

東海大學資訊管理研究所
碩士學位論文

以資料探勘分析影響公司獲利能力之財務指標-
以上市電子零組件公司為例

**An Analysis of Financial Indicators Affecting Corporate
Profitability with Data Mining- A Case Study of Electronic
Component Companies in Taiwan**

指導教授：姜自強 博士

研究生：林明璿 撰

中華民國 106 年 8 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 林明璿 君所提之論文

以資料探勘分析影響公司獲利能力之財務指標-以上市
電子零組件公司為例

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：黃光良 (簽章)

委員：

姜育強
許懷中
楊朝棟
陳淑芬

中華民國 106 年 7 月 6 日

謝誌

時光匆匆地已然過去，就讀研究所的兩年時光已經悄悄的進入的尾聲，兩年的時光中經歷了很多的事，遇見了很多人，這些事與人都在這兩年的時光中帶著我進步成長，在這我所要感謝的人很多，其中之一就是我的指導教授姜自強博士，在這兩年來，教授一直秉持著讓學生往專業領域去發展，但也讓學生可以享受到學習中的自由，除了在學校學習到專業知識外也讓學生更知道如何去運用專業，因為如此的教學態度，讓我在學習如何進行研究並撰寫論文時體會到不同以往的經歷，雖然教授的教學方式很自由，但是也不忘的提醒論文撰寫方面的每個細節，也在失去研究方向時給予了不同觀點讓我可以繼續地向前邁進，這些事情的種種都讓我體會到教授的細心，所以我感謝教授這兩年的指導。

當然這兩年除了教授的細心指導，一定還有同學們的陪伴，在這兩年的時間裡我認識了許多的同學，他們在這兩年的時間裡容忍我怪異的個性與口不擇言的說話方式持續的給予我幫助，其中顯嚴在我剛進入研究所時給予我歡樂，詩楷則常常規劃行程帶著我與同學們在台灣各地遊玩，信毅也提供各種器材在研究室供同學們使用，羿勳在我論文遇到困難時不厭其煩的幫助我，當然我的同學不只是這幾個，但是我要感謝的人實在太多了，其中也包含了熱心助人的學弟妹們，系辦的學長等，所以我就以感恩的心來謝謝他們這兩年期間的陪伴。

而在這兩年的時光中不只是老師同學學長和學弟妹的陪伴，也包含了家人的支持與朋友的陪伴，我感謝我的家人不只是這兩年而是這二十幾年來的支持，願意讓我可以專心地完成人生的求學階段，雖然其中發生了許多的磨擦有爭吵與淚水，但是一切都挺過來了，在這裡我要感謝精一與彥碩和許多陪伴著我走過的低潮與寂寞的朋友，並在感謝的尾聲，我要謝謝母親努力的容忍我的任性不管發生甚麼事都永遠的支持我，謝謝你們。

林明璿 謹誌於

東海大學資訊管學系 2017 年 07 月

摘要

論文名稱：影響公司獲利能力財務指標分析-以上市電子零組件公司為例

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 08 月

研究生：林明璿

指導教授：姜自強

論文摘要：

本研究主要是以財經資料來進行分析並探討出影響公司獲利能力的財務指標並了解財務指標和獲利能力間相互的影響方式為主要的研究目的。

目前大多的投資人在進行投資行為時會藉由觀看財務報表中的指標來選擇投資項目，但傳統拿來作為參考的財務指標對於投資結果沒辦發進行準確的預測，所以本研究希望可以藉由資料探勘的方式來進行分析，試著找出有別於傳統更能預測良好投資結果的財務指標。

本研究藉由各種分析結果來進行相互探討後發現營業成本、營業利益、每股盈餘是影響獲利能力的主要關鍵並了解公司規模與獲利能力的關係不大，並藉由 13 年間的年的股票數值驗證了本研究的結果具有一定的準確性，發現營業利益及每股盈餘越高公司獲利能力越好。

關鍵字：獲利能力、資料探勘、營業成本、營業利益、每股盈餘

Abstract

Title of Thesis : An Analysis of Financial Indicators Affecting Corporate Profitability with Data Mining- A Case Study of Electronic Component Companies in Taiwan

Name of Institute : Tunghai University, Institute of Information Management

Graduation Time : (08/2017)

Student Name : Ming-Chin Lin

Advisor Name : Tzu-Chiang Chiang

Abstract :

This study is based on financial information to analyze and explore the impact of the company's Corporate Profitability Financial Indicators, and understand Financial Indicators Mutual influences the way and Corporate Profitability is the main purpose of the study.

At present most of the investors in the investment behavior by viewing the financial statements of the index to select investment projects, but the traditional Financial Indicators used as a reference for the investment results do not send accurate prediction, so this study hope by Data Mining the way to carry on the analysis, and try to figure out what is different from the traditional can predict good investment results of Financial Indicators, based on the analysis results to explore each other after the discovery of Operating costs, Operating Profit, Earnings Per Share Corporate Profitability is the key and not understanding the relationship between firm size and Corporate Profitability, And through 13 years of stock value to verify the results of this study, with certain accuracy, found that Operating, Profit and Earnings Per Share higher, the company Corporate Profitability better.

Keywords : Corporate Profitability, Data Mining, Operating costs, Operating Profit, Earnings Per Share

目次

頁次

摘要	I
Abstract	II
目次	III
表次	V
圖次	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	4
第三節 論文架構	5
第四節 預期成果	5
第二章 文獻探討	6
第一節 台灣經濟新報資料庫(TEJ)	7
第二節 資料探勘(Data Mining)	13
第三節 資料包絡分析法(DEA)	16
第四節 決策樹(Decision tree)	21
第五節 相關分析(Correlation Analysis)	25
第六節 隨機森林(Random forest)	26
第七節 分群分析(K-Means)	29

第三章 公司獲利能力財務指標分析方法設計	31
第一節 研究流程	31
第二節 研究決策單位及主要資料說明	32
第三節 資料包絡分析法(DEA)投入選項及產出選項	33
第四節 主要分析資料	35
第四章 分析結果與解釋	37
第一節 資料包絡分析法結果與說明	37
第二節 決策樹分析結果與說明	44
第三節 依照隨機森結果變數分析之決策樹結果與說明	50
第四節 隨機森林分析結果與說明	52
第五節 相關結果與說明	53
第六節 分群結果與說明	55
第五章 結論與研究限制	60
參考文獻	62

表次

頁次

表 1-1 台灣每年股票交易市場的交易情形.....	3
表 2-1 台灣經濟新報資料庫(TEJ) 資料提供國家.....	11
表 2-2 台灣經濟新報資料庫(TEJ) 主要項目.....	12
表 3-1 91 家電子零組件產業公司圖表.....	32
表 3-2 經營能力變數敘述性統計表.....	33
表 3-3 獲利能力變數敘述性統計表.....	34
表 4-1 2002 年到 2015 年 DEA 與股價分析狀況.....	40
表 4-2 91 家電子零組件公司資料包絡分析法(DEA)之結果.....	42
表 4-3 決策樹分析驗證.....	44
表 4-4 決策樹分析結果統整.....	45
表 4-5 三年各 DEA 最大值.....	46
表 4-6 隨機森林之決策樹分析結果統整.....	51
表 4-7 隨機森林分析結果.....	53
表 4-8 主要分析資料 CRONBACH ALPHA 值.....	53
表 4-9 2013 年相關分析分析結果.....	54

圖次

	頁次
圖 1-1 台灣股票各類股成交比重	2
圖 1-2 電子業個類上市公司的數量	2
圖 2-1 KDD 流程圖	14
圖 2-2 資料包絡分析法的示意圖	20
圖 2-3 決策樹示意圖	24
圖 2-4 隨機森林流程圖	28
圖 2-5 分群分析流程圖	30
圖 3-1 研究流程	31
圖 4-1 各公司 DEA 示意圖	38
圖 4-2 三年股市交易情形曲線圖	38
圖 4-3 各公司三年股市交易情形曲線圖	39
圖 4-4 2013 年決策樹結果	47
圖 4-5 2014 年決策樹結果	48
圖 4-6 2015 年決策樹結果	49
圖 4-7 2013 年隨機森林之決策樹分析果	50
圖 4-8 2014 年隨機森林之決策樹分析果	50
圖 4-9 2015 年隨機森林之決策樹分析果	51
圖 4-10 2013 年分群結果	56
圖 4-11 2014 年分群結果	56
圖 4-12 2015 年分群結果	57
圖 4-13 2013 資料標準化後分群結果	58
圖 4-14 2014 資料標準化後分群結果	58
圖 4-15 2015 資料標準化後分群結果	59

第一章 緒論

第一節 研究背景與動機

金融市場隨著時間的演進，在幾十年間蓬勃的發展，讓現在許多的民眾都可以利用金融交易的方式買賣股票、基金等，進行金融投資行為來增加金錢收入，而這種行為已經是越來越普遍了，也隨著時間的演進，經過了種種金融事件，大眾對於金融投資行為所存在的風險越來越了解，所以大眾在進行投資行為時開始會利用各種方法來避免風險發生，而其中以觀看各公司的財務報表來選定合適投資標的為做常見的作法，金融投資中，股票買賣是最為熱門的，而所謂股票是在經濟的發展中，各企業為了籌措龐大的資金進而發展出一種股票市場的交易機制，是屬於一種金融投資人用金錢購買的有價證券是經由上市公司的以集資為目的發放給投資者作為公司資本額的一部分的所有權證明，而持有相關證券的投資人可以以股東的身分獲得股利和禮品，並且股票的面額會因應公司的成長或股票交易市場的財務波動來進行價格的正向波動，但是如果公司營運狀況不佳導致交易市場的負成長，股票的價格也會因此下降，所以這就是股票的持有者必須共同承擔的風險，但雖然股票的投資會帶有一定的風險，但是以金融交易市場來說，投資股票的項目中還是最熱門，最多人投資的，所以因為股票投資較為受到大多投資者的喜愛，因此相較於其他投資行為，股票的相關資料獲取的方式也越來越多，越來越容易，在這種情形下投資者可以藉由許多管道拿到許多的財務資料，所以投資股票可以藉由資料上進行買賣上的判斷，在了解上市公司的營運狀況，及財務資料後，可以藉由資料來判斷哪間公司的股票適合投資，進而降低投資的風險，所以現在一般的投資者在選擇股票投資標的的時候，藉由觀察投資標的財務狀況和獲利指標來選擇較為績優風險性較低的股票是非常常見的。

本研究在這樣的時代背景之下，認為對公司財務指標進行分析，並找出其關鍵的指標是有意義的，所以希望藉由資料探勘的手法來分析出對於各上市公司的營運狀況有較大的影響力的財務指標。

以下是台灣股票各類股成交比重：

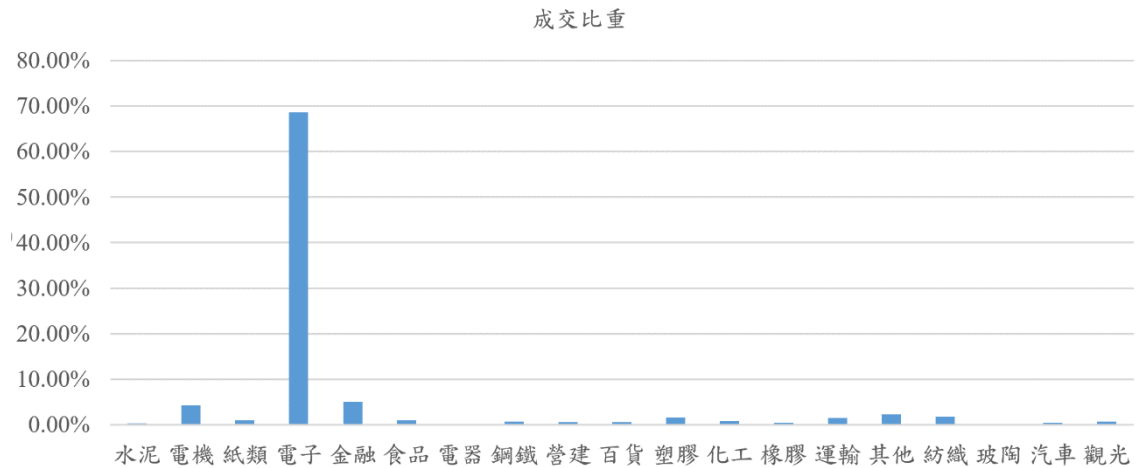


圖 1-1 台灣股票各類股成交比重

從圖 1-1 可以看出，在所有股票類別的產業中，以電子股交易類類別最大，所佔有的交易比重最多，所以本研究採用電子產業的財務資料進行分析，希望可以分析出較為接近實際交易市場的情況。

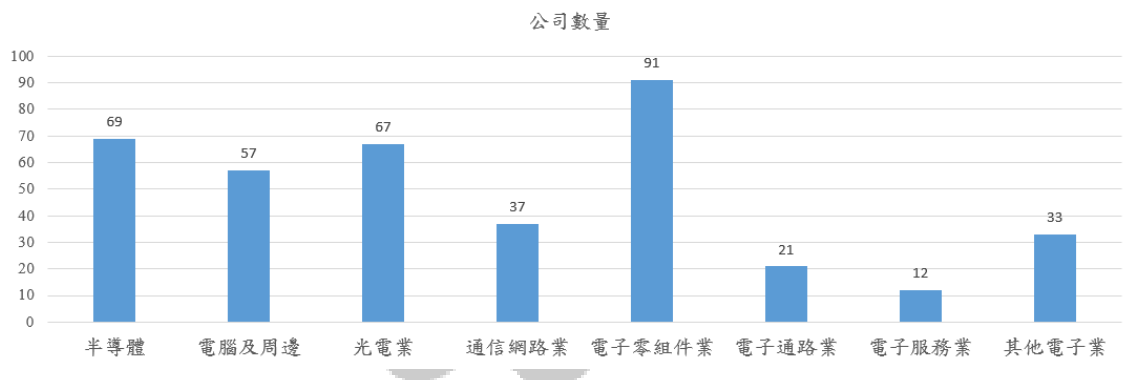


圖 1-2 電子業個類上市公司的數量

從圖 1-2 可以看出台灣 2016 年度電子業個類上市公司的數量，可以知道上電子業上市公司中以電子零組件最多，所以本研究選用了電子業中電子零組件公司的財務報表為主要的研究項目。

本研究是利用台灣經濟新報資料庫(Taiwan Economic Journal, TEJ)來進行資料的分析與探勘，而主要以電子業中電子零組件公司為主要的分析族群，其中主要是以上市公司的財務報表做作為主要分析資料，並透過資料包絡分析法(Data Envelopment Analysis, DEA)來分析各電子零組件公司財報取得可以判定公司獲利能力好壞的數值，並以決策樹、相關、隨機森林等分析方法來做資料探勘的動作並找出關鍵的變數，也就是影響公司獲利能力的財務指標，以下是本研究已經金融投資為主要的研究題材以及使用電子業中電子零組件為主要的投資目標得相關數據以及原因。

以下是台灣每年股票交易市場的交易情形：

表 1-1 台灣每年股票交易市場的交易情形

西元年	上市公司家數	上市總股數	成交總股數(單位：千股)
1986	130	24,082	39,040,917
1991	221	61,671	175,941,044
1996	382	162,680	350,738,899
2001	584	406,400	606,420,171
2006	688	549,493	732,510,321
2011	790	602,677	650,941,011

資料來源：台灣證券交易所

表 1-1 是台灣 1986 年開始到 2011 年每隔 5 年台灣股市的概況，可以看出從 1986 年開始台灣股市不論是上市公司、股數、面額等都大幅度的上升，而從成交總股數可以得知，台灣股票市場得交易量越來越大，交易越來越熱絡，所以社會大眾參與金融市場的投資行為比重越來越高，是一個合適的研究項目。

