際的需求與服務。

Perri 6 認為 1980 年代以來新公共管理的改革，主要在於追求往往受制經濟與效率等短期效益，使得政府組織更趨於功能分化與專業分工(Perri 6 ,et al, 2002)。而相對的對於實際的服務功能則未能加以同步關心，無法解決民眾切身相關的許多問題，如失業率、犯罪率、經濟衰退或環境保護等議題，無法滿足民眾的期望。全觀型政府正是為了解決上述「功能性裂解治理」所提出的因應對策，Perri 6 認為全觀型政府是過資訊科技所促成的唯一窗口服務機制，而這些唯一窗口是將人類的思考作為生活的各個事件。也就是政府業務的分工，應該從民眾的生活事件角度來加以重組。如何使政府的功能進行整合的工作，提供民眾一個整合型的服務，而不是奔波於各個部會、單位，這就是全觀型政府所追求的目標。

因此結合全觀型治理與政府開放資料監督政府的永續經營能力，正是本研究的目標。

第二節 資料目標—資料包絡分析法

壹、包絡分析法(DEA)之定義

運用 DEA 來評估政府效率主要事由 Charnes, Cooper, & Rhodes(1978)指出採用(Charnes, Cooper, & Rhodes, CCR)將 Farrell (1957)的效率評估觀念推廣，並建立一般化之數學規劃模式，以衡量在固定規模報酬下，多項投入與多項產出之生產效率。BCC 模式：Banker(1984)指出將 CCR 模式修正為變動規模報酬的假設下衡量決策單位之相對效率。

資料包絡分析法(DEA)是一種引進生產函數概念的效率評估方法，此法源於(Farrell, 1957)

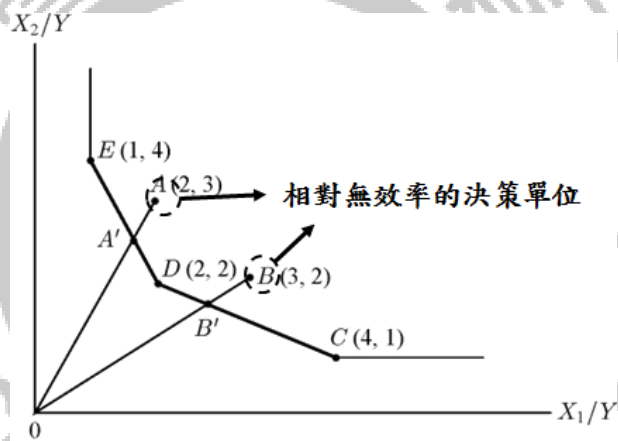


圖 2-2 DEA 相對無效率的決策單位圖

貳、資料包絡分析法(DIDEA)之基本假設

本研究使用資料包絡分析法來做為目標變數的分析方法，但是使用資料包絡分析法應受到相關文獻的支持，所以本研究參考了許多學者的研究，應用資料包絡分析法 (DEA) 探討職業球隊投入成本與績效之關係(吳翠治、齊德彰、李科翰，民 100)，資料包絡分析本身大多用於績效方面的分析，而此研究利用了資料包絡分析來職業球隊投入成本與績效。

資料包絡分析法之整合式規模報酬關係—以臺灣道路維護為例(李綱，民 101)，台灣的道路維護一直是大多人之間的詬病，此研究以台灣道路為研究主

題，利用了資料包絡分析來分析出台灣道路的維護效率來提供政策上的參考，是在研究成果也屬於研究限制上的建議，但是也說明了台灣道路維護上經費成本上的狀況，此研究的目的為透過資料包絡分析法，瞭解各基層道路養護工務段之執行績效，進而針對各項資源之投入及產出成果，做最有效之比較分析，並引為組織再造檢討，預算資源分配等公路養護政策之參考，而研究成果方面有二，其一是在公路養護的資料包絡分析結果中，投入、產出項目應改要有變動，確認傳統 DEA 分析採用之預算項目重複計算部分投入、產出項，所以建議應予剔除，這屬於研究限制方面，而其二是各地區之鋪面瀝青修復經費，其單價並無明顯差異，代表以經費表達各工務段努力之產出程度應無地域差別，此研究以鋪面瀝青修復經費來估算出各地的維修狀況沒有很大的差別。

參、資料包絡分析法(DEA)之定義

資料包絡分析法(DEA)主要是以經濟學中柏拉圖最適境界 (Pareto Optimality) 做為主要觀念來判斷效率，並將所有作為主要資料的決策單位 (Decision Making Unit, DMU) 中的資料投入和產出(徐志明、彭育秀，民 102)，本研究是以經營能力作為投入，獲利能力作為產出，並以因此求得各公司財報變數中，是否有可以明顯的影響利用資料包絡分析法(DEA)所算出的數值。

資料包絡分析法(DEA)投入產出的觀念，起始於於 Farrel (1957) 指出所提之非參數邊界分析，Charnes ,et al .(1978) 指出 利用 1957 年 Farrell 學者的確定性非參數法 (Deterministic Non-parametric Approach)，並納入數學規劃 (Mathematical Programming) 技巧，在所有受評估的決策單位中建構效率前緣 (Efficiency Frontier)，位於邊界上的 DMU 即為最有效率的，而未落在邊界上的 DMU 則為無效率，且 此效率為一種相對效率的概念。(劉文斌、左晉璋與彭克仲，民 99)。

肆、各相關研究對於資料包絡分析法(DEA)的看法

資料包絡分析一直是財務分析方面常常利用到的分析手法，而所以也就造成了有許多學者對於資料包絡分析有多樣的想法。

以下是個學者對於資料包絡分析的看法：

蔡榮發、黃文中(民 93)資料包絡分析法是績效評估的工具之一，其目的在於希望透過決策單位元 (Decision Making Units, DMU) 衡量多項產出與多項投入間相對的表現，並透過權重計算各 DMU 的產出與投入比值。

(胡志堅、黎漢林，民 93) 資料包絡分析法可以處理多項投入、產出之評估問題，其評估效率之結果係為一綜合指標，易於描述接受評估企業之績效表現。

劉祥熹、朱珮宏(民 93)資料包絡分析的投入導向模式為對投入量可加以控制者，即是將現有產出值固定來計算投入要素可以縮減之部份，如對產出量可加以控制者，即可採用產出導向樣式，亦將現有投入量固定來計算產出要素可以擴張之部份，正確導向模式之選擇，端視決策單位對投入產出要素之控制能力而定。

資料包絡分析法(DEA) 正可解決投入或產出不易量化的廠商經營效率分析

伍、目標函數：

目標是找出一組伴隨每種產出的繫數 u 和一組伴隨每種投入的繫數 v ，從而給被評估的服務單位最高的可能效率。

$$\max E_e = \frac{u_1 O_{1e} + u_2 O_{2e} + \dots + u_M O_{Me}}{v_1 I_{1e} + v_2 I_{2e} + \dots + v_N I_{Ne}}$$

第三節 決策樹(Decision tree)

壹、使用決策樹(Decision tree)所做之相關研究

決策樹演算法其主要功能是使用已分類之資料數據來建立一個樹狀結構。利用決策樹模型來做樣本外的預測。決策樹由一個訓練集合(Training Set)產生，當資料不完整或是含有雜訊時，建構決策樹時會產生過適化的問題。若樣本的屬性過多，決策樹演算法容易選用到和種類不相關的屬性，因此決策樹建構完成後還需要做適當的修剪(吳植森，民 99)。

貳、公式

Entropy = 系統的凌亂程度，使用演算法 ID3, C4.5 和 C5.0 生成樹算法使用熵。這一度量是基於信息學理論中熵的概念。

1. ID3 算法 (Iterative Dichotomiser 3) 奧卡姆剃刀闡述了一個信息熵的概念：

$$I_E = - \sum_{j=1}^m f(i,j) \log_2 f(i,j)$$

參、決策樹(Decision tree)的修剪

策樹計算完成之後，可以利用的修剪 (tree-pruning)的步驟來讓決策樹的大小進行改變，因為如過決策樹太大，就會很造成過度學習 (overfitting) 的情況發生，所以修剪決策樹是有助於決策樹類提升推算到算其他資料分類的能力，決策樹是用上的優點可以很清楚的知道資料分類的依據和規則，也可以很容易處理連續性或特定類別的資料，並不會有過多的計算過程就可以進行決策樹的分類，當然決策樹還是有缺點的，樣本數目太少或這是所分類出的類別過多時容易造成資料上的錯亂。

伍、決策樹(Decision tree)的分類方法

決策樹(Decision tree)的分類，是依照投入的資料中已知的類別屬性來進行資

料分類，而決策樹的分類是建立可以讓研究者更加了解各種資料的特徵與屬性，而決策樹也可以依造所分的類別來進行預測。

本研究使用 R 語言的 rpart 決策樹演算法，將資料分類出一個數狀結構圖型，運用決策樹可以將本研究 15 個變數中取出重要的分類因子，將其分類因子產生出決定目標變數的關鍵數值進行區分，使本研究的 15 個分類變數產生鑑別和區分程度。

第四節 隨機森林

隨機森林可以用於排序出研究中的所有變數的對於目標變數影響程度的排名，可以讓研究者知道較為值得進行分析的變數為何，或者去驗證其他分析結果的準確性。

以下是各學者對於利用隨機森林進行分析的相關研究：

基於隨機森林和轉導推理的特徵提取方法(邱一卉、米紅，民 99)

此研究的研究目的為利用隨機森林的相似性矩陣對訓練樣本和測試數據進行轉導推理，在此基礎上進行多維尺度變換得到降維後的低維特徵，此研究是利用隨機森林進行特徵選取，來分析出更好的特徵提取方法，研究結果上也成功的提出此研究的研究結果為，利用隨機森林的相似性矩陣對訓練樣本和測試數據進行轉導推理，在此基礎上進行多維尺度變換得到降維後的低維特徵，使用該方法充分利用帶標籤訓練樣本全部信息和未標籤的測試數據的屬性，不僅降低了訓練器的輸入維數，更突出地表達了數據的空間分佈，從而較好地改善模型的泛化誤差，是一種行之有效而又簡便的特徵提取方法。

隨機森林於政府開放資料尚未開端，本研究是將資料 15 個開放式資料數據以包絡分析法為目標變數，將隨機森林的 %IncMSE (increase in MSE) 作為主要影響因數指標，若 %IncMSE 越大表示影響的 DEA 值越高。