第二節 研究目的

根據了研究的背景與動機，本次研究主要是利用資料包絡分析法來分析台灣電子業中電子零組件目前的營運狀況，本研究採用了台灣經濟新報資料庫(TEJ)中所提供的財務數據，並擷取資料庫中獲取 2013 年以後至 2015 年的數據進行分析，因為 2013 年到 2015 的的財務資料最為完整，並且公司數量都為 91 家，為了了解近年台灣電子股票市場的波動，本研究對於公司財務情形的判定是以資料包絡分析法(DEA)來做為分析藉由此總分析方法可以利用將財務資料分為投入產出換算出可以代表公司營運狀況的數值並作為目標變數來進行其他的分析作為分析得基準，其中資料包絡分析法(DEA)以經營能力為投入，獲利能力為產出，來了解各家的財務波動，而得到相關數值，這些數值可以作為目標變數來知道目前電子產業中電子零組件的營運狀況，而本研究也因應包絡分析法(DEA)之分析結果，來進行各種統計分析方法。

本研究目的如下：

1. 使用上市電子公司中以電子零組件為主要營業項目的公司的財務報表為主要研究資料，希望模擬出台灣投資市場現況。
2. 以資料探勘的方式分析公司的財務報表，來得到影響公司獲利能力之財務指標。
3. 以財務指標與各種分析結果做比對，了解財務指標對獲利能力的影響狀況。

第三節 論文架構

本研究主要是利用 R studio 作為資料包絡分析法(DEA)、隨機森林、決策樹、分群等分析手法的分析軟體，在相關分析中，使用了(Statistics Analysis System, SAS)中(Enterprise Guide, EG)中作為分析軟體，本研究的第一章包含了緒論、研究背景與動機、研究的目的與論文架構，第二章文獻探討中包含了，隨機森林、資料包絡分析法(DEA)、決策樹、相關等等的相關研究及本研究對於這些研究的看法，第三章公司是獲利能力財務指標分析方法設計，主要以分析之資料說明和分析目標變數所使用之資料等，第四章是本研究所使用的分析方法以及分析的結果和所分析出的相關數值及解釋，第五章是本研究藉由所分析出之數值所得出之結論。

第四節 預期成果

判定公司獲利能力，依照前人做過的財務研究可以運用資料包絡分析法得到判定的數值，而決策數可以作為找尋影響獲利能力之財務指標，本研究希望在投資市場分額較重的電子業作為研究目標利用資料包絡分析法進行判定公司獲利能力的判定並以決策樹找尋影響獲利能力之財務指標，再藉由相關、隨機森林、分群的分析手法來尋找財務指標的關聯性、影響程度或者是分析依據，本研究預期可以從 39 種財務變數中，找到最能判斷獲利能力的財務指標，並了解分析出的財務指標在 39 種財務變數中個別的相關性和判斷獲利能力的依據。

第二章 文獻探討

隨著時間的演進，人們獲取金錢方式也越來越多，而其中投資行為也越來越普遍，不論是股票、基金還是期貨，都有大量的民眾進行投資，而隨著科技的衍進投資行為已經可以在網路上進行，不用親自到證券公司或通過電話才能進行，而也因為網路的發達獲得投資公司的資訊越來越方便，所以如能夠透過資料探勘的技術進行分析，應該能獲得不錯的結果，所以投資資料上的資料探勘是值得去深入研究的。

而本研究會以電子業中電子零組件為主要分析資料是因為，在台灣所有上市公司中電子類公司的公司數量最多，而其中以電子零組件的投資額最大，較能夠模擬出台灣產業和投資情形的現狀，並且本研究選擇以年底財務資料作為研究，因為一般公司在發布財務報表時，是以年底公布的財務資料較為完整齊全，在分析後結果可以較為準確，所以本研究以上市電子零組件各公司年底所公布的財務報表為主要的研究資料。

研究資料來源選定以台灣新報資料庫(TEJ)為來源，主要原因台灣新報資料庫(TEJ)提供之資料完整性較為良好，作為主要研究來源相對穩定，各家控股公司財報完整性較差，利用多家證券公司所提公司資料，較容易出現有資料依據單位不同，甚至出現空值，會造成分析結果出現誤差，所以本研究統一使用台灣經濟資料庫做為研究主要資料的來源，確保結果的穩定性與準確度。

本研究閱讀了許多與研究使用的分析方法與資料的研究，以下的文獻探討部分，主要有本研究以所閱讀之相關文獻，並敘述出本研究認為其研究目的以及其分析結果進行的評論和與本研究比較後的看法，而本研究的相關文獻如下。

第一節 台灣經濟新報資料庫(TEJ)

壹、利用台灣經濟新報資料庫(TEJ) 所做之與研究目的之相關研究

台灣新報資料庫有 20 多年的歷史已經有許多的學者利用了該資料庫中所提供的資料做了相關的研究，而本研究為了瞭解研究主題的可行性，藉由參考其他學者的相關研究並希望可以藉由相關的研究確立可行性，並在閱讀其他相關研究的期刊論文來讓研究方法上更加的嚴謹，依照前面所述，本研究閱讀並參考了其他學者的研究，以下是與台灣新報資料庫有相關的研究，下列是講述本研究閱讀了這些相關研究後所認為其研究的目的，比對其研究結果的看法。

在一篇股權結構與信用風險、盈餘管理方面的研究中，本研究認為其目的在於公司股權結構對於信用風險盈餘管理方面的影響，認為此研究的目的有二，其一是當如果公司的信用風險較高時，其盈餘管理的程度是否較為嚴重，其二是股權結構對於信用風險和盈餘管理兩者間是否存在關聯性並互相有所影響，這兩個研究主題是在探討信用風險、盈餘管理、股權結構之間的關係，是屬於探討公司內部的經營績效方面的研究，和本研究主要探討公司外部投資者的方向較為不同，但是其中也是利用了台灣新報資料庫所提供的資料來做主要的分析，而此研究的研究結論是認為，此企業的信用風險越高的時候，公司的管理階層基本越有可能會進行盈餘管理，反之信用風險越低的時候越不可能進行盈餘管理，所以信用風險是決定公司是否進行盈餘管理的關鍵，可以看出此研究的研究結果是指信用風險與盈餘管理是主要有相關影響的，而在股權結構方面較為獨立對於股信用風險與盈餘管理並沒有非常顯著的影響（劉彥甫，民 101）。

在利用模糊平對電子公司進行財務危機預測的相關研究中，本研究認為次研究跟本研究一樣是研究台灣電子業方面的研究，但是主要是研討模糊平等趨勢對於電子業股票市場的財務危機方面的研究，其研究目的有二，一是使用了模糊邏輯的方式消除傳統評等方式所可能造成的誤差值，二是以歸納的方法了解財務危機發生之模糊趨勢與規則，此研究主要是探討利用較為模糊有容錯空間的方式來進行財務方面的預測，希望可得知在模糊邏輯的方式下，財務預測是否會更加準確，此研究的研究成果方面得到了含有誤差的分數加總而得知評等分數來進行預測，則發生誤測的機會更大，代表較為模糊也就是較不精確含有誤差的數值對於預測股市趨勢的準確度會較高，而第二個研究結果方面，模糊評等趨勢對台灣上市電子公司呈現良好的財務危機預測結果，所以也證實了此研究對於第一個研究成果方面的結論，含有誤差除了對於預測趨勢有所改善也對於財務危機有所改善，此研究利用了模糊邏輯的方式來進行財務分析，並藉由取得資料的樣本來確定準確度，和本研究用資料包絡方分析法取得各公司的經營績效有相同點（黃焜煌、卓統佑，民87）。

在一篇以電子公司合併及分裂的宣告方面的研究，相較於本研究都相同是對台灣電子公司進行研究，但是除了上市公司也包含了上櫃公司的部分，其研究目的有四，一是用來分析臺灣電子產業購併宣告之的效果，二是比較臺灣電子產業的購併宣告中，目標公司所位於國內和國外的不同處，三是探討臺灣電子產業分割宣告的效果，四是比較臺灣電子產業購併與分割宣告效果的差異，在這的時期台灣中小型的電子公司也逐漸的沒落，所以不乏各公司間產併購等事件發生，而併購不論是國內間的互項併購，也包含了國外併購過內的事前發生，所以此研究依照國內外併購見的差異，來證明台灣市場的狀況，來分析出台灣市場的因為併購所發生對於市場的效應，在研究成果中，其一證明了證明臺灣股市中已經是半強式效率市場的判斷已經成立，並依照國內外的交易市場的比對台灣的股市市場確立台灣已經屬於半強式效率市場，而成果其二購併宣告前股市會提早反應的影響，但是在獲利性並沒有存在顯著的差異，依照研究成果指出台灣市場會對於要併購公司的財務狀況

出現反應，但是這些反應不會對於獲利方面有所影響，研界結果其三是對於平均累積異常報酬方面，整個事件觀察期皆無顯著之反應，但由於報酬率呈現負向，代表雖然異常報酬沒有大幅度的變動，但是對於整題來說，報酬卻是有再下降的，研究成果其四，投資人認為購併宣告效果之影響大於分割宣告，身為投資者一家公司在要被併購比起分割要更加的重視，影響效果較大（蕭文姪、顏慧明、謝昌隆、周德佳，民 103）。

在一篇在用資料探勘手法建立預測破產模型的研究中，本研究認為此研究的目的有三，一是比較不同類型的屬的預測績效探討屬性篩選對於預測正確率的影響，二是經過篩選找出更具代表性的指標，然後再進行訓練與測試模型收斂出更具代表性的財務指標，三是分析 AUC 以評估何種模型較適合破產預測，此研究主要是利用破產預測模型來預測公司破產的可能，而也利用不同類型的財務指標來瞭解利用財務指標來預測公司的破產可能性，而此研究的研究結果在財務指標方面沒有明確的指出何種財務指標有明確預測破產的可能性，而在研究的模型方面提出了 3 種結果其一是簡化維度可得到更高的準確率與降低錯誤率，其二是財務指標加入做機器學習優於其他的分類組合其三是 CART 與 MLP 的表現接近，考慮機器學習時間成本推薦使用 CART，其中大多是建議其他研究者在做相關的研究可以試著簡化維度和加入機器學習方面來改善研究的結果與方式（鄭茂松，民 105）。

在在用多種資料探勘技術建立股票投資模型的研究中，本認為此研究相較於本研究的題材較為相近，跟本研究一樣利用了資料包絡分析法作為主要研究的分析方法之一，此研究的提出的目的有五，一是運用資料包絡法評估候選股票之發行公司財務狀況以作為選股之參考，和本研究一樣是利用包絡分析但是相較本研究此研究主要是一樣使用資料包絡分析來預測經營效率但是投入產出方面變數較少，二是以蜂群演算法來求出投資模型中各檔股票之投資權重，以蜂群演算法分析可以知道各項數值對於投資方面所佔的比例，三是利用類神經網路預測股票之收盤價以便對各檔股票次周之漲或跌作出判斷所以可以知道個投資者以收盤價來判斷是否購買股票的狀況，四是制定四個股票買進與賣出之策略，配合類神經網路所建

構之股價預測模型以決定最佳之股票買進或賣出之時機，此研究利用了分析的結果和類神經網路來制定股票買賣的較好的組合與時機，建議投資者如何購買股票比較符合效益，五是透過計算在投資時間區間之整體投資報酬率，並與股票大盤投資報酬率或銀行定存利率進行比較，以驗證此研究所提出的整合性投資策略之可行性與有效性，而本研究是以資料包絡分析法和決策樹進行研究在以隨機森林和相關分析等進行驗證，而此研究單純以投資報酬率和銀行定存利率進行比較得出正確率，此研究的研究結果有三，其一為運用資料包絡法來作為選擇股票購買的依據並選擇出績效較為優良的股票，再透過蜂群演算法的分析方法來得出各股的投資效益權重，並且在利用類神經網路的分析方法來預測出股票的收盤價來針對股票的漲跌作出判斷，並制定出有效的交易策略來實際配合股價預測的模型來決定買進或賣出時機，並試著藉由這樣一連串的動作來判斷除選擇績優股的有效策略以及依據，其二為以四個個案為例，其中各案一結果是報酬率優於台灣銀行定存 6 個月的報酬，但是其餘三個各案例實證的比率均遠低於台銀定存 6 個月的報酬，尤其是在跌幅近 3% 時間點，可以看出台灣銀行定存 6 個月利率較佳，這個結果表示對於較為保守投資人而言，可以將手頭現金定存銀行這樣報酬更為穩當，並且不和事作任何的股票的投資行為其三是此研究所建構之投資組合最佳化程序動作，確實可以協助投資人於大盤類股趨向於正報酬時可以優於大盤類股之績效，但於大盤類股趨向於負報酬時，也可以達到風險控制之目標，此研究所提供的四個研究結果表明研究所分析出的模型與財務指標的確可以準確的分析出股市的漲跌與分析數值的關係（徐志明、彭育秀，民 102）。

在一篇影響傳統財務指標的研究中，本研究認為此研究提出的研究目的有二，其一是以傳統的財務指標及影響傳統的財務指標的相關因素，來進行兩組樣本排序並已資料的極值得分點進行差異的分析，判斷其研究分析出的結果是否會有顯著的差異，其二是試圖從績效方面的評估與相關理論的質性探討，次研究運用了一套完善的統計方法，整理並試著找出衡量績效的因素指標，並透過分析計算出各項指標影響的權重比例，以試著建立一個可以解釋、評估和預測績效的結果，並讓可以作為後續用於績效評估的研究及企業投資者的參考依據以及運用方法，此研究利用了各種分析方法來分析出影響績效的財務指標和因素指標，所以與本研究的目的相同以資料探勘手法來找尋影響公司績效的指標，但是此研究也經由相關理論來建構出預測公司績效的方法，此研究的成果在於尋找出影響績效因素指標和績效指標之權重，其中影響績效因素指標依序為每股比率、其他比率、獲利能力、成長率、成本費用率、償債能力、經營能力，績效指標之權重依序為其他比率、每股比率、成本費用率、獲利能力、償債能力、成長率、經營能力（黃尹佐、陳筑螢，民 96）。

貳、台灣經濟新報資料庫(TEJ) 介紹

台灣經濟新報資料庫(TEJ)，是成立於 1990 年 4 月間，主要是屬於金融財經類別的資料庫，是一種專門提供證券金融市場基本分析所需的資訊。

參、台灣經濟新報資料庫(TEJ) 經營宗旨與資料提供國家

資料提供國家：

表 2-1 台灣經濟新報資料庫(TEJ) 資料提供國家

台灣	中國	香港	韓國	泰國
新加坡	馬來西亞	菲律賓	日本	

肆、台灣經濟新報資料庫(TEJ) 主要項目與資料庫內容簡介

表 2-2 台灣經濟新報資料庫(TEJ) 主要項目

財務資料庫	原始財報	資本形成及股利
公司相關資訊	股價資訊	總體經濟指標

資料庫內容簡介：

Profile：總體經濟學、原物料、匯率、利率等。

Company：月盈餘、產銷組合、董監事、資本形成、長期投資、短期投資等。

Finance：上市櫃財務指標，報表。

金融業：自有資本、資產品質、自有資本明細等。

金融控股：子公司營收、月產銷組合、轉投資、財務報表。

TEJ 股價：股價、信用交易、外資法人、集保庫等。

TEJ 審計品質分析資料庫：影響變數、衡量變數、事務所相關統計。

TEJ 公司自理：集團組成、持股、董監酬勞、社會責任、董監事兼任情形。

TDR：公司基本資料、收盤價、財務、證券進出明細。

FUND：基金資料庫。

PUB：財務資料、股價、董監事、月營收、資本形成。

第二節 資料探勘(Data Mining)