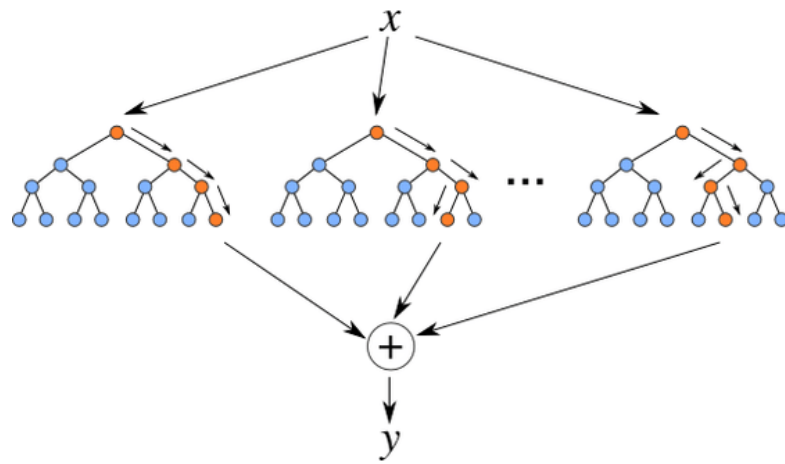


圖 2-3 隨機森林



第三章 分析政府經營能力方法

第一節 研究流程

本研究流程是運用地方政府的開放資料，分別以社會、經濟與環境的議題下選取 2013 年至 2015 年 15 個變數，進行包絡分析法將資料分為投入以及產出，求得政府效率以此效率為目標變數，再運用決策數和隨機森林模型求得重要的指標，影響政府效率的指標，將 22 個縣市進行分群觀察各縣市之間的關聯性，最後分析研究結果，使效率較不佳的城市向效率佳的城市學習。

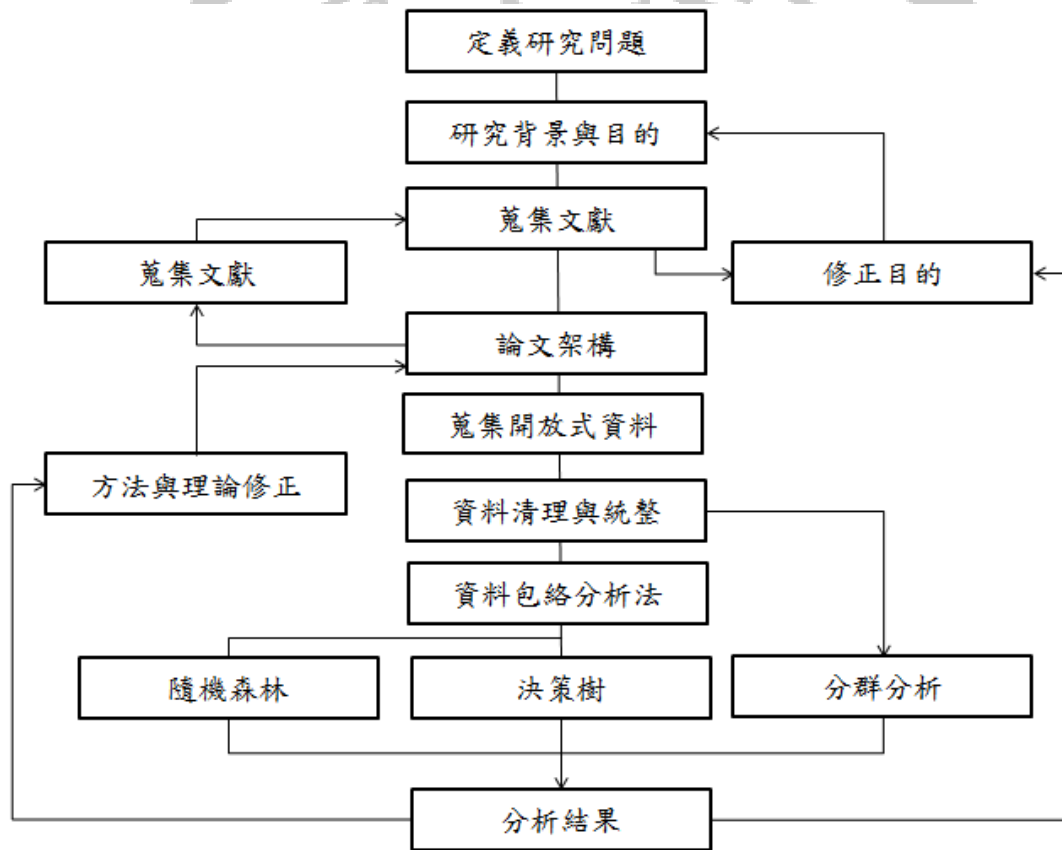


圖 3-1 研究流程

第二節 研究主要變數說明

本次的根據(張鈺琪, 民 100)計算政府效率採用歲出、人口數、身心障礙人數、犯罪率等等資料。本研究分為社會、經濟與環境的三大方面去檢視地方政府的生產效率，以下為 15 個變數的欄位說明：

表 3-1 變數說明

構面	類別	定義
社會	人口數	指在某地區設有戶籍之中華民國國民，於統計標準日不論其是否住在戶內，均為該地區之人口數。
	弱勢族群福利	老人長期照顧、安養機構實際進住人數+兒童及少年安置及教養機構現有收容人數+特殊境遇家庭緊急生活扶助人次+身心障礙人數(人次)
	刑事案件	刑案發生件數係指各警察機關受理民眾告訴、告發、自首或於勤務中發現之罪犯件數。
	就業人口	資料標準週內年滿十五歲從事有酬工作者或從事十五小時以上之無酬家屬工作者。(千人)
經濟	平均每戶可支配所得	平均每戶可支配所得(元)
	營利事業銷售額	營利事業登記之公司行號依規定申報或核定之銷售額。(千元)
	歲出	依預算法第六條規定稱歲出者，謂一個會計年度之一切支出(百萬元)
	平均每人稅賦	平均每一人之稅捐實徵淨額。(不包括關稅、礦區稅及公賣利益) 公式： $(\text{稅捐實徵淨額} / \text{年度中人口數})(\text{元})$

環境	空氣中總懸浮微粒濃度	浮游於空氣中之微粒（包括 10 微米以上浮游粒子）。各地區濃度為該地區一般人工監測站各測站濃度之算數平均值。各測站濃度為該測站各連續 24 小時監測值之幾何平均值；81 年以前每年測定月數未達 8 個月，82 年以後每年測定次數未達 16 次之測站及其監測資料不納入計算。
	垃圾清運量	由廢棄物清理執行機關或公私處所自行或委託清運至垃圾清運量處理場(廠)之垃圾清運量(公噸)
	事業廢水罰鍰金額	指環保單位執行污染(包括畜牧廢水、其他事業廢水)稽查，對違反法令規定之受查對象，給予處分罰金之總額。
	平均每人環保經費	平均每人環保經費：行政院環保署提供之各級環保單位歲出預算決算統計資料，主要統計各類別預算及決算總計經費，包括一般行政、綜合計畫、研究及發展、空氣污染防治、噪音振動防制等計畫。
	電力售電量	係指台電公司售予包用電力、低壓電力、高壓電力、特高壓電力等用戶之售電量。(百萬度)
	土地面積	土地面積係指行政區域之土地面積，包含海埔新生地及附屬島嶼。土地面積：(平方公里)
	都市計畫區面積	依法完成法定程序的都市計畫區域範圍之總面積。(平方公里)

資料來源：本研究整理

第三節 數據分析目標

本文將政府開放資料，自開放式資料主計處下載原始資料，將資料進行分類，分類為三類：社會、經濟與環境三類，在將資料進行資料清理，選取本次研究所需要的部分資料，選取 2013 年至 2015 年各縣市資料運用，將資料分為投入、產出及不良產出如(表 3-2~3-4)所示。

表 3-2 產出敘述性統計

2013 年 -2015 年	產出		
	弱勢族群福利(人)	營利事業銷售額(千元)	平均每戶可支配所得(元)
最大值	171,193.000	12,810,834,067.000	1,314,031.000
最小值	451.000	4,318,736.000	636,162.000
平均數	54,580.470	1,786,163,440.848	894,040.667
標準差	47,642.092	2,745,399,916.730	163,718.021

表 3-3 不良產出敘述性統計

2013 年 -2015 年	不良產出			
	空氣中總懸浮微粒濃度	垃圾清運量(公噸)	事業廢水罰鍰金額	刑事案件(件)
最大值	88.660	427,356.000	149,962.000	54,387.000
最小值	29.700	2,112.000	19.000	58.000
平均數	56.628	148624.364	13,484.621	13,379.167
標準差	13.045	130,258.295	22,582.567	13,909.368

表 3-4 投入敘述性統計

2013 年 -2015 年	投入		
	就業人口(千人)	歲出(百萬元)	電力售電量(百萬度)
最大值	1,945.000	169,580.240	22,535.000
最小值	3.000	3,000.850	26.000
平均數	504.697	45,293.541	6,551.621
標準差	520.909	47,184.288	7,314.824

本次研究將三年的資料分別做 DEA-SBM 模型分析，本次研究因為有環境中不良因子的效率問題，所以以傳統的 DEA-SBM 模式當中投入越少、產出的越多，效率會越高，但在本自研究當中因為環境效率的問題，可能會有爭議。例如：空氣懸浮粒子濃度、垃圾清運量以及事業廢水罰鍰金額等非需求產出，若列入產出變數，則產出越多效率越高的假設並不成立。本次研究採用 Cooper 等人(2000)所提出的 SBM 修正模式，將非需求產出納入 SBM 模式中。

非需求產出模式可分為兩種，一種是 Bad Outputs(不良產出)模式，此模式將產出項目區分為好的(需求)產出及不良的(非需求)產出，兩者之間沒有相互關係，另一種是 Non Separable(不可切割)模式，此模式的好的(需求)產出及不良的(非需求)產出彼此不可切割。本次的好的(需求)產出及不良的(非需求)產出是可以切割的因此選擇 Bad Outputs(不良產出)模式。

不良的產出模式評估系統內有 n 個決策單位(SBM)，每個 SBM 都有三種投入及產出要素：投入、好的產出及不良產出，分別以三個向量表示： $X \in R^m, Y^g \in R^{s1}, Y^b \in R^{s2}$ ，將 X, Y^g, Y^b 三個向量矩陣定義為： $X=[x_1, \dots, x_n] \in R^{m \times n}, Y^g=[y_1^g, \dots, y_n^g] \in R^{s1 \times n}, Y^b=[y_1^b, \dots, y_n^b] \in R^{s2 \times n}$ ，並假設 $X > 0, Y^g > 0, Y^b > 0$ ，其生產可能集合可定義為：

$$P = \{(x, y^g, y^b) | x \geq X\lambda, y^g \leq y^g\lambda, y^b \leq y^b\lambda, \lambda \geq 0\}, \quad (3.1)$$

根據上式(3.1)的定義，可將 SBM 模式修正如下式：

$$\begin{aligned} \min p^* &= \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^{\text{投入}}}{x_{io}}}{1 + \frac{1}{s_1 + s_2} \left(\sum_{r=1}^{s_1} \frac{s_r^{\text{產出}}}{y_{ro}^g} + \sum_{r=1}^{s_2} \frac{s_r^{\text{不良產出}}}{y_{ro}^b} \right)} \\ \text{s.t.} \quad x_o &= X\lambda + s^{\text{投入}} \\ y_o^g &= Y^g\lambda - s^{\text{產出}} \\ y_o^b &= Y^b\lambda + s^{\text{不良產出}} \\ \lambda, s^-, s^g, s^b &\geq 0 \end{aligned} \quad (3.2)$$