以下提到有關資料探勘相關之研究皆為本研究進行參考並閱讀後，所整理並論述其目的及表達其與本研究之同異性及目的與結果之看法。

壹、各學者資料探勘資料探勘(Data Mining)之定義

本研究是以資料探勘的手法來進行分析、不論是資料包絡分析法、決策樹、隨機森林等都是資料探勘的分析方法，而何為資料探勘，本研究試著依照各方的學者論文中所提到的定義來進行說明，其中有學者在研究中提到，資料探勘是從大量的資料庫中不斷探索，以發掘隱藏在資料中的知識與規則(蔣德威、王佩琪、簡春雄、阮定憲，99)。許多國外學者也在研究中提到了定義了資料探勘的部份，認為資料探勘是依據使用或研究學者用的分析方法所需要的資料，從使用或研究學者所選定的來源資料庫中選擇出研究用的分析方法合適使用的資料，並加以經過處理、轉換和探勘甚至是評估的一連串的動作(Fayyad, Piatetsky Shapiro & Smyth, 1996)。學者 Berry & Linoff 在研究中定義了，資料探勘是針對較為大量的研究所需的資料，在利用了自動化或者是半自動化的分析手法來進行分析，並以尋找出有意義的或者是有研究成果的關係或者是法則(Berry & Linoff, 1997)。Cabena et al 學者定義的資料探勘是將先前是未知的、且有效的資訊從大型的資料庫中抽取出的過程，並且將抽取出的資訊中有用的部分提供給決策者來做出決定性的決策。(Cabena, Hadjnian, Stadler, Verhees, & Zanasi, 1997)。學者 Han and Kamber 定義資料探勘是從儲存在可以用於儲存體大量資料的資料庫、資料倉儲或者其它資訊媒介中，發現特別的特徵值的一連串過程。(Han & Kamber, 2000)。學者 Fayyad 等人定義資料探勘是協助使用者或研究者找出想要知道的資訊或資料之間的特徵值(Fayyad et al, 1996)。學者 Frawley 認為在過去的文獻中對於資料探勘的定義為從資料庫中抽取出不明確、前人未知或是存在有用的資訊這一連串的過程(Frawley, 1991)。學者 Grupe and Owrang 定義資料探勘是指從已經存在的資料庫當中挖掘出專家仍未知的新事實。(Grupe & Owrang, 1995)。

在了解眾多學者對於資料探勘的定義後本研究認為眾多學者對於資料探勘的定義大多是指從資料集中利用各種方法來取得有一定意義的成果，而套用現在的大多人對於資料分析的用語，應屬於大數據(Big Data)分析，故本研究趨向於使用於大數據分析，這樣的說法會導致範圍較為受到局限，所以還是傾向是指資料被從數據庫中抽取並進行研究的一連串過程。

以下為本研究依照資料探勘的過程製作了流程圖，又被成為資料庫知識發現(Knowledge Discovery in Database, KDD)，是資料探勘的過程，以下是本研究所做的資料探勘流程圖：

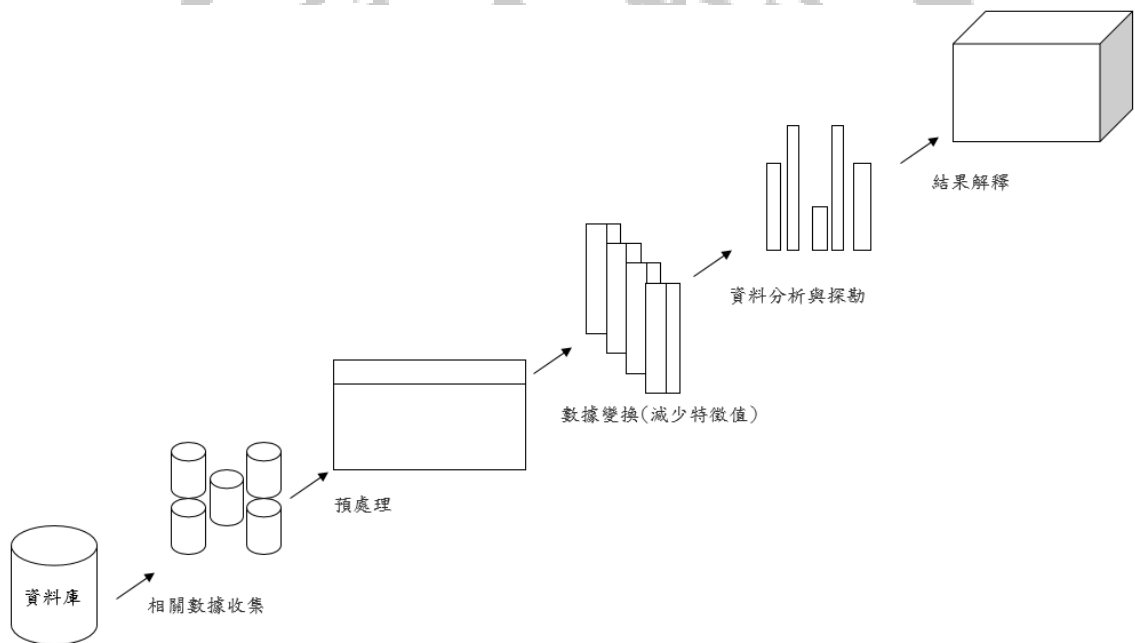


圖 2-1 KDD 流程圖

本研究以此流程圖來表示資料探勘 KDD 的步驟，其一是從資料庫擷取相關的資訊，其二是將資料做預先的處理將資料中遺失、重複、不一致、不完整的資料進行刪除，其三是將資料中的特徵值減少，讓可以再進行分析時可以有更佳準確的結果，其四是利用分析手法譬如，決策樹、分群等方法進行資料分析，最後再進行解釋來試著讓資料變成有用的知識。

貳、資料探勘(Data Mining)的步驟

本研究除了 KDD 的步驟講解，也額外講解資料探勘較為細部的部分，相關講解如下。一般而言大部分研究是將資料探勘分為四的步驟來進行：

第一步是資料整合與修正，其中是將資料庫中的資料進行抽取，並將錯誤或著遺失、重複、不一致、不完整(存在空值)的資料進行刪除。

第二步是資料選擇與轉換，對於抽取出的資料中相關的欄位進行資料格式的轉換來符合接下來的研究，讓研究可以順利進行。

第三步是粹取的動作，其一是選擇適當的資料探勘方法來進行分析，譬如資料包絡分析法、決策樹等都是資料探勘的分析方法。

第四步是評估和展示解藉由資料探勘所分析出來的資料進行檢視與判斷，並依照研究的目的去嘗試解釋並依照分析出的結果提出解釋，並了解到是否能和預測的結果相符。

此資料探勘的步驟是依照李建億和蔡芳遠的研究文獻中進行發想，已此研究所提除的相關步驟再經由前面各方的學者所提出的定義加以修改，並希望可以更貼近資料探勘研究方面的較為容易理解的狀況（李建億、蔡芳遠，民 93）。

第三節 資料包絡分析法(DEA)

壹、使用資料包絡分析法(DEA)所做之相關研究

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是否有受到相關文獻的支持是需要考證的，也因此本研究尋求利用資料包絡分析來進行分析的相關研究並藉由這幾篇研究來確定分析出目標變數的可行性，而這些研究大多都屬於分析於投資標的或財務指標等盡量貼近本研究的主題。以下就是本研究所參考的相關研究，在應用資料包絡分析選擇股票投資標的之研究——以上市食品產業為例的研究中，利用的資料包絡分析來分析出較為影響經營績效的財務指標和投資標的，此研究的研究目的有二，其一是經由資料包絡分析，分析食品產業的企業經營績效且尋找出影響經營績效的財務指標，並提供金融投資者在選擇股票投資目標上的一個參考方式，其二是透過此經營績效與股價報酬方面的分析，讓投資者能夠較為容易的評估想要投資的標的，所以此研究的研究步驟第一項為利用資料包絡分析法來分析出可以影響經營績效的財務指標，第二個是以經營績效和股價報酬的分析結果，確立可以讓投資者較為容易評估投資標的方法，此研究的研究結果是較為偏向研究限制的方向進行分析，其研究成果是，以食品產業研發了新的具有核心競爭力的產物，可能會對於營業狀況造成瞬間並有相當規模的影響，因此若可以加上質化資料方面的分析，將可以有利於投資者利用研究成過進行第二階段的決策分析，且選擇除正確的股票投資標的，而此研究中標明了利用質化資料較有利投資者進行正確的投資分析（劉文斌、左晉璋、彭克仲，民100）。

本研究在參考在應用資料包絡分析法來探討職業球隊投入成本與績效之關係這篇研究中發現，在資料包絡分析法本身通常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，為此，此研究的目的為應用資料包絡分析來了解並探討出投入與產出之間效率，並利用了資料包絡分析來瞭解美國三大職業運動球隊的選手薪資狀況，其中以球員的薪資與其他相關的費用來做

為投入，而使用薪資投入後所得到相對的成果也就是球員對球隊的勝率、價值與收入的影響做為產出，而在研究的成果方面得到球隊在與球員薪資和相關費用的花費上必須去注意到其投入成本與球隊價值之間的效率，如此才能獲得更能符合成本效益的結果，而此研究是以國外得球隊分析數據來做為分析資料，也在分析結果上有了結論，球員的價值的確是影響球員薪資的主要變數（吳翠治、齊德彰、李科翰，民 100）。

在以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本研究
的看法是認為，在臺灣的道路維護一直是大多人之間的詬病，而此研究以臺灣道路
為研究主題，利用了資料包絡分析來分析出臺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，但是也說明了臺灣道路維護上經費
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路
段的執行績效也就是養護的成果，來針對各項的資源投入及產出成果，做相對有效
果的比較分析，並讓相關組織作為檢討依據與方向，在預算資源分配方面的分析結
果作為公路養護政策的方向給予參考，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、產出的項目中應改要有所變動，在確認傳統資料包絡
分析在採用了預算項目有重複計算部分應該要禁行刪除，但是這屬於研究限制的
方面，而其二是在各地區所花費的鋪面瀝青修復經費，其單價並沒有的明顯差別，
這是代表以經費來看各工務路段的努力之產出程度也就是養護的效益並沒有地域
上的差別，所以此研究以鋪面瀝青修復經費來分析出各地的維修狀況並沒有因為
部分地區所獲取的資源或經費較多而有較為良好的狀況，所以經費多寡不是道路
修繕是否完好的主要原因，所以可以經由這方面所得出的數據來給各地機關給予
養護工經費上的建議，並發揮出其應該要有的效益，讓投入產出可以達到平衡（李
綱，民 101）。

貳、資料包絡分析法(DEA)之定義

學者徐志明、彭育秀在研究中提到資料包絡分析法主要是以經濟學中的柏拉圖最適境界(Pareto optimality)，做為主要觀念來作為效率方面的判斷，並且將所有作為主要資料的分析單位中的資料分為投入和產出，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(徐志明、彭育秀，民 102)，學者劉文斌等人的研究中提到資料包絡分析法將資料分為投入和產出的觀念是起始於 Farrell (1957)所提出的非參數邊界分析這是一開始最為傳統的決策樹分析，後續的學者 Charnes, Cooper, & Rhodes (1978)年利用了納入數學規劃的技巧在 Farrell 在 1957 年所提出的確定性非參數法中，在所有受評估的決策單位中建構了效率前緣，並在位於邊界上的決策單位即是為最有效率的數值，而未落在邊界上的決策單位則為無效率的數值，且在此效率是一種相對的效率概念(劉文斌、左晉璋、彭克仲在，民 100)。

參、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，本研究認為資料包絡分析法是用於績效評估的工具之一，其目的在於希望透過決策單位來衡量多項產出與多項投入間相對的表現，並透過權重計算各決策單位之間的產出與投入比值，但是因為資料包絡分析是屬於較基礎的財務分析方法，所以也就會有許多學者對有不同的看法，所以本研究參考了其他學者對於資料包絡分析研究與看法，以下就是其他學者對資料包絡分析的看法。

學者胡志堅、黎漢林在研究中提到了資料包絡分析法是可以處理多項的投入、產出在評估方面的問題，在其評估效率的結果視為一項綜合的指標，容易且可以用於評估企業的績效表現在（胡志堅、黎漢林，民 90）。

學者王艷輝、郭亞軍、王麗在研究中認為，資料包絡分析法避免了進行分析的決策者給予分析結果主觀上所出現的問題，而且利用資料包絡分析法對投入產出的指標有較大的包容性，因此這種分析方法處理並評價問題時比一般常見的統計方法更加優良（王艷輝、郭亞軍、王麗在，民 93）。

肆、資料包絡分析法(DEA)的兩種模式

學者 Charnes 等人在 1978 年在 Farrell 學者的理論之下，提出了 CCR 模型，而這個 CCR 模型就是資料包絡分析法最初的模型，而 CCR 模型也是現今最多人使用的在管理科學的模型，而本研究的資料包絡分析法就是使用 CCR 模型，CCR 模型可以最直接獲得投入產出之間效率的數值，對於財務研究方面，可以進行非常直觀的應用，而以下就是 CCR 包絡分析模型的介紹，分為最初的 CCR 模型以及後來考慮到規模報酬的 BCC 模型(Charnes et al, 1978)。

資料包絡分析法主要有兩種模型其一是 CCR 模型，學者羅瑞霖、廖冠傑，在 2006 年的研究中提出，CCR 模型主要是以固定規模報酬來做為最基本的基礎，生產邊界是以線性的規劃方法並可以去計算到每一個決策單位的相對效率，而如果越靠近邊界值代表效率越好，也就是越接近 1，而如果沒在邊界上代表效率越差，所以 CCR 模型的效率值都在 0 到 1 之間，而資料包絡分析法中的 BCC 模型是學者 Banker 等人於 1984 年提出，BCC 模型主要的功用是用來彌補 CCR 模型的不足，在最初始的 CCR 模型是以固定規模報酬為基礎提出的，但是在現實中投入和產出的情形也有可能是屬於規模報酬遞增或規模報酬遞減的狀況不一定是固定規模報酬，所以 BCC 模型是可以幫助使用者，可以檢視自己投入產出方面的配置缺失(Banker, Charnes & Cooper, 1984)。

本研究繪製了資料包絡分析法的示意圖，讓資料包絡分析法可以圖形化更容易理解，下圖就是資料包絡分析法的示意圖。

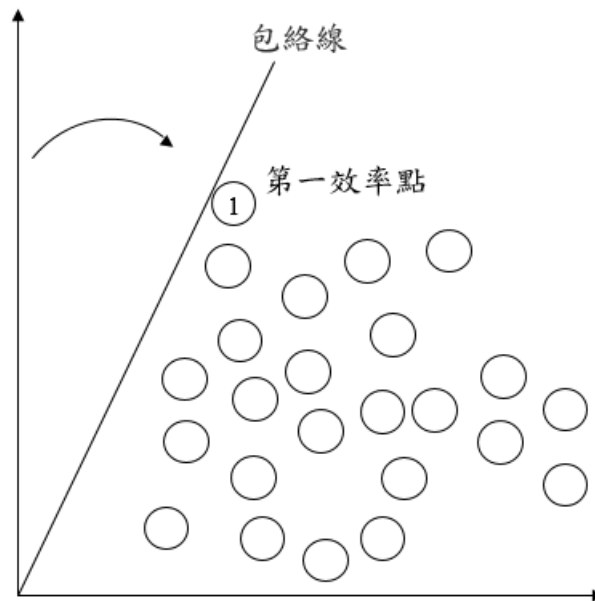


圖 2-2 資料包絡分析法的示意圖

資料包絡分析法是一種評估單位效率的方法，如圖解釋因為，其中產出越大投入就越小，所以產出/投入所形成了線代表效率最大值，此線被成為包絡線，因此要尋找最有效率的決策單位，包絡線會以順時鐘移動，直到與第一個決策單位碰觸才會停止，次單位為有效率的決策單位數值為1，並以此點推算出其他決策單位的效率。