在上式(3.2)中， $s^{\text{投入}}$ 為投入過剩數量， $s^{\text{不良產出}}$ 為不良過剩數量，這兩項應予以縮減； $s^{\text{產出}}$ 為好的產出之短缺數量，應予以增。 $0 < p^* \leq 1$ ，若令上述線性規劃式之最適值 $(\lambda^*, s^{\text{投入}*}, s^{\text{不良產出}*}, s^{\text{產出}*})$ ，則當 $s^{\text{投入}*} = 0, s^{\text{不良產出}*} = 0, s^{\text{產出}*} = 0$ 時， $p^* = 1$ ，代表 DMU_o 在不良產出模式中具有效率。

假如 $p^* < 1$ ，代表，代表 DMU_o 在不良產出模式中為無效率 SBM，要成為有效率 SBM，必須縮減投入及不良產出的過剩量，並擴增好的產出的短缺數量。式(3.3)為不良產出模式，不具有效率的 SBM 調整方式，其中 $(\widehat{x}_0, \widehat{y}_0^{\text{產出}}, \widehat{y}_0^{\text{不良產出}})$ 代表效率邊界之投射點。

$$\begin{aligned} \widehat{x}_0 &= x_0 - s^{\text{投入}*} \\ \widehat{y}_0^{\text{產出}} &= y_0^{\text{產出}} + s^{\text{產出}*} \\ \widehat{y}_0^{\text{不良產出}} &= y_0^{\text{不良產出}} - s^{\text{不良產出}*} \end{aligned} \quad (3.3)$$

運用 SBM 模型可以產出本研究政府效率指標。

第四節 決策樹分析

生產理論中有關技術效率 Farrel(1957)乃奠基在生產技術部改變之下，衡量的 SBM 之產出投入離生產邊界(production frontier)之程度；並以估計出之效率指數作為評估效益。而本研究的政府效益之評估最為決策樹的目標變數，使用生產效率

的當作指標評估，藉由決策樹評估 2013 年至 2015 年能夠影響政府效能的重要指標。

運用 R 的 rpart 套件，rpart 是 CART 分類回歸樹，分類迴歸樹(CART, Classification and Regression Tree)由 Brieman 在 1984 年提出 CART，以反覆運算的方式，由根部開始反覆建立二元分支樹，直到樹節點中的同質性達到某個標準，或觸發反覆運算終止條件為止。運用決策樹將 SBM 的效率指數作為目標變數。求得模型的重要指標以及模型的準確率。



第四章、政府經營能力分析結果

第一節 包絡分析法(DEA-SBM)結果

本文將三年的資料分開做 SBM 模型，將好的產出弱勢族群福利、刑事案件、營利事業銷售額、平均每戶可支配所得；不良產出空氣中總懸浮微粒濃度、垃圾清運量以及事業廢水罰鍰金額；投入就業人口、歲出、電力售電量進行 DEA-SBM 分析結果如下表：

表 4-1 2013 年 DEA-SBM 分數

No.	City 縣市	Score 分數	Rank 排名	1/Score
2	臺北市	1	1	1
12	雲林縣	1	1	1
13	嘉義縣	1	1	1
15	臺東縣	1	1	1
16	花蓮縣	1	1	1
19	新竹市	1	1	1
20	嘉義市	1	1	1
21	金門縣	1	1	1
22	連江縣	1	1	1
10	彰化縣	0.9819397	10	1.0183924
14	屏東縣	0.9802676	11	1.0201296
18	基隆市	0.9745176	12	1.0261487
1	新北市	0.9627735	13	1.0386659
3	桃園市	0.9575404	14	1.0443423
7	宜蘭縣	0.9537752	15	1.0484651

11	南投縣	0.9311214	16	1.0739738
8	新竹縣	0.8770389	17	1.1402003
5	臺南市	0.8515812	18	1.1742861
6	高雄市	0.8185517	19	1.22167
17	澎湖縣	0.7920523	20	1.2625429
9	苗栗縣	0.7795309	21	1.2828229
4	臺中市	0.7170804	22	1.3945437

從表 3-4 可以看出來 2013 年時，有 9 個縣市的 DEA-SBM 指標為 1 最高的效率，而其餘的 13 個縣市可以改善他們的投入產出達到好的效率，台中市的效率指標是最低的，從表 3-5 可以看出來台中市的效率是最低的，弱勢族群福利、營利事業銷售額、平均每戶可支配所得都比較缺乏，在投入就業人口、歲出與電力售電量之下台中市的產出比較不足，不良產出的事業廢水罰鍰金額也需要減少。(其他縣市如附錄)

表 4-2 台中市投入產出分數

SBM	1/DEA_Score			
I/O(投入/產出)	Data(原始資料)	Projection(相對值)	Difference(差距)	差距百分比%
臺中市	1.394543748			
就業人口	1,276	1,276	0	0.00%
歲出	109,538.02	109,538.02	0	0.00%
電力售電量	17,505	17,505	0	0.00%
弱勢族群福利	119,908	167,216.9517	47,308.95169	39.45%
營利事業銷售額	3,687,509,962	5,142,393,962	1,454,884,000	39.45%
平均每戶可支配所得	928,938	7,589,818.795	6,660,880.795	717.04%

空氣中總懸浮微粒濃度	78.52	546.4220774	467.9020774	595.90%
垃圾清運量	378,209	457,434.2076	79,225.20762	20.95%
事業廢水罰鍰金額	23,030	16,200.98869	-6,829.011307	-29.65%

將 2014 年的資料一樣使用 SBM 模型將資料產出效率分數，不一樣的是去年 2013 年效率為 1 的縣市有 9 個而 2014 年多一個基隆市表示台灣由政府經營效率有逐年上升得趨勢。如下圖 2013 年至 2015 年三年的平均分數，三年都是最佳分數 1 的縣市為藍色顯示，而綠色的縣市以及藍色的縣市是在平均以上的縣市。

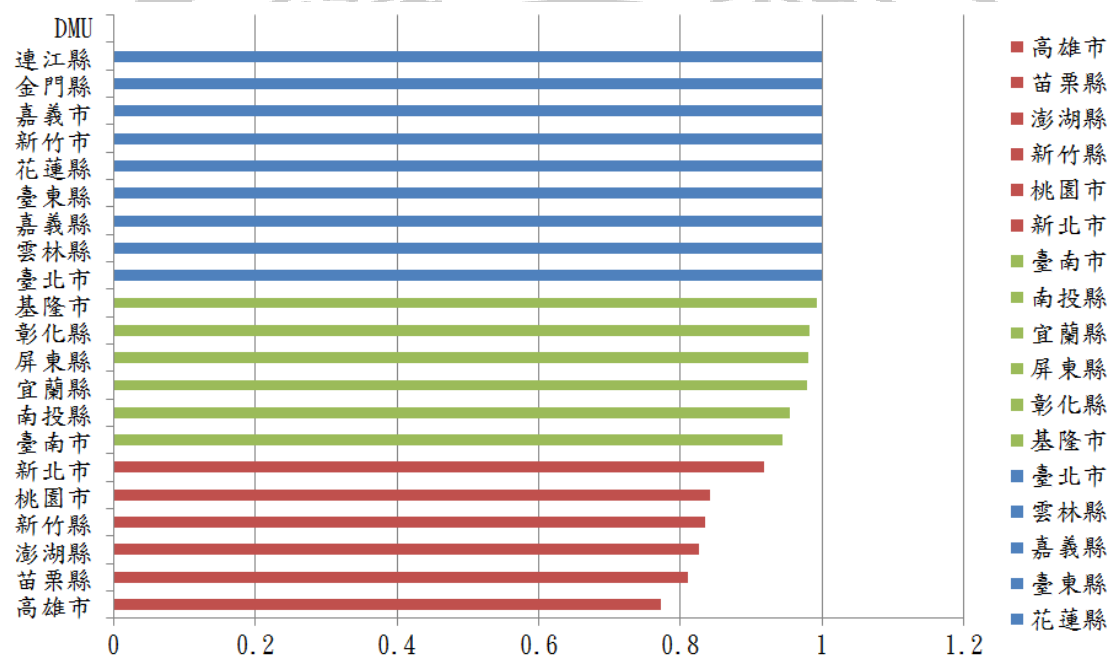


圖 4-1 2013 年至 2015 年平均 DEA-SBM 模型分數

表 4-3 2014 年 DEA-SBM 分數

No.	City 縣市	Score 分數	Rank 排名	1/Score
2	臺北市	1	1	1
12	雲林縣	1	1	1
13	嘉義縣	1	1	1
15	臺東縣	1	1	1

16	花蓮縣	1	1	1
18	基隆市	1	1	1
19	新竹市	1	1	1
20	嘉義市	1	1	1
21	金門縣	1	1	1
22	連江縣	1	1	1
14	屏東縣	0.9839405	11	1.0163216
5	臺南市	0.9817848	12	1.0185531
7	宜蘭縣	0.9807762	13	1.0196006
10	彰化縣	0.9616386	14	1.0398917
11	南投縣	0.9330617	15	1.0717405
1	新北市	0.8848463	16	1.1301397
3	桃園市	0.8835447	17	1.1318046
17	澎湖縣	0.8279055	18	1.2078674
9	苗栗縣	0.8076365	19	1.2381807
8	新竹縣	0.7724447	20	1.2945911
4	臺中市	0.7535297	21	1.3270877
6	高雄市	0.7480282	22	1.3368479

2014 年比 2013 年的縣市多一個為基隆市，2015 年的 DEA-SBM 分數有較多的縣市是標準縣市分數 1 共有 14 個縣市，比 2014 年多了台南市、宜蘭縣、彰化縣和南投縣，由下圖可觀察出每年政府的效率逐年漸好，到 2015 年有 14 個縣市的政府效率為最佳狀態。