第四節 決策樹(Decision tree)

壹、使用決策樹(Decision tree)所做之相關研究

以下提到有關決策樹相關之研究皆為本研究進行參考並閱讀後，所整理並論述其目的及表達其與本研究之同異性及目的與結果之看法。

決策樹常常被利用於各種分析中，是一種很典型的資料探勘新方法，其中資料分析，本研究也參考了決策樹在財務分析方面的研究，以下是各學者對於決策樹分析方面的相關研究，決策樹不只是可以運用於特定領域譬如研究，以決策樹探討連續處方部分的影響的研究中提到，決策樹作為資料探勘常用的分析方法，可以用於各種領域學科上的分析，其中此學者利用了決策樹作為主要的分析方法，探討如果未開立連續處方簽的因素（蔡佳玲、洪新原、袁繼銓，民 93）。

本研究認為，決策樹是一種資料探勘的方法，經由決策樹分析資料可以知道該資料中有那些部分可以改進，有學者在研究中以建構模糊決策樹研究半導體部分的資料挖礦用來提昇良率而研究此研究藉由決策樹分析半導體產業關於良率改進方面的議題，半導體業一直是支撐台灣經濟的一個重要產業，其中半導體製造上的良率一直是一個問題此研究設法利用決策樹來解決這個問題。(簡禎富,林昀萱,鄭仁傑,2007)，有一篇用特徵選取呵資料包絡分析法以決策樹來探討基金部分的研究，這篇研究與本研究有相似之處，同樣是使用資料包絡分析法來取得目標變數，但是本研究是研究公司經營效率上的問題，而此研究是探討基金選購分面的問題，而此研究利用了資料包絡分析和決策樹的分析手法根本研究又相似之處，但是使用了分析手法較少（吳植森、廖柏森、余忠祐，民 99）。

在 2010 年的一篇研究中，使用決策樹來分析台灣電子股漲跌，這篇研究跟本研究一樣，選用了電子業股市為主要的分析對象，此研究主要是利用分析手法來探討電子業股票市場的行為模式，而本研究是在探討電子類公司獲利能力的指標，此研究與本研究選用的題材相同，但研究目的不一樣，但是都有使用的決策樹分析（蔣德威、王佩琪、簡春雄、阮定憲於，民 99）。

貳、決策樹之定義

本研究在決策樹方面的定義是以學者所講述的部分，再加入本研究的見解，並且將論述的部分加以講解。

決策樹的分析方法經常用於進行預測方面，而決策樹也可以是一種建立分類的方法，使用決策者所需要分析的資料利用歸納的方式來產生類似於樹狀結構分析方法，決策樹的歸納方法是將所有要進行分析的資料採用有規則的方式進行分類，並且藉由相似的特徵來進行分類並將相同的結果的放到同一類，因為決策樹的分析方法會將資料依照不同的變數以有規則的方式產生分析結果所以決策樹可以藉由此種分析方法來分析目標的特質和相同點(張俊郎、陳志豪、蔡玉純，民 96)。

有學者在 2013 年以決策樹的外觀來定義決策樹，此研究鎖定義的決策樹中，以是以樹狀的結構進行講解，在決策樹進行分支的節點中間是代表測試的條件變數也就是分類的依據，而節點下的分支代表依照條件變數分類後的結果，而決策樹的最後的節點則代表分類後所得到的分類的結果(翁政雄、洪令莊、呂培豪、陳學瀚、郭家佑、施博惟、謝孟哲，民 102)。

本研究在兩篇研究中，看到了兩位學者有差不多的見解，所以進行了統整，在這 2 篇研究中提到研究中提出目前的資料探勘分析方法中決策樹是非常常見的，而該方法吸引人之處在於有規則的利用樹狀圖來顯現規則與路徑，另外也因為是有規則的分析，決策樹可以用於文字方面的表達或是經由轉換變成資料庫語言，讓資料庫中有相同特徵的資料可以一起被收尋到，決策樹分析模式主要演算法包括三種方法 ID3、C4.5 及 CART 等(Han & Kamber, 2000)。

參、決策樹的使用

所謂決策樹在分析時要有目標變數，而決策數會以目標變數來進行資料的分類，並且藉由數值上的計算來進行分類，但是能夠進行分類的結果是有限的，資料量過少是無法進行分類的，所以到達分類極限時，決策樹分析就會停止。

肆、決策樹的修剪

在決策樹經由分析計算完成之後，可以再利用修剪的方式來讓所分析出的決策樹大小進行改變，因為如果決策樹分析的大小過大，就很容易會造成過度學習的情形發生，所以在修剪決策樹方面適當的修剪是有助於決策樹類進行分類的，而決策樹的優點是可以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、決策樹的示意圖

本研究為了瞭解決策樹分析的分析結果取得的分析方法，特別繪製的示意圖，讓決策樹可以更好的理解，以下是本研究做作的決策樹示意圖。

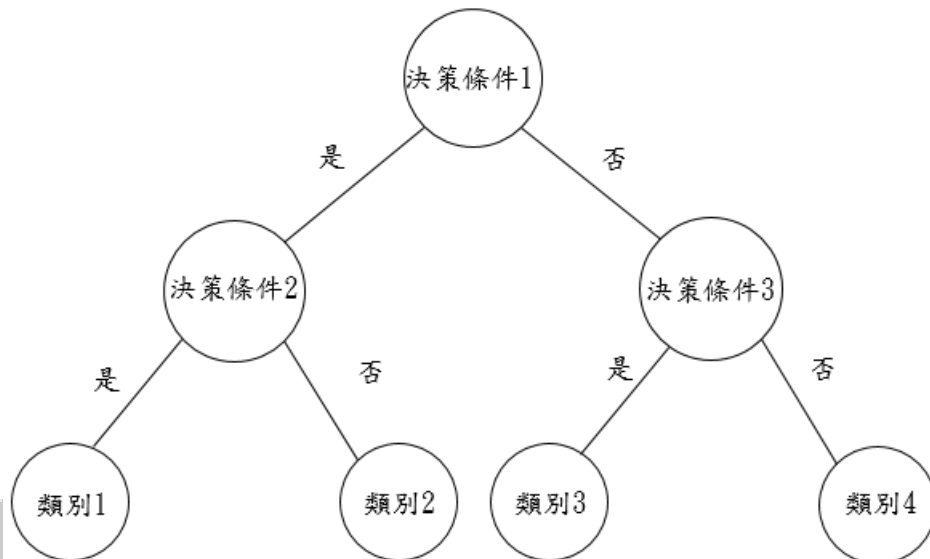


圖 2-3 決策樹示意圖

決策樹在輸入資料後會把資料分析成樹狀圖，其中資料受決策單位判斷不成立的會往右，成立的會往左，而相較輸入的目標變數越大的值會往右，本研究是用 DEA 作為目標變數，代表類別 4 是獲利能力最好的單位，而以上的決策單位就是影響獲利能力的主要條件，也就是本研究的結果。

第五節 相關分析(Correlation Analysis)

壹、使用相關分析(Correlation Analysis)所做之相關研究

以下提到有關相關分析之研究皆為本研究進行參考並閱讀後，所整理並論述其目的及表達其與本研究之同異性及目的與結果之看法。

相關分析常常用於對比各項變數上的關聯，是在資料探勘中屬於較為基本的分析手法，在大多數的研究中都會使用到相關分析，但是在財務分析相關的研究卻少之又少，而以下就是本研究參考各學者對於相關分析研究的部分：

本研究找尋了兩篇有關於相關分析的研究，一是研究大專院校柔道教練與選手的組織承諾，二是運動教練壓力與組織承諾的相關研究研究，這些研究對於相關分析方面的參考資料都屬於運動產業方面，雖然都不是用於財務分析，但是相關分析都屬於變數上的對比，對各項研究領域都有分析價值，分析領域較為廣泛，所以本研究也希望透過相關分析來了解藉由財務指標所分析的結果，在藉由相關分析是否可得知各數值之間的差異或者是關聯性（陳南福、陳明宏、湯湧，民 102；蔡育佑，民 95）。

貳、相關分析(Correlation Analysis)之基本定義

相關分析通常被稱為 pearson 相關，是利用測量兩個變數之間，是否有相關性的分析方法，本研究認為相關分析可以很直觀了解各數值之間的關聯性，所以也可以來做為財務分析，本研究覺得如果可以找出一個較為獨立的財務指標，是否就可以更加準確預測財務分析的結果。

第六節 隨機森林(Random forest)

以下提到有關隨機森林之研究皆為本研究進行參考並閱讀後，所整理並論述其目的及表達其與本研究之同異性及目的與結果之看法。

壹、使用隨機森林(Random forest)所做之相關研究

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，大部分的研究都是利用隨機森林可以比較個像數值的特徵值來做出比較來進行研究，以下是各學者對於利用隨機森林進行分析的相關研究。

在一篇研究中，用了隨機森林和轉導推理的方法做特徵提取，這是一篇主要以隨機森林為主要研究分析方法，此研究的研究目的為利用隨機森林的中的相似性矩陣讓資料分為訓練樣本和測試樣本來進行轉導推理分析，並在此分析的基礎上來進行多維尺度來變換並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，利用了隨機森林分析所得到的低維特徵的方法得到了良好的結果，此方法將帶有標籤的訓練樣本中全部資料和未標籤的測試樣本的屬性都充分利用到了，不僅是降低了輸入訓練數值上的的輸入數值，更空間分佈方面的表達能力有更佳優良的表現，在此研究成果上改善了模型的泛化誤差，證明了這種特徵提取方法是可行且良好的（邱一卉、米紅，民99）。

本研究發現一篇與本研究像是的研究是使用隨機森林來研究選股的分析，跟本研究的目的是以分析出所未來價格會成長值得投資的股票，並提出一種新的選購股票的分析方法和指南，此研究是以隨機森林算法為主要，來建構可量化的選股模型，並解決股票投資上選擇購買股票的困難點，而在此研究中可以明確的從研究目的了解到此研究是利用隨機森林來建構出可以量化選股的模型，此研究的研究結果較為屬於講解研究限制，此研究的研究結果，其一是在進行選購股票時，如果所選購股票可以是同一種類別，可以剷除產業之間的差異，讓在選購股票的投資者可以不考慮到產業之間的差異所帶來的誤差，其二是在研究中，有夠過兩個不同產業之見進行對比，讓普適性值更容易進行分析，其三是在選擇股票樣本得時候，過大的樣本造成製造業上的統計分析較為困難（曹正鳳、紀宏、謝邦昌，民103）。

在一篇隨機森林選股的應用研究中提到，此研究的目的是使用隨機森林的分析方法達成量化選股的研究，跟本研究和上一篇曹正鳳,紀宏,謝邦昌在 2014 年的研究的目的與相同，但是此研究的國家是中國而此研究主要是選用了中國較為成熟的股票是進行研究，並希望可以藉由各財務指標間的相關性來制定選股系統，並且是使有隨機森林分析來建構選股模型，想因此來讓研究成果可以適應中國的股票市場（王淑燕、曹正鳳、陳銘芷，民105）。

貳、隨機森林(Random forest)分析過程

隨機森林是透過資料的隨機分組，再藉由分組結果各別進行決策樹分析，並藉由投票的過程來進行變數的排序，本研究依照隨機森林的分析過程製作了流程圖，是隨機森林分析的過程，以下是本研究所做的隨機森林流程圖：

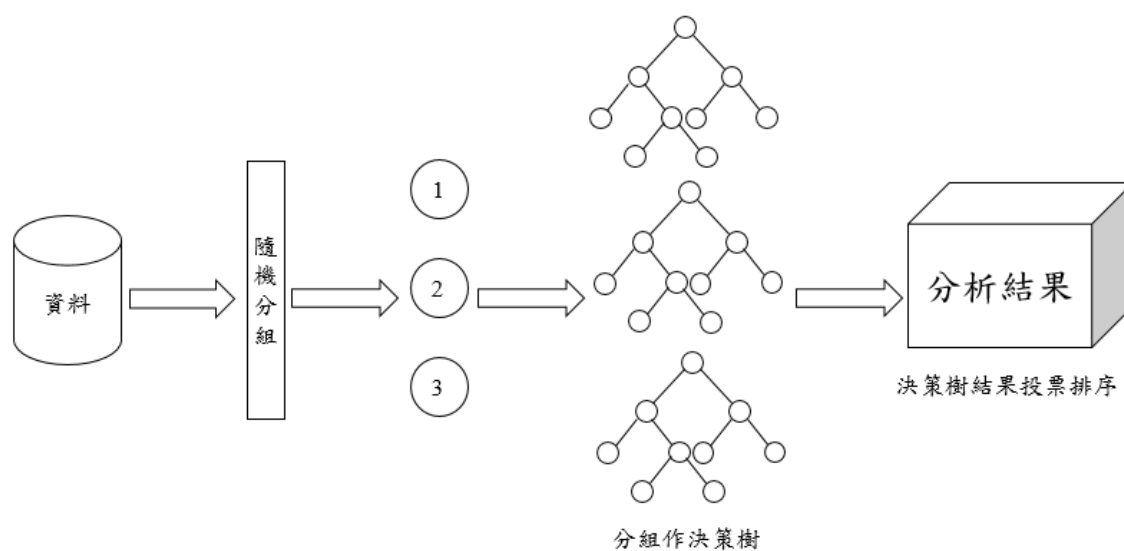


圖 2-4 隨機森林流程圖

參、隨機森林(Random forest)介紹

隨機森林是由 Leo Breiman 這位學者在於 2001 年所提出的一種組合分類器的分析方法，是由類似決策樹的分析方法來進行分析並投票出結果(Breiman Leo, 2001)。

第七節 分群分析(K-Means)

以下提到有關分群分析研究皆為本研究進行參考並閱讀後，所整理並論述其目的及表達其與本研究之同異性及目的與結果之看法。

壹、使用分群分析(K-Means)所做之相關研究

分群分析是使用是可以藉由各變數來對就對象進行族群的分析，分群的結果可以看有哪些研究對象屬於離群值，本研究閱讀了分群分析的相關研究，以下是本研究閱讀的分群分析相關研究，在分群分析來分析病人安全文化復原力的研究中，是以分群分析來分析醫療體系方面的研究，而此研究以分群分析來進行分析以了解醫療人員工作的現狀，此研究的研究目的如下，也可以說是研究流程，此研究利用了分群分析，來計算各集群的平均值，並分析出有集群高於平均值或是低於平均值有哪幾個，再透過了統計分析來計算出各個集群人口總量的變數進而分析出每個集群之間的特性，此研究的結果上表明，在 30 到 40 歲且已婚並且在工作的年資上較短並且就讀專科畢業又剛好在內科的加護病房工作的人口工作壓力較大（林苡甄、李怡慶、黃志璿、吳信宏，民 105）。