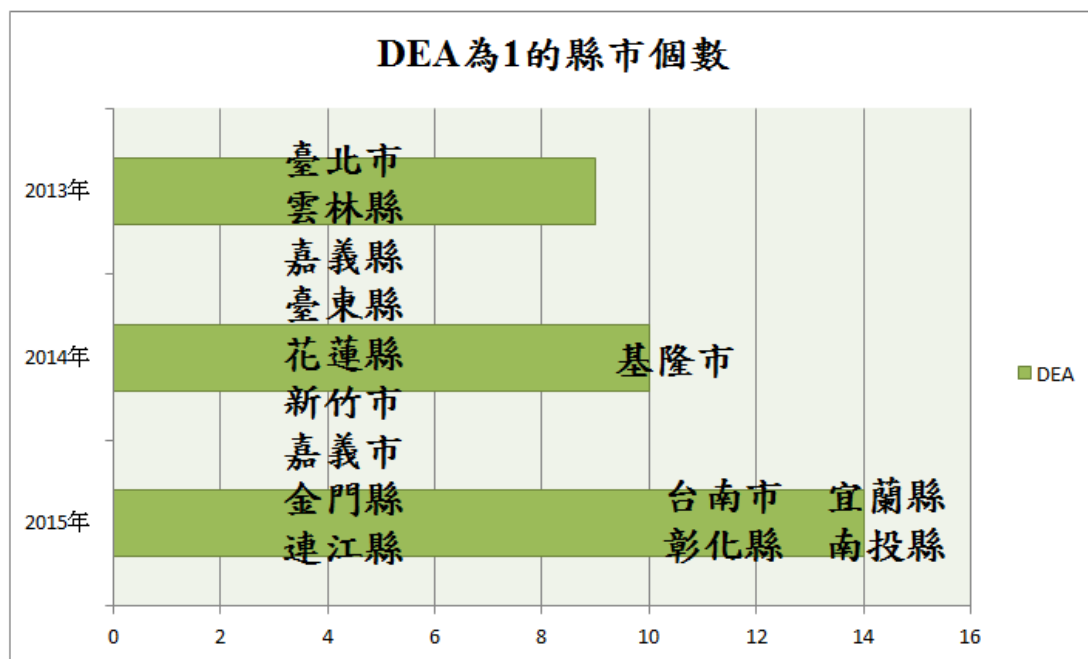


圖 4-2 2013 年至 2015 年 DEA-SBM 為 1 的縣市個數

表 4-4 2015 年 DEA-SBM 分數

No.	City 縣市	Score 分數	Rank 排名	1/Score
2	臺北市	1	1	1
5	臺南市	1	1	1
7	宜蘭縣	1	1	1
10	彰化縣	1	1	1
11	南投縣	1	1	1
12	雲林縣	1	1	1
13	嘉義縣	1	1	1
15	臺東縣	1	1	1
16	花蓮縣	1	1	1
18	基隆市	1	1	1
19	新竹市	1	1	1
20	嘉義市	1	1	1

21	金門縣	1	1	1
22	連江縣	1	1	1
14	屏東縣	0.9780446	15	1.0224482
1	新北市	0.9038394	16	1.1063913
17	澎湖縣	0.8563788	17	1.1677076
8	新竹縣	0.8526078	18	1.1728722
9	苗栗縣	0.846423	19	1.1814424
6	高雄市	0.7511832	20	1.3312331
4	臺中市	0.7303366	21	1.3692317
3	桃園市	0.6859151	22	1.4579063

本文將三年的投入以及產出進行相關分析，投入：就業人口、歲出及電力售電量；產出：弱勢族群福利、刑事案件、營利事業銷售額、平均每戶支配所得、空氣中懸浮粒子濃度，為了要驗證本次 DEA-SBM 模型的效率所以將投入與產出進行相關分析，分析結果如下投入與產出的相關係數皆為正的表示本次的 DEA-SBM 效率模型驗證出示有效的。

表 4-5 2013 投入產出相關分析

Correlation

	就業人口	歲出	電力售電量
弱勢族群福利	0.9834586	0.9516295	0.7353816
刑事案件	0.9817079	0.9713729	0.6722868
營利事業銷售額	0.7332882	0.88404	0.5224317
平均每戶可支配所得	0.2384603	0.3505443	0.2902469
空氣中總懸浮微粒濃度	0.2230551	0.157402	0.3110857
垃圾清運量	0.9588915	0.8757333	0.8591033

事業廢水罰鍰金額	0.5453328	0.4864767	0.7304795
----------	-----------	-----------	-----------

表 4-6 2014 投入產出相關分析

Correlation

	就業人口	歲出	電力售電量
弱勢族群福利	0.9856022	0.9555294	0.7316672
刑事案件	0.975036	0.9841015	0.6779537
營利事業銷售額	0.7336414	0.876446	0.5112754
平均每戶可支配所得	0.3269372	0.4183818	0.3949147
空氣中總懸浮微粒濃度	0.1657047	0.0649798	0.3683987
垃圾清運量	0.9554692	0.8789468	0.866556
事業廢水罰鍰金額	0.7034672	0.530863	0.805354

表 4-7 2015 投入產出相關分析

Correlation

	就業人口	歲出	電力售電量
弱勢族群福利	0.9838802	0.9615803	0.7331352
刑事案件	0.9747361	0.9748189	0.6670553
營利事業銷售額	0.7308049	0.8572635	0.5020175
平均每戶可支配所得	0.3930483	0.5002897	0.3902779
空氣中總懸浮微粒濃度	0.1175177	0.0018733	0.3764371
垃圾清運量	0.935856	0.905125	0.8786199
事業廢水罰鍰金額	0.6672337	0.5580792	0.6871659

第二節 決策樹(Decision Tree)分析結果

表 4-8 決策樹分支點

2013 年	2014 年	2015 年	2013 年~2015 年
事業廢水罰鍰金額	垃圾清運量	垃圾清運量	垃圾清運量
土地面積	都市計畫面積	都市計畫面積	都市計畫面積
都市計畫面積	事業廢水罰鍰金額	事業廢水罰鍰金額	刑事案件
人口數	人口數	人口數	電力銷售量
平均每戶可支配所得			營利事業銷售額
			空氣中懸浮粒子濃度

本研究將 2013 年至 2015 年分開使用決策樹分析，並且也全部分析一次結果可以看出其中事業廢水罰鍰金額、垃圾清運量、都市計畫面積和人口數都有高比例的重複影響政府效率影響。

各縣市業廢水罰鍰金額根據違反水污染防治法罰鍰額度裁罰準則修正草案有效遏止業者排放廢水超標、任意繞流排放、非法稀釋、未具備足夠之功能及設備，以及未維持廢（污）水（前）處理設施正常操作等違規行為都會受政府罰鍰，在工業發展過會使經濟或是政府各方面效能提升，但是在環境保護的面向例如廢水的排放也因為工業的發展而造成環境汙染，因此從決策樹的樹分支中可以看出事業廢水的汙染會隨著城市發展而產生環境汙染。

垃圾清運量在決策樹中也是占很重要的分支 2014、2015 年的決策樹都是在最上層的分支，2014、2015 年的垃圾清運量約大於 313,700(公噸)為政府效能正向成長，但是因為政府提倡環保回收所以從 2001 年至今垃圾清運量的生產量有隨著資源回收上升而下降，如下圖所示。

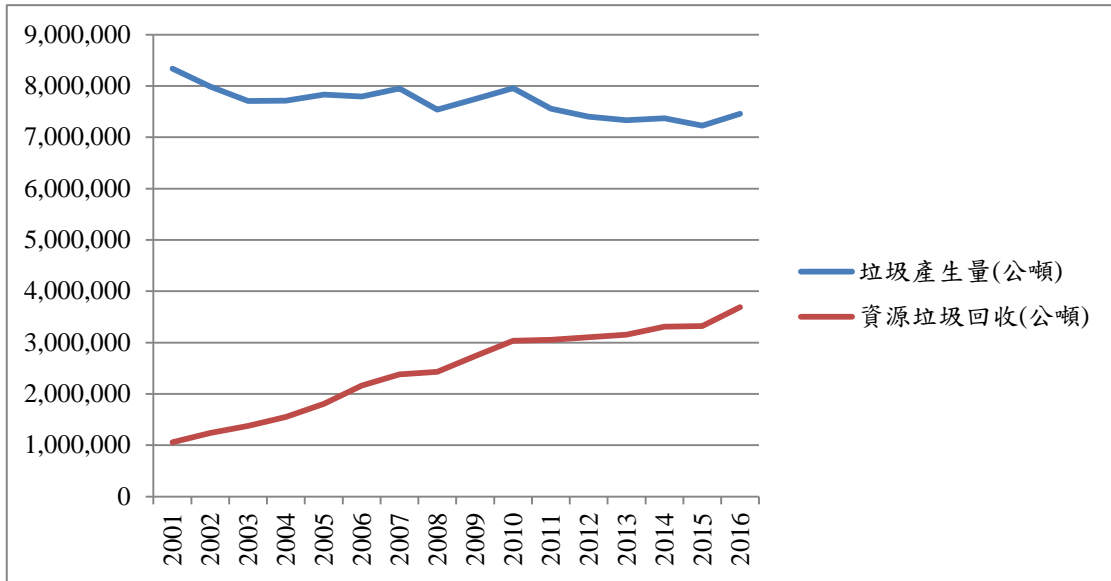


圖 4-3 2001 年至 2016 年垃圾清運量生產量及資源回收量折線圖

所以若某縣市垃圾清運量大於 313,700(公噸)跟城市規模以及人口數有關和工業發展也有關係，以 2014 年為例大於標準的縣市有新北市、桃園市、台中市及高雄市，2014 年的平均每人每日垃圾清運量為 0.383(公斤)，也就是每年每人垃圾清運量 $0.383 \times 365 = 139.795$ (公斤)，從下表可以看出，桃園市、台中市與高雄市的垃圾清運量生產量都超過各縣市每人生產量的平均值。

表 4-9 預測垃圾清運量生產誤差表

	人口數	預測生產量	垃圾清運量	誤差
新北市	3,966,818.000	554,541.3223	388,898.000	-165,643.322
臺北市	2,702,315.000	377,770.1254	274,166.000	-103,604.125
桃園市	2,058,328.000	287,743.9628	353,137.000	65,393.037
臺中市	2,719,835.000	380,219.3338	384,775.000	4,555.666
臺南市	1,884,284.000	263,413.4818	256,680.000	-6,733.482
高雄市	2,778,992.000	388,489.1866	408,745.000	20,255.813

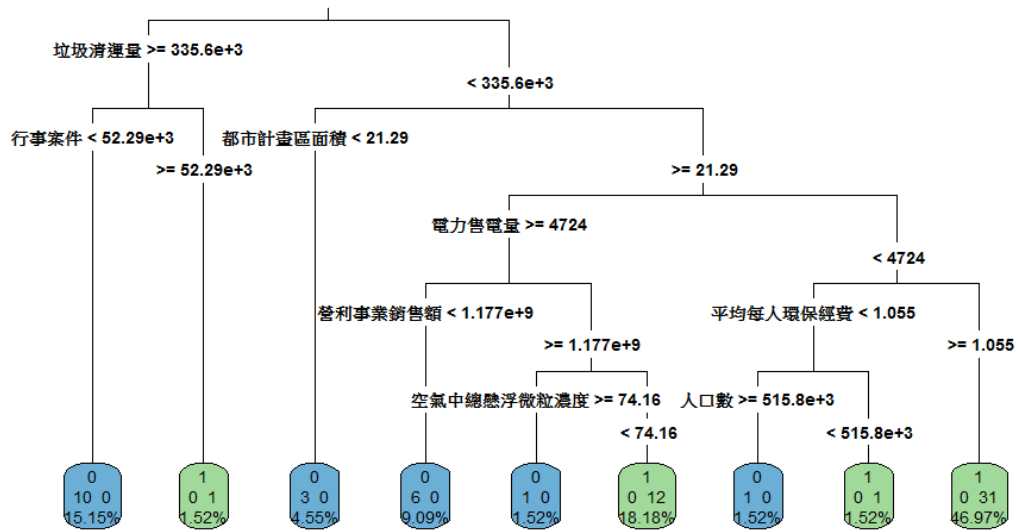


圖 4-4 2013 年至 2015 年決策樹圖

從 2013 年至 2015 年得決策樹可以看出垃圾清運量為第一個分支點，接下來刑事案件、都市計畫面積、電力銷售電量、營利事業銷售額和空氣總懸浮粒子，對影響到這三年的政府經營效率，本研究將三年的 22 個縣市做 70% 的驗證集和 30% 的測試集，發現此模型的正確率為 97.87%，幾乎可以完全預測此三年的政府經營效率。

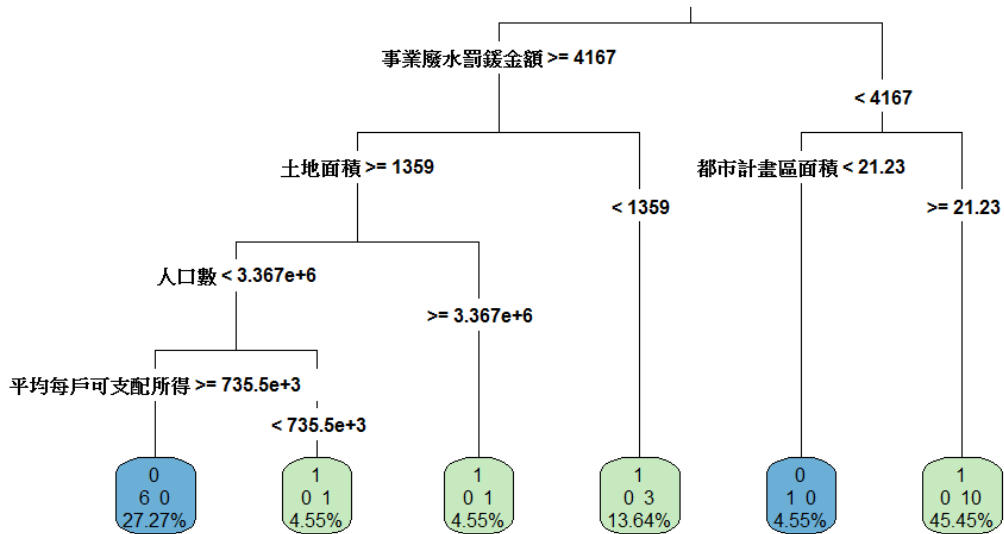


圖 4-5 2013 年決策樹圖

2013 年的主要分支為事業廢水罰鍰金額、土地面積、都市計畫面積、人口數和平均每戶可支配所得和三年的其中事業廢水罰鍰金額和都市計畫面積是重複的，所以可以推斷此兩項變數對於政府的經營效率有影響。第一個分節點為事業廢水罰鍰金額 ≥ 4167 ，事業廢水罰鍰金額為不良產出，若事業廢水罰鍰金額越小則政府效率越佳，第二層分節點有兩個分別為都市計畫面積和土地面積，都市計畫面積 < 21.23 (平方公里)只有澎湖縣所以在左邊為 0，土地面積大於等於 1359 平方公里，表示若土地面積大於 1359 平方公里越好，第三層為人口數 < 336400 人往左，表示人口越多越好，最後一個分支是平均每戶可支配所得 ≥ 735500 元，若小於的則往右邊為政府效率佳，這個比較例外是為了要排除屏東縣。

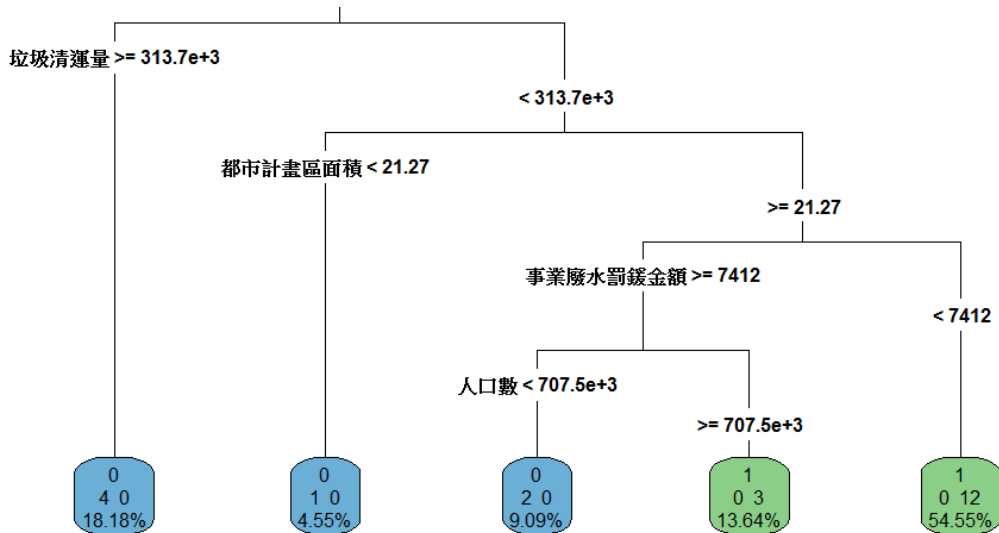


圖 4-6 2014 年決策樹圖

2014 年的主要分支為垃圾清運量、都市計畫面積、事業廢水罰鍰金額與人口數和三年的其中垃圾清運量與都市計畫面積是重複的，所以可以推斷此兩項變數對於政府的經營效率有影響。2014 年垃圾清運量最大的縣市為高雄市，而台灣六都平均每人每日垃圾清運量第一名桃園市:0.49(公斤);第二名高雄市:0.42(公斤);第三名台南市:0.38(公斤)，從決策樹的圖形可以看出來因為垃圾清運量為不好的產出，所以越小越好。第二個分支點為都市計畫面積小於 21.27(平方公里)唯有澎

湖縣的都市計畫面積小於 21.27(平方公里)，第三個分支為事業廢水罰鍰金額 ≥ 7412 ，若事業廢水罰鍰金額越多代表城市的政府經營效率為越差因為事業廢水罰鍰金額為不良產出，最後一個分支為人口數 $< 707,500$ 人，城市中人口數越多能帶給政府的各方面產出就越多，也可以藉此提升政府效能。

表 4-10 2014 年垃圾清運量敘述統計表

	垃圾清運量
總和	3,272,669.0
最大值	408,745.0
最小值	2,363.0
平均值	148,757.7

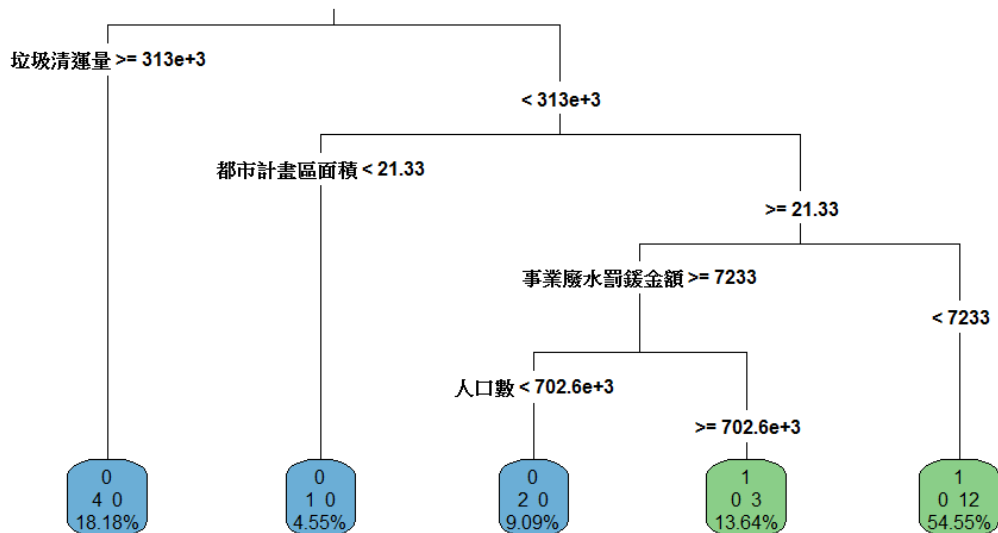


圖 4-7 2015 年決策樹圖

2015 年的決策樹圖與 2014 年的重要分支是一樣的指標但是贖時不同，也是在垃圾清運量、都市計畫面積、事業廢水罰鍰金額與人口數做分支。垃圾清運量為第一個分支點，因為垃圾清運量為不良產出，所以如果越低代表政府效率越好，因此將垃圾清運量小於 313000 公噸的縣市向右邊政府效率佳的地方走，而都市面積跟 2014 年一樣是排出澎湖縣，事業廢水罰鍰金額也是越少越好，而最後一層人口數則是越多越好，對於政府的整體營運效能有幫助。

第三節 隨機森林結果

從以上決策樹可以看出需多重要的指標會影響政府效率，使用隨機森林也可以驗證出決策樹對於政府的效率指標有相同的重要指標，能夠影響到政府的效率，隨機森林運用多個決策樹的分類器，並且其輸出的類別是由個別樹輸出的類別的眾數而定，而會產生 IncMSE 值，IncMSE 越大代表影響到目標的關係越明顯，使用隨機森林演算出來的指標分別為平均每人稅賦、垃圾清運量、都市計畫面積、事業廢水罰鍰金額以及電力售電量，其中垃圾清運量、都市計畫面積與事業廢水罰鍰金額在決策樹當中出現過很多次，也證明了決策樹的準確信。從下圖表可以看出每一個數值對於 DEA-SBM 目標政府效率的影響數值，越高的則是影響力越大。

表 4-11 隨機森林

	%IncMSE	IncNodePurity
平均每人稅賦	14.72	1.81
垃圾清運量	13.88	1.69
都市計畫區面積	12.43	1.59
事業廢水罰鍰金額	12.18	1.43
電力售電量	12.01	1.3
土地面積	9.44	0.75
就業人口	9.17	0.77
人口數	8.64	0.94
營利事業銷售額	8.09	0.74
歲出	7.65	0.64
弱勢族群福利	7.15	0.36
平均每戶可支配所得	6.32	0.45
空氣中總懸浮微粒濃度	4.08	0.25

刑事案件	4.04	0.41
平均每人環保經費	2.2	0.4

%IncMSE：均方誤差遞減意義下的重要性，若此指標值越大則說明此自變量對因變量的影響程度越大，若為 0 則說明此自變量對因變量沒有任何關係，若為負值則說明此自變量對因變量的變化可能有起到誤導的作用。

IncNodePurity：精確度遞減意義下的重要性。計算方法是殘差的平方和（非負），若此指標值越大則說明此自變量對因變量的影響程度越大，若為 0 則說明此自變量對因變量沒有任何關係。

(資料來源：<https://kknews.cc/zh-tw/tech/x5zlao.html>)