在以資料探勘分析選股模型的研究中，本研究發現此研究使用了決策樹、分群分析等分析手法，目的是以藉由分析來分析股市購買的規則，跟本研究的的研究方向大致相同，此研究使用 14 的項財務資料，並利用資料探勘的技術對研究樣本進行關聯法則、群集分析、決策樹分類法等分析，並且藉由分析的的結果進行歸納，並整理出容易理解的財務分析規則，讓選購股票的方式大幅度的簡化，此研究的研究成果有 3 點，其一是證實了依照分析的結果來選購股票，其獲利的程度比起大盤獲利的水準還要高，代表分析出的研究成果是可行並用現實的，在分析會影響上市公司的財務波動上了解股價的漲跌可以歸納出總資產報酬成長率、每股盈餘這兩項是較為影響財務波動的，所以此研究的目的上總資產報酬成長率、每股盈餘為主要進行投資時需要觀看的財務指標，而此研究也驗證了公司的財務報表的确會影響投資者們的投資意願，可以看出只要是財務報表數值優良的公司，獲得投資者的投

資是很自然的所以，股價是否會往上增幅主要原因是，公司的工作績效要好，繳交出良好的財務成績單，所以努力工作是為重要的（程守雄、陳亦伶，民 98）。

貳、分群分析(K-Means)的過程

研究依照分群分析的過程製作了流程圖，以下是本研究所做的隨機森林流程圖：

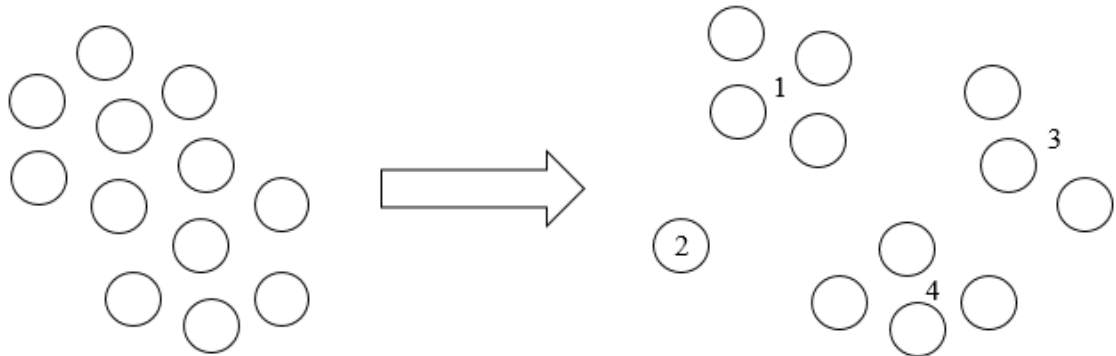


圖 2-5 分群分析流程圖

分群分析是屬於非監督式學習，所以分析時不會設定特定的變數進行分析，所以在分析者可以設定分析得群集數量，但是卻不能選定分群變數得情況下，分析者可以依照分群結果來進行推導，並了解在非監督式學習下，分群分析結果的背後，所代表的意義。

參、分群分析(K-Means)的介紹

分群分析是學者 MacQueen 在 1967 年時所提出的，一種分群模式的演算法則，是以固定的公式來進行分析，而要達到分群的最佳化，必須在分析前設定好 K 值 (MacQueen, 1967)。

分群分析(K-Means)的公式如下：

$$\operatorname{argmin} \sum_{i=1}^K \sum_{x_j \in S_i} \|x_j - u_i\|^2 \quad (3-1)$$

第三章 公司獲利能力財務指標分析方法設計

第一節 研究流程

本研究主要以電子業中，電子零組件公司做為主要分析對象，對象選擇上市公司為主要的分析目標，本研究選擇上市公司為主要研究目標原因是分析對象的經營規模差距太大會造成分析結果失去準確性。

研究的分析方法採用了資料包絡分析法、決策樹、隨機森林、相關、分群分析，以資料包絡分析法對公司的經營能力分析，得出目標變數，再經由決策樹分析出影響公司經營能力的關鍵變數，本研究以隨機森林來驗證決策樹分析結果。

本研究的以相關分析，分析各變數間的關聯性，以分群分析了解研究目標的族群組合。

1. 本研究一開始選定研究主題，希望以各種的財務變數分析出財務指標，而在研究目標資料的選定上，選定了以電子業中，電子零組件為主要的分析目標，在經由研究目標中選定較合適的上市公司作分析。
2. 分析所使用方法，參考眾多學者的研究報告選定了以資料包絡分析法和決策樹作為主要分析方法，後來希望藉由其他的分析方法來驗證資料包絡分析法和決策樹的分析結果，選定以隨機森林、相關、分群為驗證的分析方式。

下圖為本研究的研究流程圖如圖 3-1：

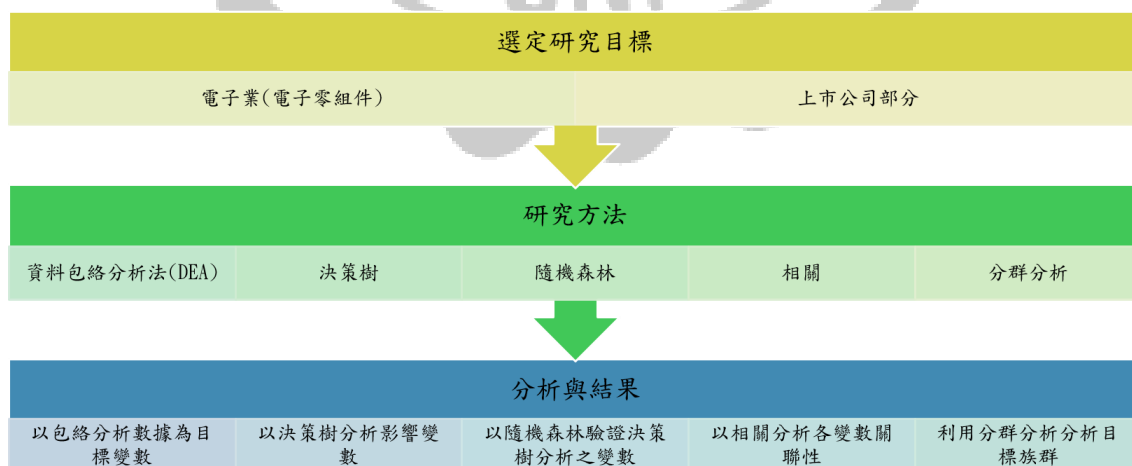


圖 3-1 研究流程

第二節 研究決策單位及主要資料說明

本研究的主要研究範圍為台灣 91 家電子零組件產業，整理為下表，下表為主要分析的 91 家公司公司代號以及公司名稱：

表 3-1 91 家電子零組件產業公司圖表

代號	名稱	代號	名稱	代號	名稱	代號	名稱
1471	首利	3037	欣興	2437	旺詮	6108	競國
1582	信錦	3042	晶技	2440	太空梭	6115	鎰勝
2059	川湖	3044	健鼎	2443	新利虹	6133	金橋
2308	台達電	3058	立德	2456	奇力新	6141	柏承
2313	華通	3090	日電貿	2457	飛宏	6153	嘉聯益
2316	楠梓電	3229	晟鈦	2460	建通	6155	鈞寶
2327	國巨	3296	勝德	2462	良得電	6165	捷泰
2328	廣宇	3308	聯德	2467	志聖	6191	精成科
2355	敬鵬	3321	同泰	2472	立隆電	6197	佳必琪
2367	耀華	3338	泰碩	2476	鉅祥	6205	詮欣
2368	金像電	3376	新日興	2478	大毅	6213	聯茂
2375	智寶	3432	台端	2483	百容	6224	聚鼎
2383	台光電	3501	維熹	2484	希華	6251	定穎
2385	群光	3533	嘉澤	2492	華新科	6269	台郡
2392	正崴	3550	聯穎	2493	揚博	6282	康舒
2402	毅嘉	3593	力銘	3003	健和興	6412	群電
2413	環科	3605	宏致	3011	今皓	6449	鈺邦
2415	錫新	3607	谷崧	3015	全漢	8039	台虹
2420	新巨	3645	達邁	3021	鴻名	8046	南電
2421	建準	3653	健策	3023	信邦	8103	瀚荃
2428	興勤	3679	新至陞	3026	禾伸堂	8213	志超
2429	銘旺科	4545	銘鈺	3032	偉訓	8249	菱光
2431	聯昌	4999	鑫禾	5469	瀚宇博		

本研究的資料來源為台灣經濟新報資料庫(TEJ)，分別為上述電子零組件公司 2013 到 2015 年的經營能力及獲利能力中財務分析資料之相關財務資料變數進行包絡分析法(DEA)運算。

第三節 資料包絡分析法(DEA)投入選項及產出選項

本研究以 DEA 分析法運用在電子零組件公司的多項財務指標，依照台灣經濟新報資料庫(TEJ)所提供的財務變數，分為經營能力與獲利能力來進行分析，而在分析前，先選取了相關的財務比率，並依照完整性、可衡量的、可解構性、不重覆性、最小化等五種特性，進行所選財務比率的刪減。

選取的屬性如下表所示：

壹、經營能力

1. 應收帳款週轉率：是指在一段時期內應收帳款轉化為現金資產的平均次數，一般來說，應收帳款週轉率越高，平均收賬期越短，說明應收賬款的收回越快。
2. 存貨週轉率：適用於反應商品存貨週轉的次數，一般來說存貨週轉的次數越多，代表公司的變現能力越高，並且也代表產品生產與銷貨速度是否達到平衡。
3. 固定資產週轉次數：固定資產週轉率可以反應出公司一定期間內的收入，可以的替換幾次固定資產，所以也可以反應出公司在一定期間內利用固定資產創造營收的能力。

投入敘述性統計：

表 3-2 經營能力變數敘述性統計表

經營能力變數	最大值	最小值	平均值
應收帳款週轉次	41.33	0.49	4.03286795
存貨週轉率	36,187.15	0	156.8349931
固定資產週轉次數	18,132.85	0.01	102.2449526

貳、獲利能力

1. 淨值報酬率：是一種用來測量公司獲利能力的數值，可以用來了解一家公司工
股東用所投資的資金到底產生了多少的紅利。
2. 營業毛利率：是指營業毛利額與營業淨收入之間的比率，主要用於用於對企業
獲利能力的分析，可以反映企業每一元營業收入中含有多少毛利額，但是獲利
能力不能單看營業收入，營業收入中還包括營業成本。
3. 已實現銷貨毛利率：公司收入考量銷貨利益後，以公司能賺得的金額來衡量
獲利能力的大小。
4. 營業利益率：公司商品或服務所賺的收入，再扣掉營運公司所需的費用，是指
公司靠主要收入所賺取的收入占有所有收入的比例。
5. 稅後淨利率：因其所表達的是每做 1 元生意，可有多少獲利。若依此準則，可
能導致大資產、小營收、高淨利率、低報酬率之經營結果。

產出敘述性統計：

表 3-3 獲利能力變數敘述性統計表

獲利能力變數	最大值	最小值	平均值
淨值報酬率	167.53	-197.94	9.673708475
營業毛利率	68.44	-94.71	17.03421477
已實現銷貨毛利率	68.44	-94.71	16.99104097
營業利益率	2,429.64	-6,152.21	2.203026846
稅後淨利率	53,874.90	-10,831.92	33.60304027

第四節 主要分析資料

在基於本研究使用了會計學中俗稱的三財表，所以必須證明使用三財表是有相關分析是有相關的正當性，所本本研究參考了在以往的研究中，有使用資產負債表、現金流量表、損益表作為主要研究資料的相關文獻，各學者利用三財表作為研究資料的相關研究如下：

在一篇以系統動力學做財務規劃的研究中利用了有三財表作為主要研究資料，此研究主要以系統動力學的方法進行模擬試算企業營運的目標，並以幫企業進行建立監度機制為主要目標，其中進行模擬試算是利用資產負債表、現金流量表、損益表作為主要研究資料（黃加賜，民 90）。

有一篇使用了損益表作為主要研究資料來預測獲利能力的研究，此研究將損益表進行分類並進行分析，並以分析結果的狀況來判斷以損益表進行財務分析的情形，此研究主要有三個分析結果，其一是以實證資料來輔佐損益表的分類進行證明的方式，是可以讓各別的盈利資訊相互的組成有持續性的資料，其二是在考慮各別的盈利資訊相互的組成的資料持續性的部分，在獲利解釋能力和預測能力方面都會比在沒有進行個別盈餘組成因素方面的顧慮時還要高，其三是在獨立的考慮到投資淨收入的部分之後，論獲利的預測能力將會比沒有單獨的考慮投資的淨收入時還要高（張仲岳，民 90）。

有一篇利用資產負債表分析銀行風險的研究中，發現，此研究的目的是探討全球商業銀行對資產負債表外業務活動對銀行風險的影響，而此研究利用的資料是資產負債中表業外的部分（吳孟紋，林怡茹，民 105）。此研究結果部分如下：

在發生金融危機的時候，商業銀行在調整資產負債表外業務活動的部分會有明顯的下降，而也因此可能會因此降低金融危機發生的風險嚴重程度。

在不同地區影響風險的因素不一樣，如在美洲地區的銀行進行衍生性金融商品業務會降低獲利波動風險，但在歐洲地區，這效果不明顯，亞洲地區，衍生性金融商品也會降低獲利波動風險及信用風險，但反而會增加破產風險呈現正相關。

資產負債表種有 19 種：現金及約當現金、短期投資、應收帳款及票據、存貨、其他流動資產、長期投資、固定資產、其他資產、資產總額、流動負債、長期負債、其他負債及準備、負債總額、資本公積、法定盈餘公積、特別盈餘公積、未分配盈餘、股東權益總額、負債及股東權益總額。

現金流量表 9 種：稅後淨利、來自營運之現金流量、投資活動之現金流量、融資活動之現金流量、本期產生現金流量、折舊、攤提、固定資產(購置)、支付現金股利。

損益表 11 種營業收入淨額、營業成本、營業毛利、營業費用、營業利益、利息支出、營業外支出合計、所得稅費用、本期稅後淨利、每股盈餘(元) 稅前息前淨利。

本研究是依照台灣新報資料庫中資產負債表、現金流量表、損益表所得之資料，並以企業的動態性財務規劃--以系統動力學方法初探中的建議，將變數加以簡化，檢選出其中 39 種變數為主要研究對象（黃加賜，民 90）。

本研究會以會計學俗稱的三財表為主要的分析變數，主要是以各項對於財經分析的相關論文都以三財表為主要的分析變數，本研究以資產負債表、現金流量表、損益表主要分析對象是為了確保分析結果的可用性以及準確性，其中本研究利用的資料，資產負債表 19 種，現金流量表 9 種、損益表 11 種共 39 種資料作為變數。

第四章 分析結果與解釋

第一節 資料包絡分析法結果與說明

本研究是以資料包絡分析法(DEA)的分析結果作為其他分析得目標變數，本研究的進行資料包絡分析法(DEA)是以 CCR 模式進行分析進而求得 2013 到 2015 年 91 家公司共 273 筆數值，資料包絡分析法中 CCR 模型的數學運算式如下：

$$E_K = \frac{\sum_{j=1}^n u_j^k Y_j^k}{\sum_{i=1}^m v_i^k X_i^k}$$

下表為 2013 年至 2015 年，91 家電子零組件公司財務報表經由資料包絡分析法 (DEA) 之結果：

從此表可以看出，各家公司財務報表中投入和產出部分的資料經由資料包絡分析法進行分析後的數值，表中的 DEA 代表該公司名稱中該年的經營狀況，在 DEA 前部分就是該 DEA 資料分析的西元年，而此表分析出的 DEA 數值分布，在圖 4-1 本研究會進行圖形化的整理。

從 DEA 的分析結果和決策樹結果營業成本、營業利益、每股盈餘中進行比較依照前 10 家公司的中決策樹 3 種數值的分配情況知道，在決策樹結果的數值特定情形下在個情形下 DEA 值會較高，就在營業成本數值較 91 家公司中數值較小，而營業利益及每股盈餘數值較 91 家公司中數值大這這種情況下 DEA 數值較大。

所以藉由 DEA 和決策樹的結果可以知道在財務報表中營業成本越小，營業利益及每股盈餘越大的公司財務狀況較為良好。

為了瞭解 DEA 對於股價的反應狀況，本研究以 DEA 中，4 個分段 8 家公司的 DEA 做成橫向長條圖，股票數據做成折線圖，進行比對了解瞭解 DEA 對於股價的反應狀況以下是股價資料的折線圖與 DEA 的橫向長條圖及說明。

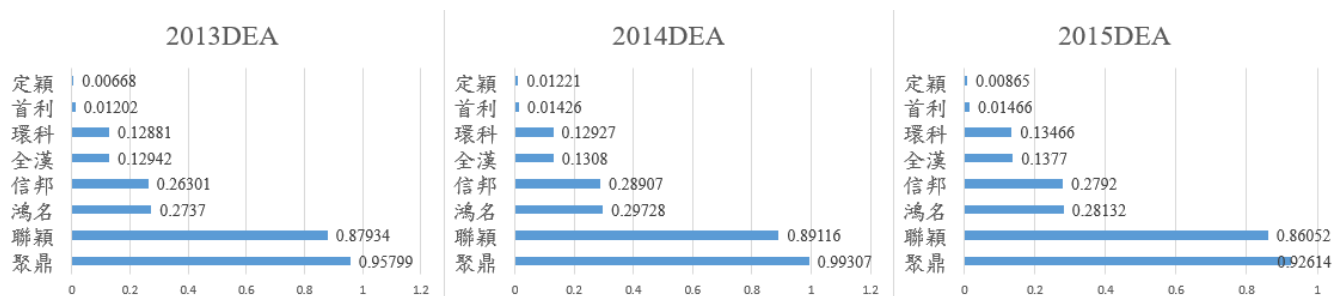


圖 4-1 各公司 DEA 示意圖

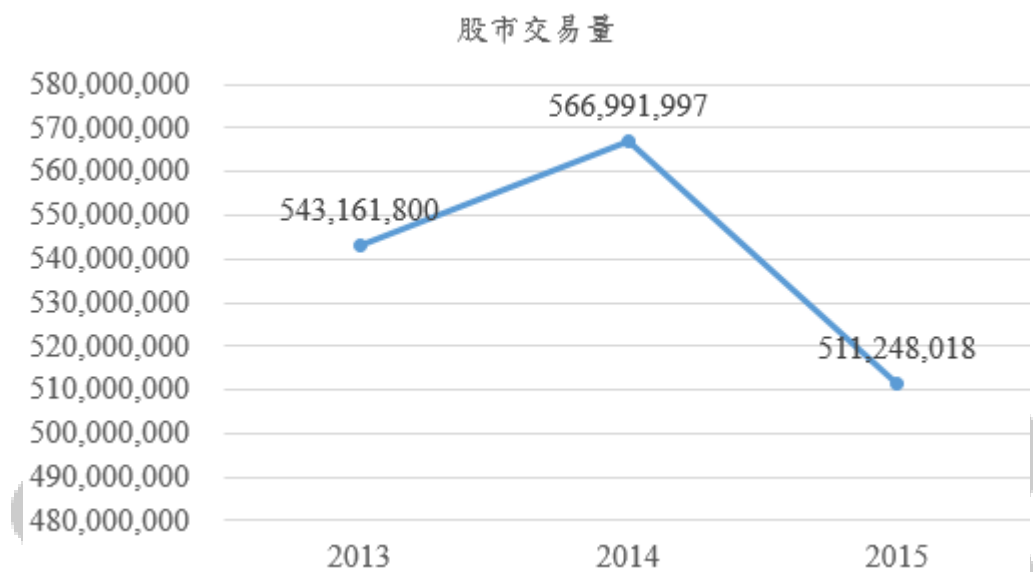


圖 4-2 三年股市交易情形曲線圖

DEA 是以年為單位，反應在股市也是以年為單位，從圖 4-2 可以知道股市從 2013 至 2015 年交易量是先漲後跌，並從圖 4-3 可知股價大多與交易量成正比，除了信邦是連續上漲，並以圖 4-1 可知 2013 至 2014 年的數值都是呈現與股票趨勢相符，而 2014 至 2015 年獲利能力較好的公司都是呈現跌幅，股票情況也大多是呈現下跌，但其中信邦所呈現的跌幅較小，股票情況卻呈現上漲，而獲利能力較差的公司 DEA 是有漲有跌，但是股票情況都是呈現下跌，代表以年為單位計算獲利能力的大小，能力大的較不受大盤趨勢影響，能力小會較容易受到影響，所以獲利能力小的公司獲利能力較不影響股價，而獲利能力大的數值如變動較小，就會顯現與趨勢相反的情況，但者也顯現出以年為研究時間單位不能非常準確股市漲跌，所以依照信邦的情況可以了解，能以更小時間單位研究，可能就可以預測得更加準確。

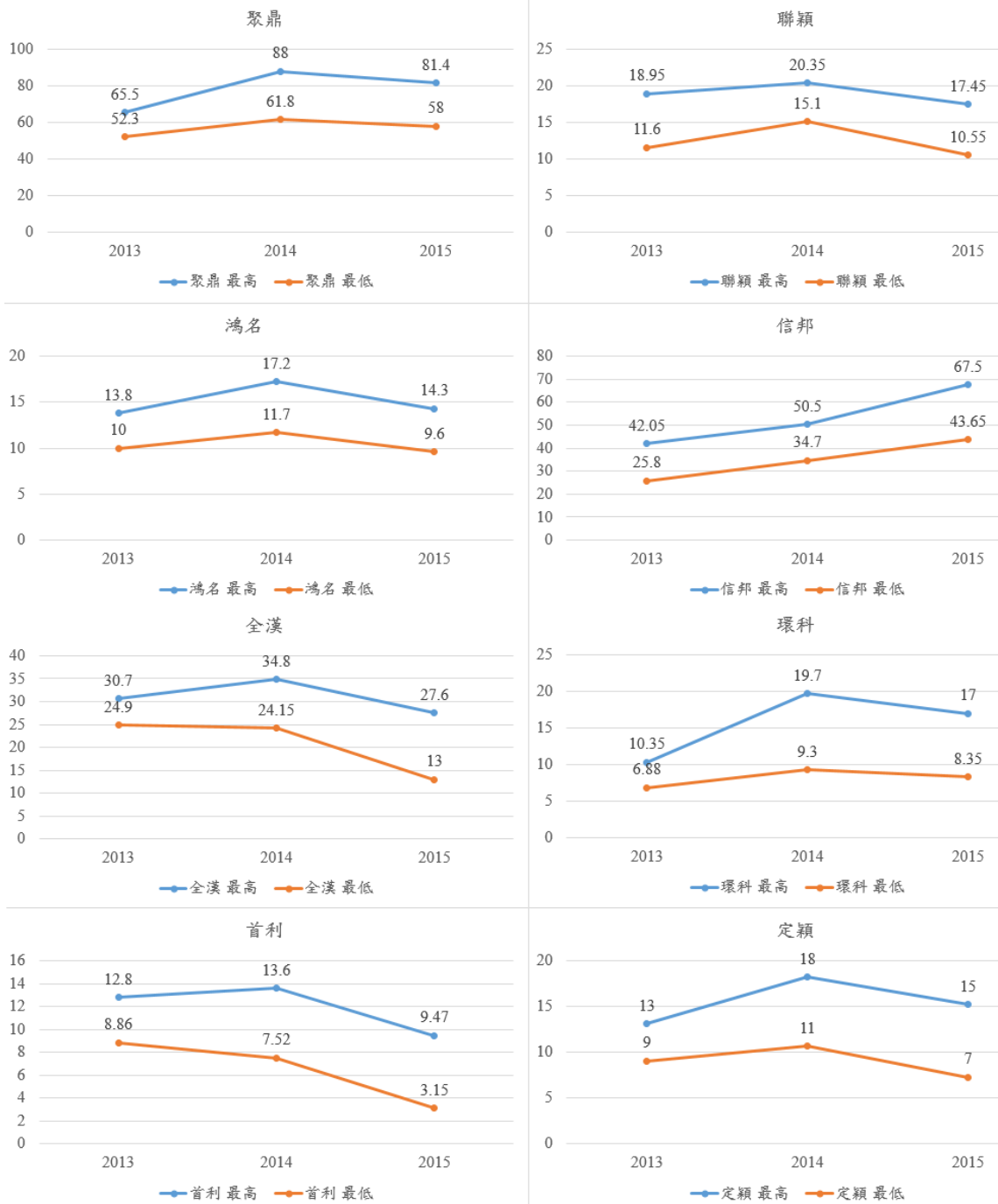


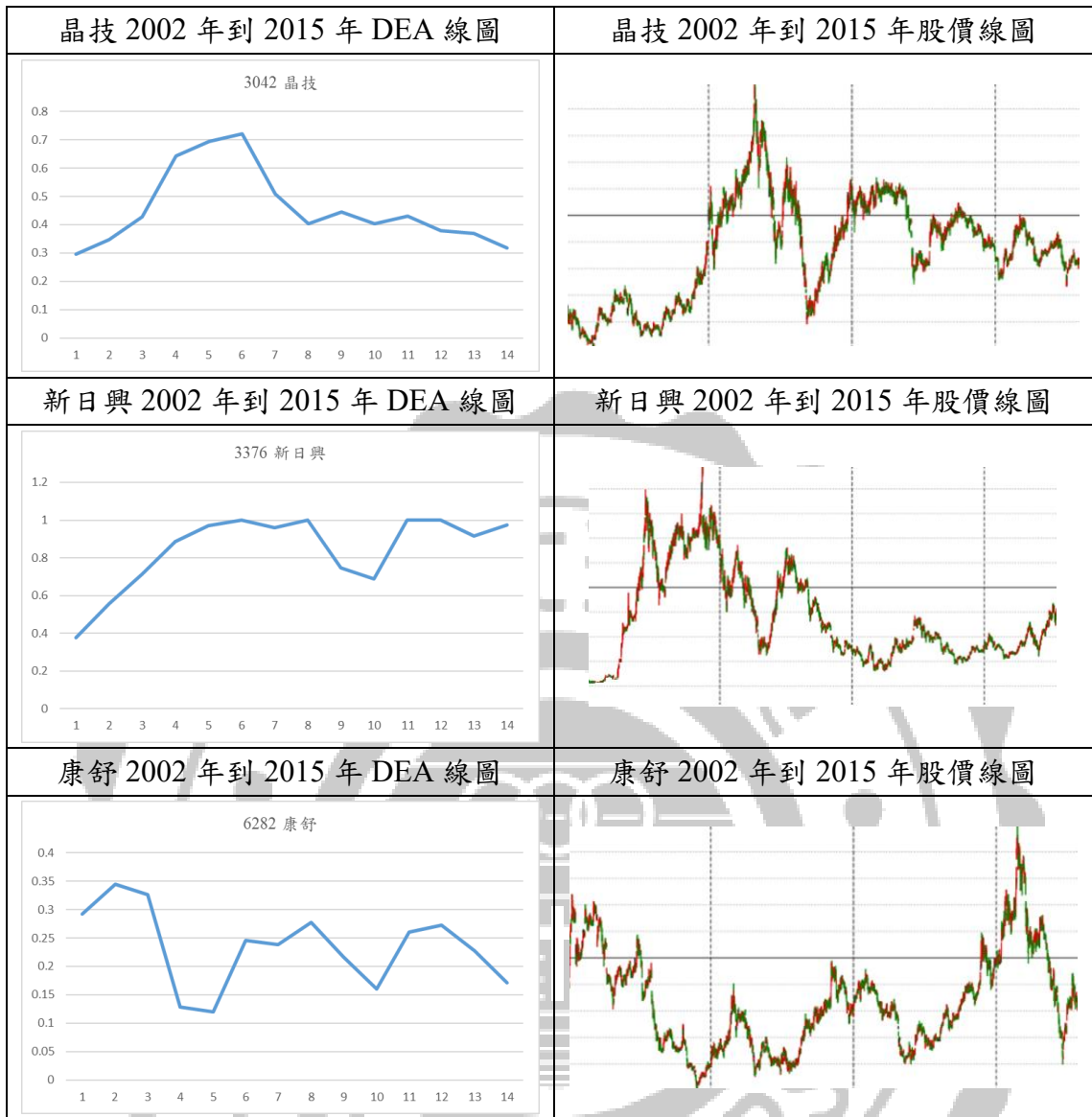
圖 4-3 各公司三年股市交易情形曲線圖

為了驗證 DEA 的分析結果所表示的可能性，如果以更小時間單位進行研究，就可能就可以預測得更加準確，所以本研究另外分析了 15 家電子零組件公司從 2002 年到 2015 年的 DEA 情況，並以該公司 2002 年到 2015 年的股票線圖作為對照，進而進行相互比對分析，而在 13 年期間的 DEA 與股價的分析情況結果如下表 4-1 所示。

本研究分析了 15 家公司 DEA 數值並做成趨勢圖來和股價進行相互對照，本研究選出了其中 5 家公司做成表格，在這表格中可以看出這 5 家公司，耀華、揚博、晶技、新日興、康舒的 DEA 與股價相互對照的情形，可以看出在 DEA 上漲的時候股價也就跟著上漲，DEA 下跌時也就跟著下跌，但是可以看出漲跌的幅度不是依照 DEA 變化的幅度來做變化，所以可以知道，DEA 可以當作股票漲跌的參考，來判斷股票的波動方向，但是卻不能判斷出股票漲跌的幅度，但這個分析結果代表，本研究利用 DEA 作為分析的數值是具由參考性的。

表 4-1 2002 年到 2015 年 DEA 與股價分析狀況

耀華 2002 年到 2015 年 DEA 線圖	耀華 2002 年到 2015 年股價線圖
揚博 2002 年到 2015 年 DEA 線圖	揚博 2002 年到 2015 年股價線圖



下表 4-2 是本研究分析出 2013 至 2015 年的分析數值，用於進行決策樹、隨機森林的分析依據，用來判斷各變數對於公司獲利能力的影響性。

表 4-2 91 家電子零組件公司資料包絡分析法(DEA)之結果

公司名稱	2013DEA	2014DEA	2015DEA	公司名稱	2013DEA	2014DEA	2015DEA
新日興	1	1	1	瀚宇博	0.17426	0.18639	0.17258
新利虹	1	1	1	敬鵬	0.17395	0.18623	0.16876
聚鼎	0.95799	0.99307	0.92614	旺詮	0.17051	0.17655	0.16683
聯穎	0.87934	0.89116	0.86052	康舒	0.15557	0.17242	0.1561
達邁	0.86587	0.86039	0.84576	泰碩	0.15421	0.16881	0.1546
志聖	0.71463	0.69735	0.68326	群電	0.15085	0.16863	0.15425
鑫禾	0.62804	0.69241	0.67009	華通	0.14565	0.15636	0.15364
鈞寶	0.61716	0.65717	0.58891	新至陞	0.14465	0.15232	0.15203
川湖	0.59279	0.59822	0.55365	立隆電	0.14425	0.15182	0.15019
揚博	0.52143	0.51729	0.54284	鎰勝	0.14077	0.1482	0.14969
希華	0.4592	0.4921	0.4994	智寶	0.13944	0.14291	0.14597
健策	0.43494	0.48821	0.48379	金橋	0.13844	0.14063	0.14309
銘鈺	0.4112	0.48491	0.45696	禾伸堂	0.13673	0.13458	0.14107
台郡	0.39893	0.47305	0.45061	嘉聯益	0.13092	0.1333	0.13862
新巨	0.39035	0.44954	0.40265	全漢	0.12942	0.1308	0.1377
國巨	0.38617	0.44916	0.39989	環科	0.12881	0.12927	0.13466
健和興	0.36587	0.40434	0.39465	建通	0.12415	0.11656	0.13063
奇力新	0.35365	0.37629	0.39419	群光	0.118	0.11195	0.13025
同泰	0.33991	0.37132	0.39108	耀華	0.11627	0.11108	0.11822
大毅	0.33032	0.34714	0.37023	建準	0.1126	0.10945	0.1157
興勤	0.3285	0.33952	0.35997	聯茂	0.11021	0.10875	0.1053