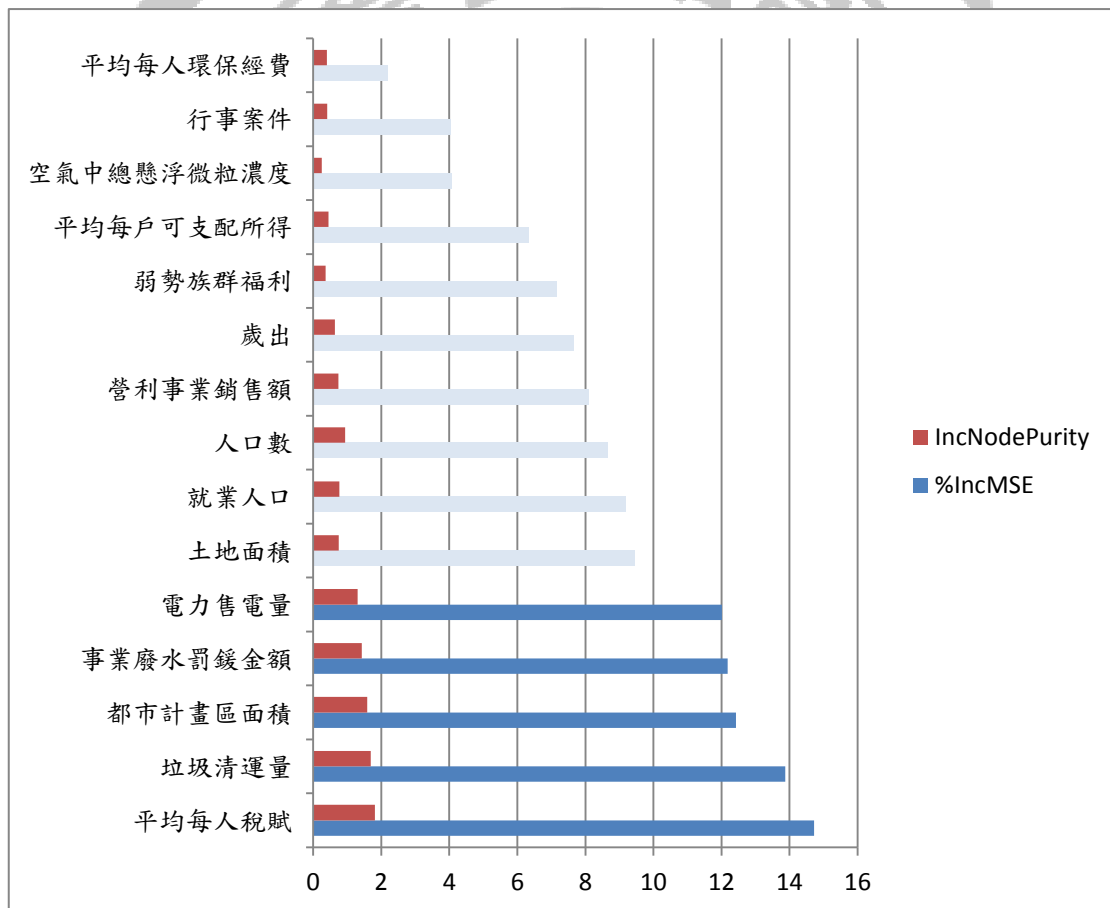


圖 4-8 隨機森林圖

第四節 分群結果

本研究將 22 個縣市分別以三年的資料作為分群，而三年的分群結果都相同，分別把 22 個縣市分為了首都台北市以及除了台北市以外的六都和其他縣市，由此可見

依據社會、經濟與環境面可以以數據的角度看出因為縣市的級別不同包過政府的福利、人口數、和經濟指標以及環境的各方面數值可以將台灣的縣市級別明顯的劃分出三個區塊。因為本研究有多重構面所以使用 PCA 降維演算法將多維陣列轉製成二維向量矩正。

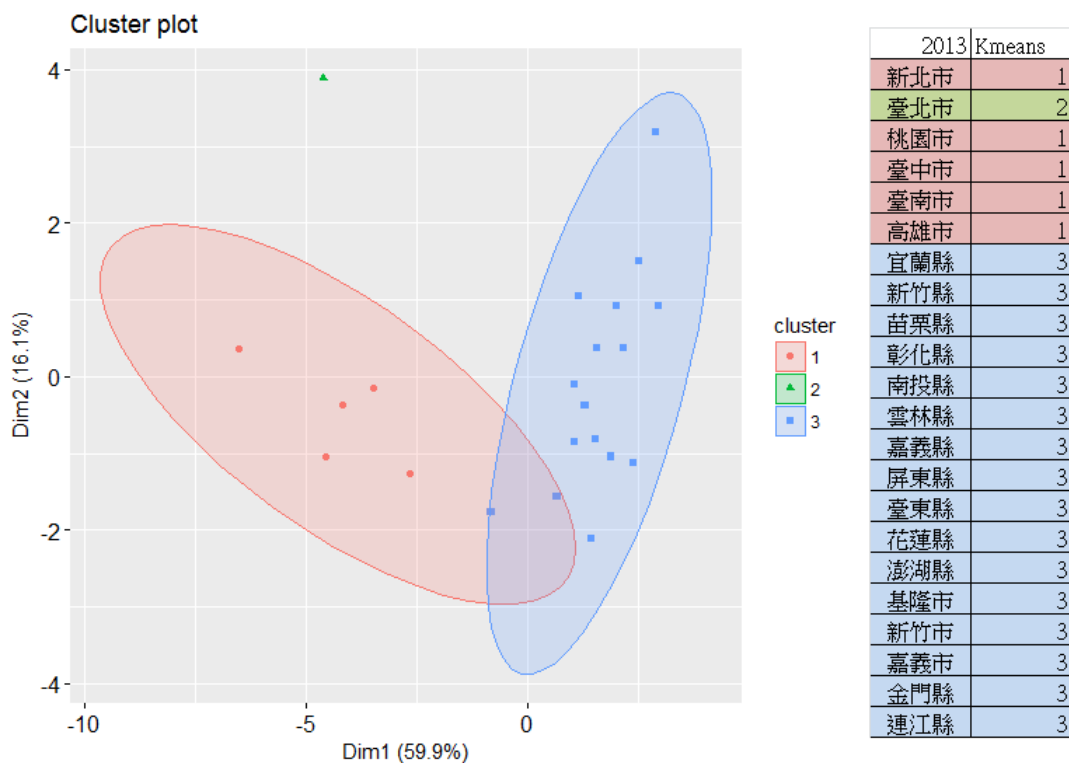


圖 4-9 2013 年各縣市分群

從分群可以看出六都的 DEA-SBM 投入：就業人口、歲出和電力銷售額，會比其他縣市來的高，由此可知如果較高的投入相對應也要有一定程度的產出對社會的福利以及處理環境保護的方面有更多的資源，目前除了原本的直轄市台北市以外其餘的五都包括新北市、桃園市、台中市、台南市以及高雄市都還沒有讓投入與產出平衡的狀態。

以 2013 年新北市為例如下表，由 DEA-SBM 所觀察出，若新北市以目前投入的情況弱勢族群福利少了 20% 左右需要再重視老人安養機構、兒童及少年安養機構、特殊境遇家庭生活扶助和身心障礙福利等等，為了要提升市府效率在經濟地

方面事業營利銷售額代表各縣市的金融概況，但是此項目並不需要增加太多僅 3.8%，而每戶可支配所得應在增加 27%，2013 年平均每戶可支配所得最高的縣市為台北市 127 萬次之為新竹市 122 萬，新北市則位於第 8，以新北市的就業人口全台最高來看新北市的每戶可支配所得相對較低，環境部分除了事業廢水需減少之外垃圾以及空氣懸浮粒子都比其他縣市較好。因此最=對於本次的分群可以看出六都除台北市之外其餘五個縣市擁有的資源較多但相對應得產出也需要提升上來，這樣市府效率可以更提升。

表 4-12 新北市需改善之產出表

SBM	1/DEA_Score			
I/O(投入/產出)	Data(原始資料)	Projection(相對值)	Difference(差距)	差距百分比%
新北市	1.038665885			
就業人口	1,910	1,616.369858	-293.6301417	-15.37%
歲出	146,754.1	146,754.1	0	0.00%
電力售電量	9,325	9,325	0	0.00%
弱勢族群福利	163,568	197,787.8973	34,219.89727	20.92%
營利事業銷售額	4,287,143,918	4,452,910,130	1,657,66212.1	3.87%
平均每戶可支配所得	931,239	1,262,107.48	1,168,983.58	27.90%
空氣中總懸浮微粒濃度	57.13	903.7886968	846.6586968	999.90%
垃圾清運量	399,825	562,764.7058	162,939.7058	40.75%
事業廢水罰鍰金額	23,726	8,099.758686	-15,626.24131	-65.86%