信錦	0.30815	0.33749	0.3347	楠梓電	0.1039	0.1087	0.10522
晶技	0.29656	0.3325	0.33334	金像電	0.09992	0.09571	0.10365
勝德	0.29472	0.32738	0.31969	鉅祥	0.09254	0.08268	0.10266
嘉澤	0.29166	0.31469	0.30993	偉訓	0.07723	0.08027	0.08185
百容	0.2858	0.31179	0.30178	廣宇	0.07156	0.07572	0.07407
志超	0.28056	0.29862	0.28556	捷泰	0.07126	0.07361	0.07254
鴻名	0.2737	0.29728	0.28132	健鼎	0.06955	0.07147	0.0665
信邦	0.26301	0.28907	0.2792	台端	0.06881	0.06478	0.05989
日電貿	0.25978	0.28781	0.27649	聯昌	0.06842	0.05662	0.05984
瀚荃	0.25805	0.27752	0.24632	正崙	0.06083	0.05034	0.05515
詮欣	0.258	0.27179	0.24393	立德	0.06015	0.04633	0.04945
欣興	0.25551	0.25846	0.24186	華新科	0.05516	0.03639	0.04789
晟鈦	0.25501	0.25534	0.23606	錫新	0.05182	0.03518	0.04692
鈺邦	0.25	0.24718	0.23525	精成科	0.04031	0.0341	0.03862
聯德	0.24841	0.24699	0.2333	飛宏	0.03992	0.03405	0.03585
柏承	0.23105	0.24574	0.21953	菱光	0.03808	0.02895	0.03262
宏致	0.21117	0.24391	0.21211	今皓	0.03364	0.02838	0.02577
台達電	0.21014	0.24259	0.21055	毅嘉	0.02386	0.02719	0.02479
競國	0.20356	0.22876	0.20884	太空梭	0.01452	0.02548	0.02121
台虹	0.20197	0.22546	0.20322	首利	0.01202	0.01426	0.01466
維熹	0.196	0.21899	0.19523	定穎	0.00668	0.01221	0.00865
谷崧	0.19102	0.21332	0.19176	力銘	0	0.00106	0
台光電	0.18959	0.21014	0.18705	南電	0	0	0
良得電	0.184	0.20295	0.18578	銘旺科	0	0	0
佳必琪	0.18334	0.18696	0.18111				

第二節 決策樹分析結果與說明

壹、決策樹分析驗證

本研究再進行決策樹分析前將決策樹的資料分為訓練樣本和測試樣本，訓練樣本佔 70%，測試樣本佔了 30%，本研究共分別做了 3 年的決策樹分析所以共有 3 次，以此做為驗證決策樹分析的準確性，以下是本研究進行決策樹分析驗證的結果：

表 4-3 決策樹分析驗證

年份	2013	2014	2015
驗證結果	100%	100%	100%

貳、決策樹分析結果

在決策樹的分析結果中可以推斷出每股盈餘、營業成本、營業利益是主要常見影響公司財務狀況的數值，而以下是這三種數值的解釋。

每股盈餘：

每股盈餘與各家公司的股價具有一定的連動性，所以也大多是衡量投資人與公司獲利的主要關鍵要素之一，每股盈餘有可以被稱為每股收益或者是每股營利。

每股盈餘的公式如下：

$$\text{每股盈餘} = \text{稅後淨利} / \text{流通在外的普通股加權股數}$$

營業成本：

營業成本是指一家公司在一定期間內，所發生在營業上的費用，譬如餐廳服務員的薪水，工作人員加班或伙食上的人事費用等，或者是餐廳進行烹飪所必須購買的食材費用，水電使用後必須繳納的費用等只要進行營業就會產生的費用這些被稱之為營業成本。

營業成本常常會與營業費用搞混，所以在此介紹營業成本與營業費用的差別。營業成本是大多是由，人力、購料、品質、水電和公司機器或用品的折舊等組成，重點是營業成本是產業在製造流程上產生的費用。

營業費用大多是在銷售折讓、處理品質、庫存損失、產品運輸、讓業務可以順利運作的費用等，重點是在管銷與銷售方面的費用。

營業利益：

營業利益是公司在營業上的收益扣除所有的營業成本後，所可以收到的利潤。

營業利益的公式如下：

$$\text{營業利益} = \text{營收} - \text{銷貨成本} - \text{營業費用}$$

藉由這三種數值的解釋可以了解，每股盈餘的用途本身就是用來判斷出公司獲利能力的關鍵數值，所以藉由分析可以分析出每股盈餘，代表本研究以決策樹分析的結果符合研究目的，達到找尋能判斷公司獲利能力的關鍵指標，而營業成本及營業利益比部分，營業成本是屬於公司在經營與運傳所花費的代價，營業利益是指公司在花費代價後所可獲得的收益在減除成本後的實際獲利，所以本研究認為營業成本和營業利益也都主要跟公司運轉的結果有較大的關係，公司運轉後所獲得的報酬非常直觀的顯現在財務報表當中，在以下表格中是決策樹分析後所獲得的主要變數，可藉由重複性來卻得較為有影響力的數值，也是決策樹主要分析的結果。

表 4-4 決策樹分析結果統整

決策樹分析結果統整		
2013	2014	2015
每股盈餘	營業成本	每股盈餘
營業成本	營業利益	營業成本
其他流動資產	短期投資	營業利益
營業利益	來自營運之現金流量	固定資產
現金與約當現金		營業毛利
支付現金股利		

藉由表格統整可以明確看出，每股盈餘、營業成本、營業利益是主要有重複性的變數，再藉由下面對於決策樹結果的樹狀的來看，符合條件中，每股盈餘較大，營業成本較小，營業利益較大的公司，財務狀況越好，為值得投資。

以下是主要分析資料中藉由決策樹分析後經營狀況較為良好的幾家公司，藉由數值分析的結果。

表 4-5 三年各 DEA 最大值

2013 年 DEA 平均值	2013 年 DEA 最大值	2013 年 DEA 最小值	2013 年 DEA 平均值
0.243482558	1	0.258000089	0.627427067
2014 年 DEA 平均值	2014 年 DEA 最大值	2014 年 DEA 最小值	2014 年 DEA 平均值
0.255778674	0.99307411	0.347138547	0.602263215
2015 年 DEA 平均值	2015 年 DEA 最大值	2015 年 DEA 最小值	2015 年 DEA 平均值
0.248089877	0.926140087	0.208836813	0.52619046

上表中的數值是 2013 到 2015 年 3 年間決策樹分析得結果中，分析出獲利能力最好的幾家公司，2013 年為 13 家，2014 年為 13 家，2015 年為 11 家，而這些公司 DEA 的數值也就是獲利能力的平均狀況，可以看出決策樹分析後獲得的共 37 家公司以 DEA 得平均值來看較全部 91 家公司來的高，所以分析結果的公司財務狀況的確較好，而 2013 年和 2015 年的 DEA 最小值都高於平均，但 2015 年卻低於平均值，決策樹分析出的公司都是獲利能力較好的，所以出現了 2015 年最小值較低的情況，代表 2015 年台灣電子業中，電子零組件的財務狀況較差有相較前兩年來說較糟糕的狀況，所以可以知道決策樹可以知道獲利能力較好的幾家公司和比較有參考價值的數值之外，可以了解各年整體產業的財務狀況是否良好。

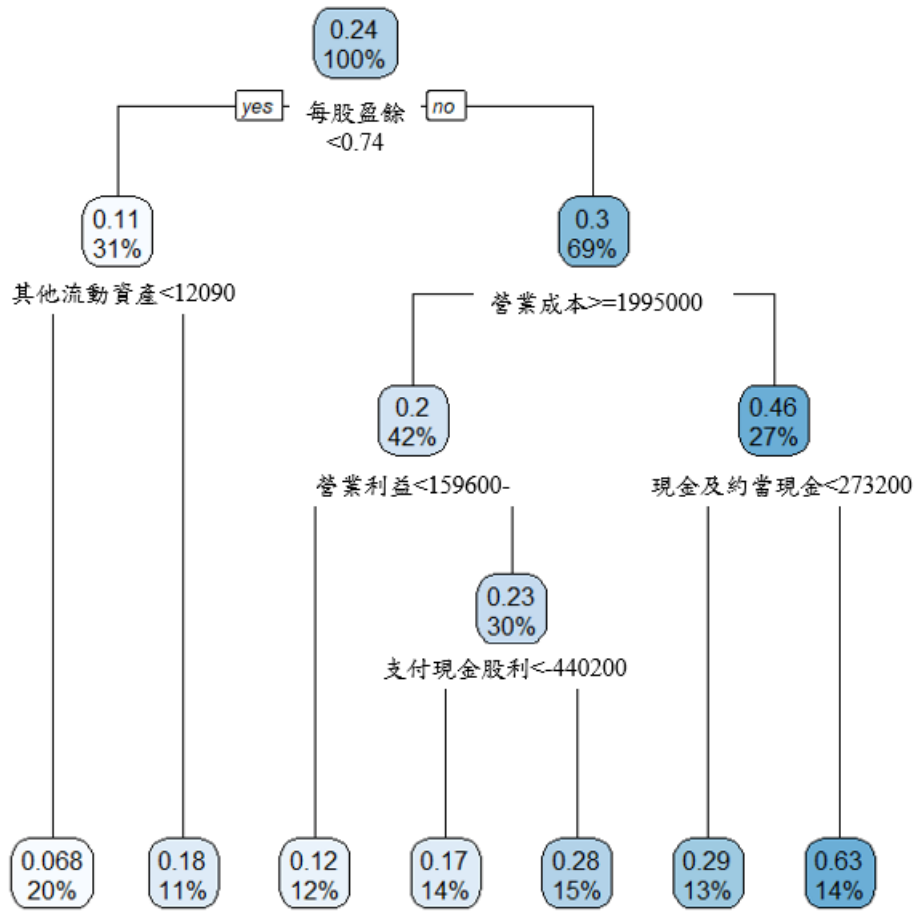


圖 4-4 2013 年決策樹結果

從 2013 年決策樹圖可以看到，決策樹分析結果中，以每股盈餘、營業成本、營業利益、現金與約當現金、其他流動資產、支付現金股利，為主要分析結點，依照 2013 年的決策樹分析結果，資料的區間於每股盈餘 >0.74 、營業成本 <1992000 、現金及約當現金的 >273200 的數值是最佳的，總共有 13 家上市公司分別為新利虹、新日興、聚鼎、達邁、志聖、鑫禾、鈞寶、川湖、希華、銘鈺、健和興、百容、詮欣，並從這幾家公司的 DEA 部分可以知道都大於該年全部 91 家公司的平均值，而此決策樹中，除了在這 3 年決策樹分析結果每股盈餘、營業成本、營業利益外，還多了較不重複的現金及月約當現金，其中約當現金是指公司短期投資中較為容易變現的部分，所以可知一家公司的現金存量也是屬於判斷公司營運狀況的指標之一。

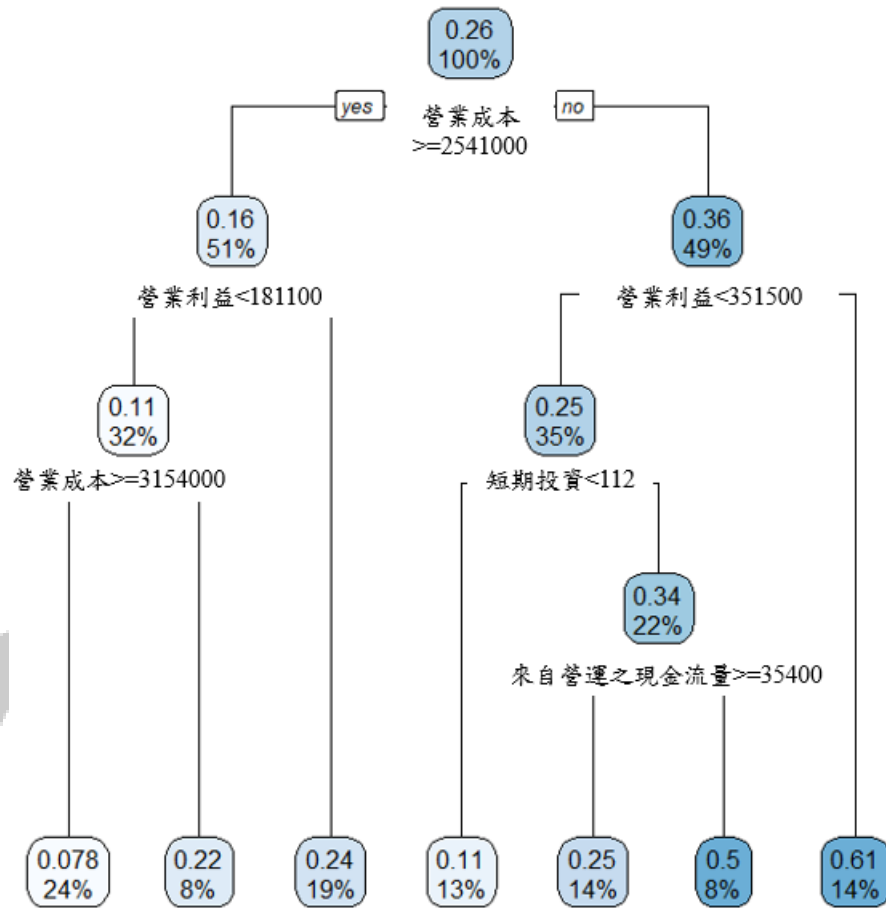


圖 4-5 2014 年決策樹結果

從 2014 年決策樹圖可以看到，決策樹分析結果中，以營業成本、營業利益、短期投資、來自營運之現金流量，為主要分析結點，依照 2014 年決策樹分析結果，主要分析 91 家公司，財務資料中資料區間營業成本 < 251400、營業利益 > 351500 的公司財務狀況較好，共有 13 家公司分別為，達邁、新日興、聚鼎、揚博、鑫禾、志聖、川湖、希華、新巨、健策、銘鈺、健和興、興勤，依照財務資料可以知道這 13 家公司的 DEA 較 2014 年的平均值大，代表分析結果是準確的，而此次分析中沒有每股盈餘的部分，但是本研究計算該分析結果的每股盈餘為 5.3，而 2014 年的每股盈餘為 2.707582418，所以可以知道每股盈餘較大財務情況越好，符合決策樹整體分析的結果。