另外兩年 2014 年以及 2015 年的分群皆相同表示各縣市分群很明顯已台北市、六都除台北市以及其他 16 個縣市，分為 3 群。

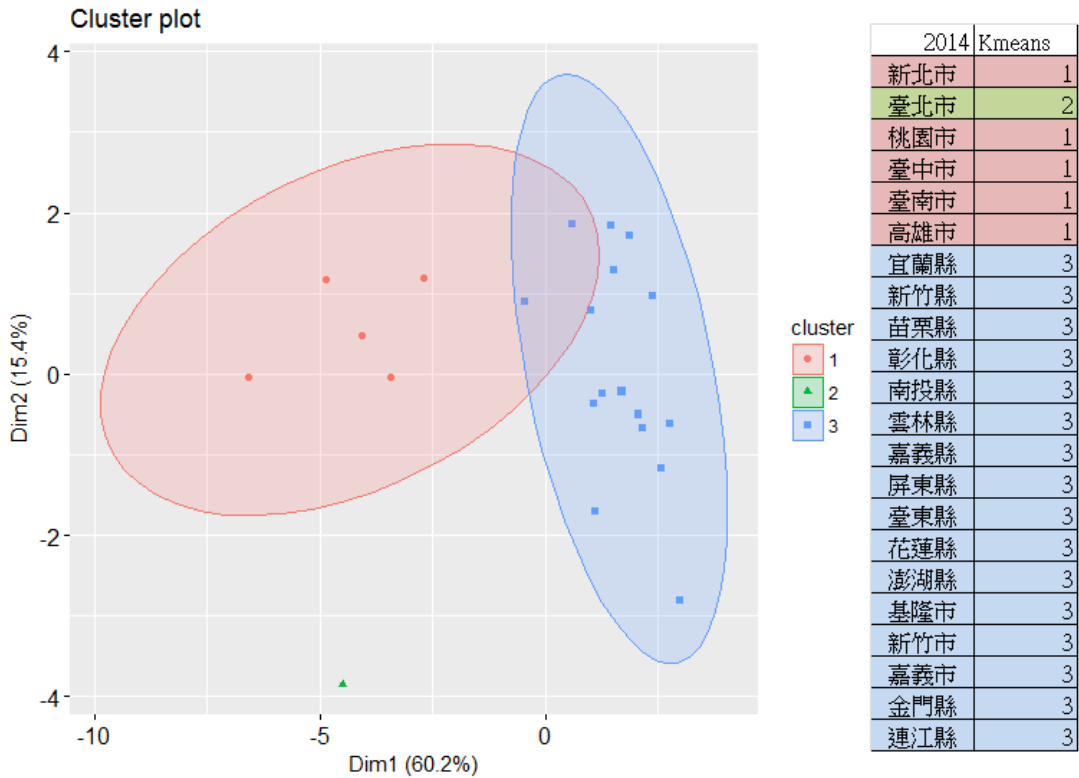


圖 4-10 2014 年各縣市分群

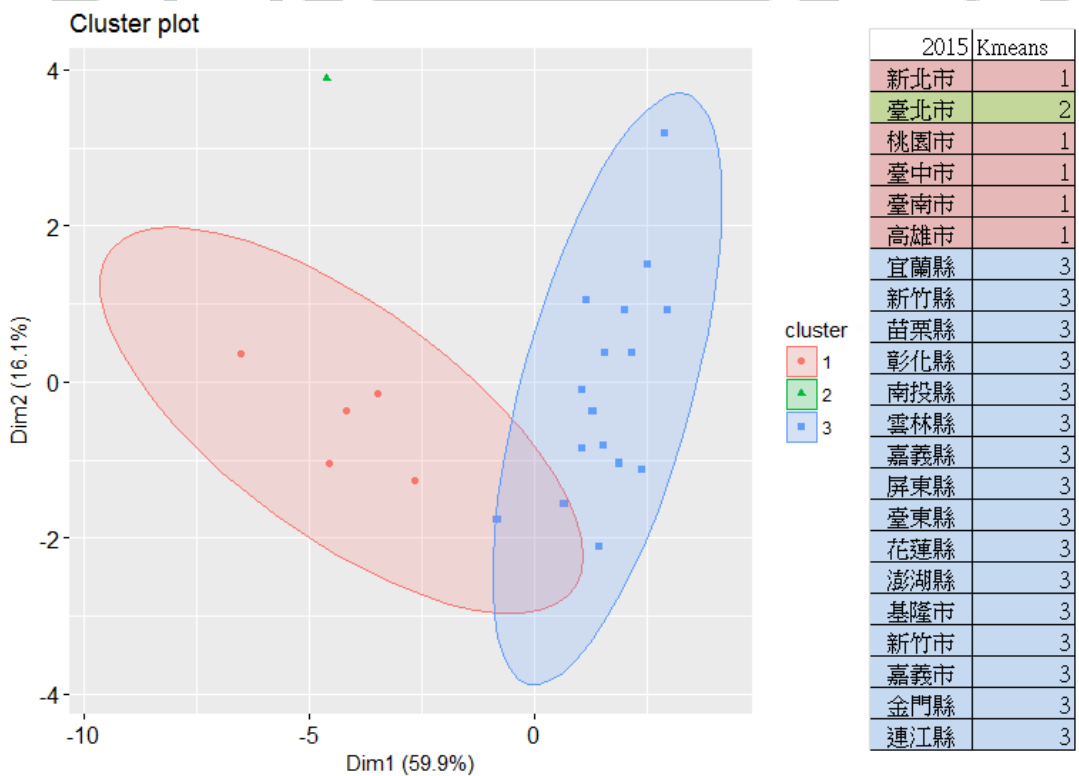


圖 4-11 2015 年各縣市分群

為了要使各縣市地方政府的效率越來越好所以提出了許多政策，包括廢棄物處理、

工業水回收及汙水處理、生育補助和節約能源等等，以下為各個政策在六都的推動下實施的成果。

廢棄物處理：至 2008 年，台灣運轉中之焚化廠有 24 座，可處理家戶及事業產生的可燃性垃圾，家戶垃圾的焚化率已達 94.6%，其餘垃圾送掩埋場處理，垃圾妥善處理率達 99.9%。實施「垃圾不落地」、「家戶垃圾強制分類回收」、「製造者延伸責任」及「源頭減量〈例如：限用塑膠袋、塑膠類免洗餐具、過度包裝〉」等政策後，每人每日垃圾清運量，從最高 1998 年的 1.14 公斤降到 2008 年的 0.52 公斤，家戶垃圾回收率達到 42%，與先進國家同步。事業廢棄物管理方面，以事業廢棄物流向電子化追蹤管理系統及建置清運車輛 GPS 追蹤管制系統，提升事業廢棄物妥善清理率，減少非法棄置，並以資源永續利用之原則，積極推動事業廢棄物再利用。依據 2008 年申報統計資料，事業廢棄物再利用率已達 75% 以上。未來將繼續推動源頭減量及資源再利用等政策，以邁向「資源循環零廢棄社會」。

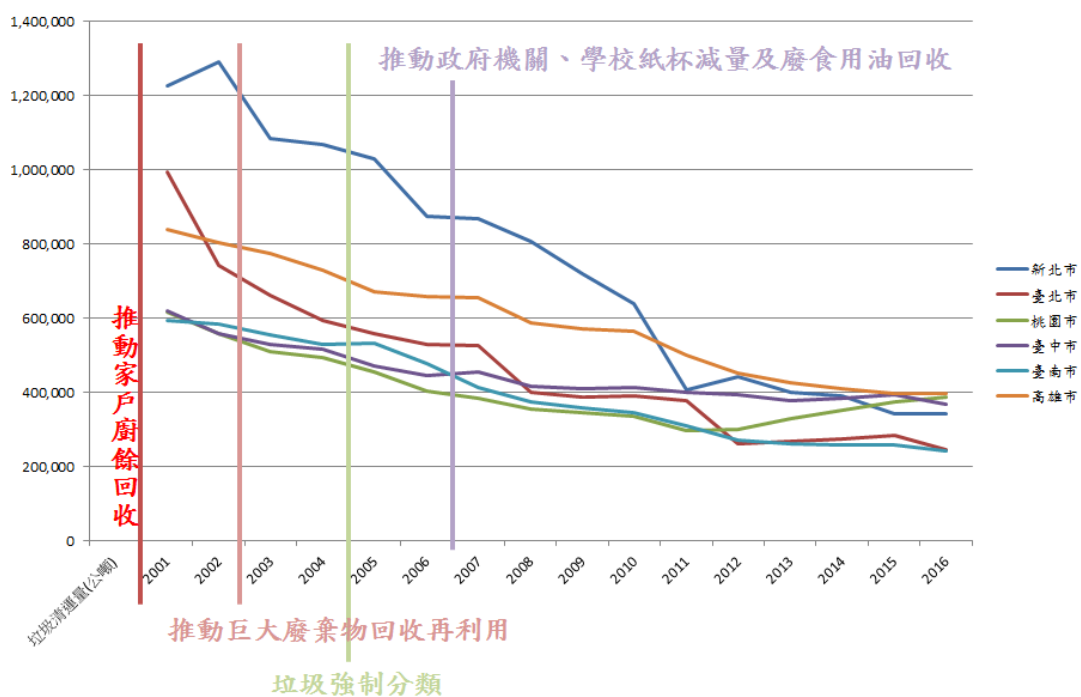


圖 4-12 六都垃圾政策時間關係圖

表 4-13 工業園區水支援回收政策

查核對象		環評承諾	
		製成用水回收率	全場回收率
竹科	環調及因應對策	88 年後興建：85%	—
		83~88 年興建：70%	
		83 年前興建：50%	
	三、五路	85%	—
	竹南基地	85%	60%
	龍潭基地	85%	75%
中科	台中園區	半導體業：85%	—
		光電業：80%	
	后里園區	半導體業：85%	—
		光電業：85%	
	七星園區	全區回收率：85%	
南科	台南園區	半導體業：85%	半導體業：75%
		光電業：85%	光電業：75%
	高雄園區	85%	75%

為了將水資源再利用，各大工業園區進行水回收再利用措施，下圖為六都事業廢水發還金額，此趨勢與經濟發展有關桃園市的工業區以及高雄市的工業區廢水污染超標得比較嚴重，為了要使水污染的問題能夠有效的處理，政府應該更重視水汙，河川以及海洋污染的嚴重性，所以應該推動相關政策保護環境。

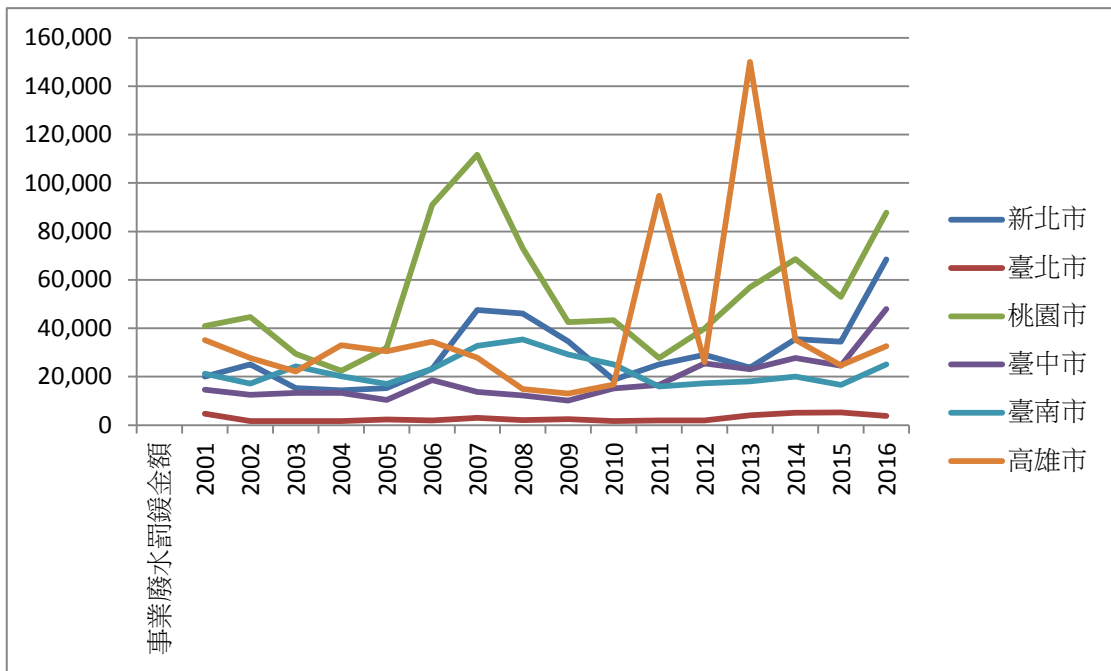


圖 4-13 六都事業廢水罰鍰折線圖

近年我國出生率持續下降，若新生兒的人數縮減人口老化勞動人口不足，會導致國家稅收不足，人民的賦稅會越來越重加上撫養比增加，年輕人的經濟壓力會更加沉重，若新生兒得數量減少會使學校倒閉許多老師會失業，並且會造成國內市場萎縮，使經濟無法成長國家財政會出現巨大的黑洞，因此為了要使生育率提升政府提出生育補助措施，希望以後生育路會持續上升，解除國家的財政危機。

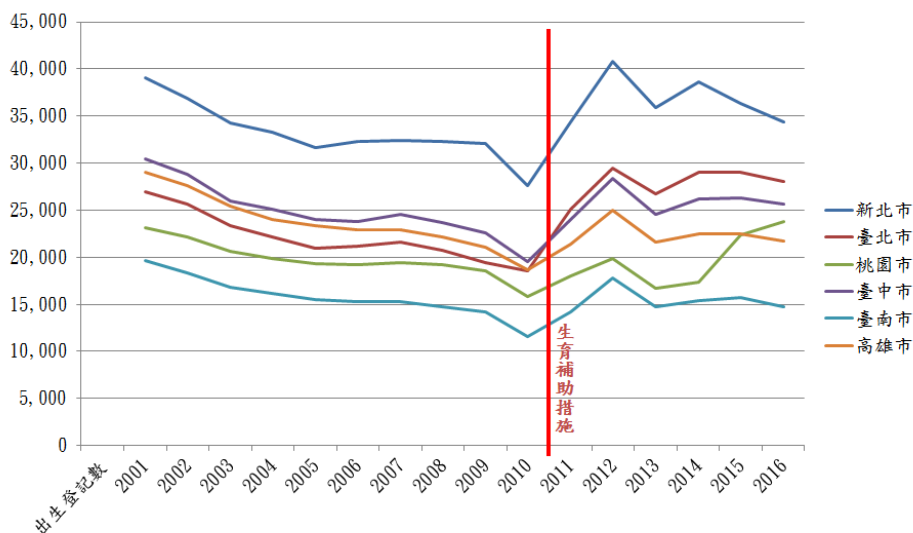


圖 4-14 六都生育補助政策下新生兒個數

政府帶頭各機關率先落實節電，並廣大納入國營企業，機關學校四省評比績優的單位頒獎表揚，各縣市棟同推動節約能源並舉夏季節電競賽，獲獎縣市除表揚頒獎之外，並給予節電推動獎金。

各產業參與節電動，能源規模大的產業為節電重要目標，服務業全面實施冷氣不外洩與室溫不低於 26 度，節電廠商經濟部給予表揚。民眾自發家庭日省一度電學校結合家庭共同節電，台電給予節電獎勵。

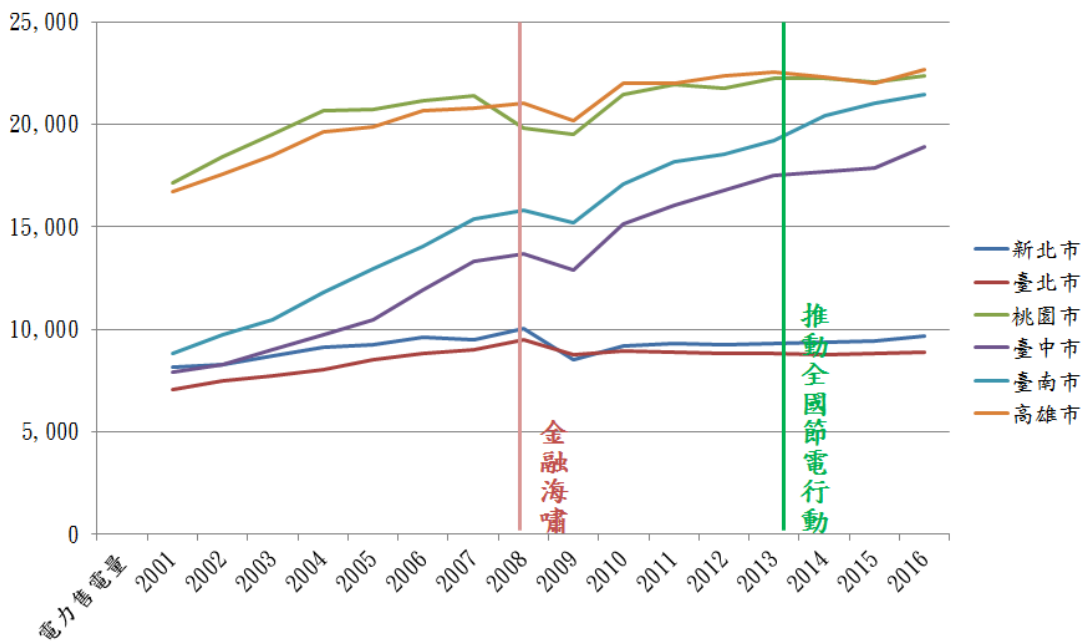


圖 4-15 六都電力售電量與政策折線圖

第五章 結論與研究限制

本文研究選取政府開放式資料 2013 年至 2015 年社會、經濟與環境 15 個變數的經營能力，以包絡分析法(DEA-SBM)分析從投入、不良產出和產出得出經營指標，在使用決策樹進行分群，並且得出三年的決策樹，分別探討每一個變數的關係和內容影響台灣的政府效率，最後得出各縣市的垃圾處理以及廢水會影響政府的效能，根據本次研究發現各縣市的不良產出，對於各縣市的投入有相當的影響。

期望未來可以對各縣市以較好的縣市為例，減少不良產出並且增加老的產出，使政府效能可以提升，影響各縣市的社會、經濟與環境的發展。以便政府能夠提升效率在社會、經濟與環境更有效提升各縣市的政府效能以及兼具環境保護的永續經營。

從此本次分析可以推斷出因為自從縣市合併將包括桃園、台中、台南以及高雄，因為土地以及原有的舊縣都涵蓋在其中因此可能會導致 DEA-SBM 模型中的效率在新的都會區當中發現有需要改進的社會、經濟與環境的議題，在 2015 年的 DEA-SBM 模型中台南市已經從需要改進更變成市府效率佳的城市。本研究分析出各地方政府在經濟發展的況下能夠對環境的保護是否能夠同時兼顧，從分析結果看出若地方政府能夠更重視空氣、水和土地污染的情況下會使市政更有永續經營的能力也在台灣的土地上能夠有良好的居住環境以及發展空間，希望未來政府能夠更重視我們的土地制定政策將我們的城市做改善並且能夠有永續經營的能力。

本研究研究限制，因 2013 年以及 2014 年的事業廢水罰鍰金額連江縣資料缺乏，但因 DEA-SBM 模型不可有數值為 0 的資料，所以以 2013 年的數值為替，所以可能影響到 DEA-SBM 指標變數。

文獻參考

中文文獻

1. 蘇進祿(民 93)。以資料包絡分析法評估鋼鐵產業經營績效之研究。成功大學高階管理碩士在職專班(EMBA)學位論文，1-82。
2. 趙李英記(民 102)。「隨機森林運用於白血病基因分類」，第九屆知識社群國際研討會。
3. 林郁珊、張智安(民 100)。應用空載全波形光達資料於波形分析與地物分類 (Doctoral dissertation)。
4. 魯濤、周晶(民 99)。基於 DMU 異質性的 DEA 研究述評及展望。生產力研究，2，251-253。
5. 李正文、陳翔修(民 97)。台灣光電產業之經營效率分析-資料包絡分析法之應用。中原企管評論，6(1)，1-30。
6. 葉家瑋(民 105)。決策樹應用於情境分類歸納。臺灣大學工業工程學研究所學位論文，1-64。
7. 陳錫欽(民 103)。利用資料探勘技術探討服務品質對捐血意願之影響-以台灣中部地區為例。嶺東科技大學資訊科技應用研究所學位論文，1-74。
8. 陳詠霖、郭玟君(民 104)。運用決策樹預測與預防台灣高犯罪風險青少年之犯罪問題。聯大學報，12(2)，139-149。
9. 郭家姣(民 103)。隨機森林在河川水位即時預報之應用。成功大學水利及海洋工程學系學位論文，1-72。
10. 卓達瑋(民 99)。隨機森林分類方法於基因組顯著性檢定上之應用。政治大學統計研究所學位論文，1-51。
11. 林焜詳(民 105)。支撐向量機與隨機森林應用於颱風時雨量預報之比較。成功大學水利及海洋工程學系學位論文，1-105。

英文文獻

12. Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399-418.
13. Cao Mengtian; Luo Luyang(2017) The policy performance evaluation of new energy automobile industry based on DEA-SBM method. 2017 3rd International Conference on Information Management (ICIM).1-5.
14. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429-444.
15. Chehata., N., Guo, L., & Mallet, C., (2009), "Airborne LiDAR Feature Selection for Urban Classification using Random Forests,"*IntArchPhRS*, 38(3), 207-212.
16. Crimmins, S.M., Dobrowski, S.Z., & Mynsberge, A.R., (2013), "Evaluating Ensemble Forecasts of Plant Species Distributions under Climate Change,"*Ecological Modelling*, 266, 126-130.
17. Gurin,J(2014)Open Data Now,McGraw Hill Education.
18. Lin, Cheng-Hsien(2015). Dynamic Performance Perspective with Matthew Effect. Diss. 56-68
19. Martin, C. (2014). Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective. *Policy & Internet*, 6(3), 217-240.
20. Odom, M. D. and R. Sharda.(1990). A neural network model for bankruptcy prediction. *IEEE INNS IJCNN* 2: 33-38.
21. Pan, Y. H. (2015). The Evaluation of Car Icons and Head Up Display Icons Design Using a Decision Tree.
22. Peck, E., Perri 6, Gulliver, P., & Towell, D. (2004). Why do we keep on meeting like this? The board as ritual in health and social care. *Health Services Management Research*, 17(2), 100-109.
23. Quinlan, J. R. (1993.)C4.5: Programs for Machine Learning. Morgan Kaufmann

Publishers,

24. Ran Wang; Sam Kwong; Xi-Zhao Wang; Qingshan Jiang(2015) Segment Based Decision Tree Induction With Continuous Valued Attributes. IEEE Transactions on Cybernetics. 1262 – 1275.
25. S. R. Nandanwar; S. B. Warkad; K. M. S. Y. Konara; M. L. Kolhe; N. P. Patidar(2016) Load management for voltage security using probabilistic fuzzy decision tree method. 2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS). 1 – 5.
26. Satya Samyukta Kambhampati; Vishal Singh; M. Sabarimalai Manikandan; Barathram Ramkumar(2015) Unified framework for triaxial accelerometer-based fall event detection and classification using cumulants and hierarchical decision tree classifier. Healthcare Technology Letters. 101 – 107.
27. Sayogo, D. S., Pardo, T. A., & Cook, M. (2014, January). A framework for benchmarking open government data efforts. In System Sciences (HICSS), 2014 47th Hawaii International Conference on (pp. 1896-1905). IEEE.
28. Schrock, A. R. (2015). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New media & society*, 18(4), 581-599.
29. Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, 27(3), 3-24.
30. Sonal Singhal; Shrey Jain; Nithin B; Bhawna Rawat; H. O. Gupta(2016) Performance evaluation of Uttarakhand electric utility data using slack based measure and Two-stage DEA-SBM modelling. 2016 International Conference on Electrical Power and Energy Systems (ICEPES)27-34.
31. Yao, C. C. (2014). The Study of a Bayes Classifier Enhanced by Least-Mean-Square Learning Approach.

32. Yisen Wang; Shu-Tao Xia(2017) Unifying attribute splitting criteria of decision trees by Tsallis entropy. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2507 – 2511.
33. Yu, H., & Robinson, D. G. (2012). The new ambiguity of 'open government'.