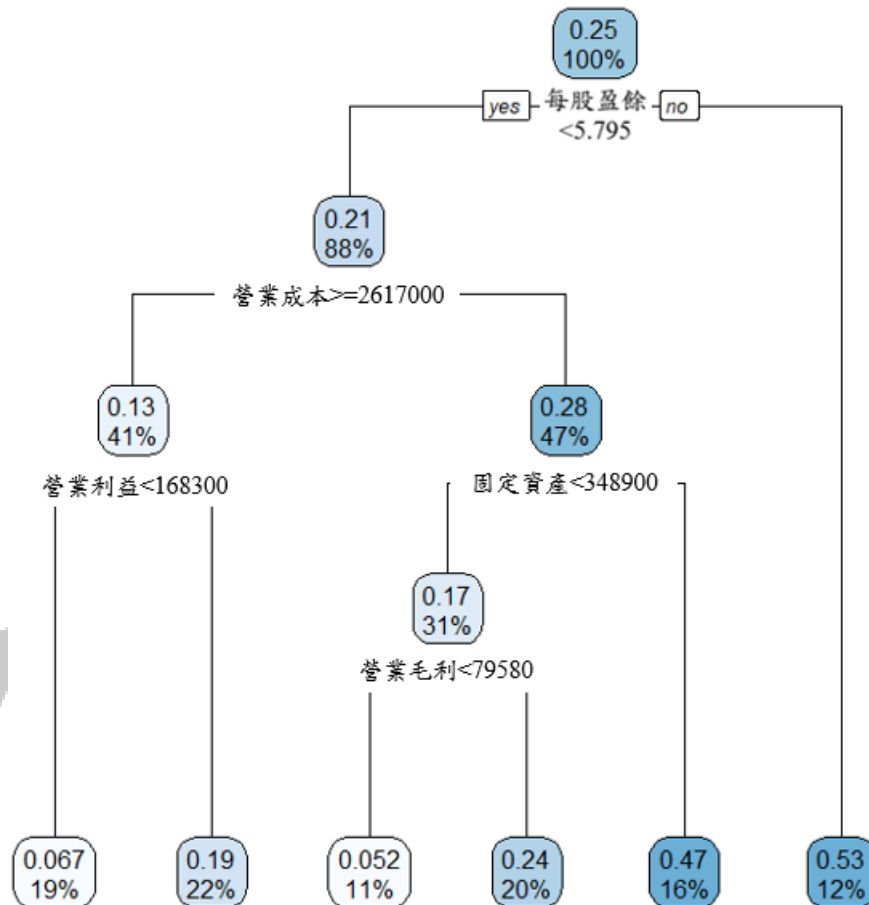


圖 4-6 2015 年決策樹結果

從 2015 年決策樹圖可以看到，決策樹分析結果中，以每股盈餘、營業成本、營業利益、固定資產、營業毛利，為主要分析結點，依照 2015 年決策樹分析結果，主要分析 91 家公司，財務資料中資料區間每股盈餘 > 5.795 的公司財務狀況較好，共有 11 家公司分別為，新日興、鑫禾、聚鼎、楠梓電、川湖、台郡、銘鈺、台光電、嘉澤、維熹、台達電，依照財務資料可以知道這 11 家公司的 DEA 較 2015 年的平均值大，代表分析結果是準確的，而其中較左邊的財務數值營業成本和營業利益依照分析出的 11 家公司的財務數值，2015 年營業成本和營業利益的平均非別為，6,239,979 和 299,193，而這 11 家公司的平均為 6,565,876 和 941,715，依照這些數值可以知道在營業利益的部分確實是越大越好，而營業成本方面雖然這 11 家公司沒有數值比較小，但是在決策樹中出現小於 2317000 的數值財務狀況較好，所以可以推斷出，每股盈餘、營業成本、營業利益的分析狀況符合預期。

第三節 依照隨機森林結果變數分析之決策樹結果與說明

下圖為 2013 年至 2015 年所作之決策樹分析的結果圖，本研究以隨機森林之分析結果進行決策樹分析和以 39 財務指標分別進行決策樹分析的結果進行相互對照，並進行分析，找出之間的差異性。

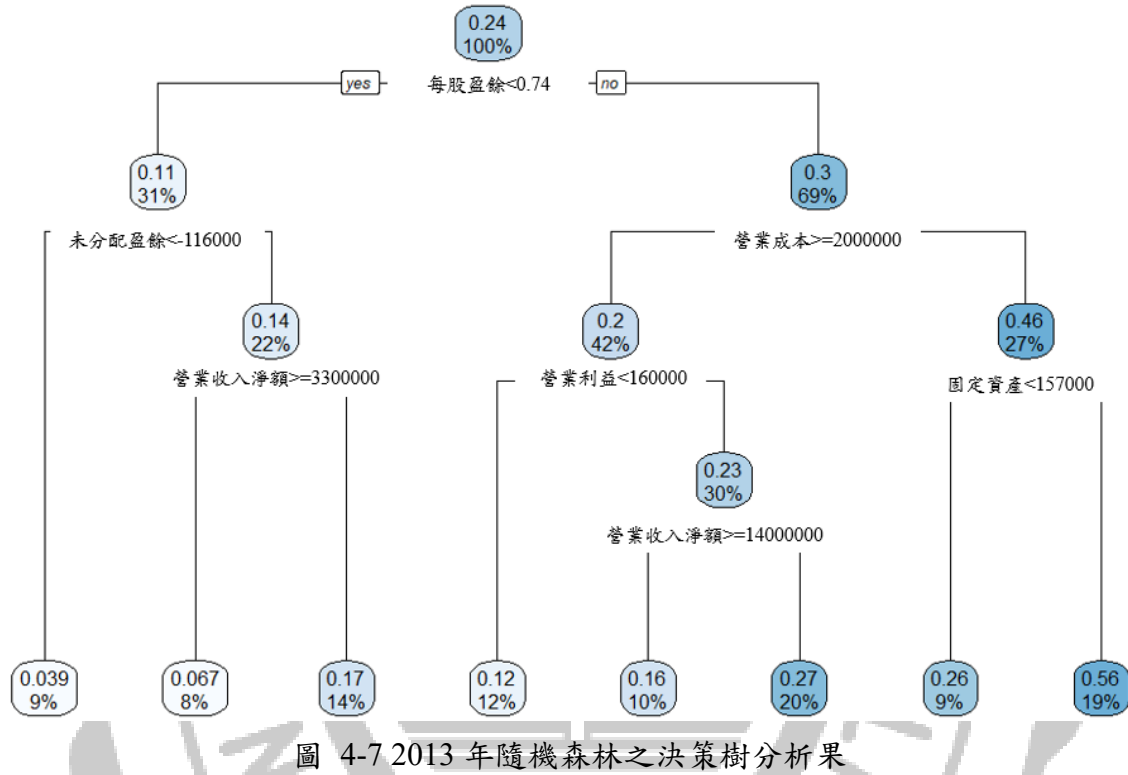


圖 4-7 2013 年隨機森林之決策樹分析果

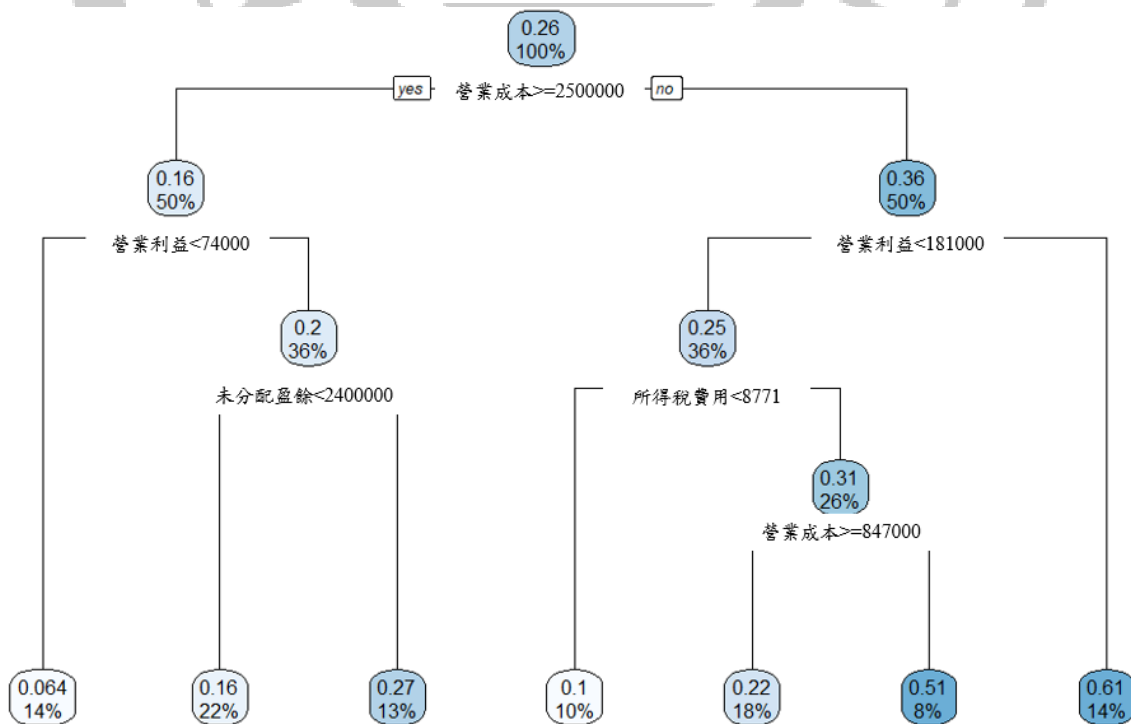


圖 4-8 2014 年隨機森林之決策樹分析果

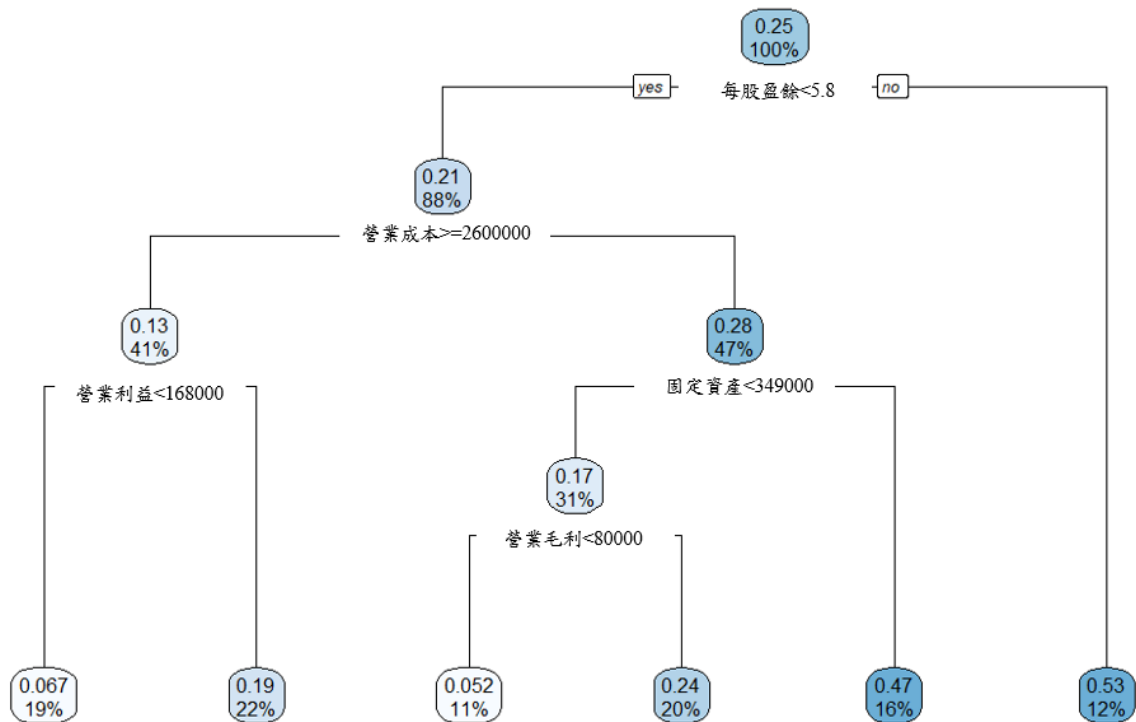


圖 4-9 2015 年隨機森林之決策樹分析果

以上是研究以隨機森林分析之結果進行決策樹分析所得之結果，其中可以看到每股盈餘、營業成本、營業利益與 39 種決策樹分析的結果大致相同，本研究整理了這 3 張決策樹分析中重複變數的表格。

表 4-6 隨機森林之決策樹分析結果統整

決策樹分析結果統整		
2013	2014	2015
每股盈餘	營業成本	每股盈餘
未分配盈餘	營業利益	營業成本
營業成本	未分配盈餘	營業利益
營業利益	所得稅費用	固定資產
營業淨收入		營業毛利
固定資產		

從以上表格中可以看出，這三年有重複性的財務指標為，每股盈餘、未分配盈餘、營業成本、營業利益，結果以 39 種變數分析的結果大致相同，其中以多了未分配盈餘，而未分配盈餘是保留盈餘中的一個項目，而保留盈餘就是尚未分配給股東的盈餘及不可分配給股東的調整數，所以並不會直接影響股票價格及其他的投資行為，所以已影響投資行為的財務指標中未分配盈餘是不需要考慮的項目，而未分配盈餘是會結算到下一個年度的盈餘中，所以以公司的經營能力來說，也是不需要去考慮的項目。

第四節 隨機森林分析結果與說明

從決策樹的分析中可以看出各財務報表中那些數值會影響公司的獲利能力，而使用隨機森林可以用來與決策樹的分析結果進行比較，了解對於影響公司的獲利能力的指標是否跟決策樹相同，來驗證決策樹的分析結果，在隨機森林分析法上，隨機森林是運用了許多的決策樹相關的分類器，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(Increase in MSE, IncMSE)值，IncMSE 越大代表影響到目標的關係越明顯，使用隨機森林分析出來的指標分別前幾項為營業成本、營業利益、存貨、固定資產、營業收入淨額、長期投資、每股盈餘，而營業成本、營業利益、每股盈餘，其中營業成本、營業利益、每股盈餘在決策樹的分析結果上是相同有出現過的，所以也證明了決策樹分析結果的準確性，下表中可以看出隨機森林所分析出的每一個變數對於本研究目標變數的影響程度，其 IncMSE 數值越高的則是影響性越大。

表 4-7 隨機森林分析結果

代號	變數	IncMSE	代號	變數	IncMSE
A2	營業成本	22.12	A36	折舊	7.81
A5	營業利益	19.51	A21	流動負債	7.44
A15	存貨	12.47	A14	應收帳款及票據	7.3
A18	固定資產	12.19	A32	來自營運之現金流量	7.05
A1	營業收入淨額	12.06	A28	未分配盈餘	7.02
A17	長期投資	11.73	A25	資本公積	6.66
A10	每股盈餘(元)	10.47	A24	負債總額	6.56
A3	營業毛利	9.59	A8	所得稅費用	6.12

第五節 相關結果與說明

壹、相關分析信度結果

本章節是相關分析信度的結果而相關分析信度的依據就是 Cronbach Alpha 的大小，Cronbach Alpha 值若小於 0.35 為低信度，介於 0.35 至 0.70 為尚可，大於 0.70 為高信度。

表 4-8 主要分析資料 Cronbach Alpha 值

2013 年 相關分析 Cronbach Alpha 值		2014 年 相關分析 Cronbach Alpha 值		2015 年 相關分析 Cronbach Alpha 值	
變數	Alpha	變數	Alpha	變數	Alpha
原始	0.88443	原始	0.881502	原始	0.876884
標準化	0.956355	標準化	0.95534	標準化	0.952471

本研究表示相關分析的信度方面的目的，是為了可以確定本研究的研究成果是可靠的，而本研究在信度上的分析表明，依照上面的表格資料，可以知道本研究在相關分析中，信度都在 0.70 以上，代表本研究的信度屬於較高，也就是本研究的相關分析是可信的，這可以知道本研究以相關分析作為研究成果的一部分是可行的，所以本研究將接下來將進行相關分析的解說，並依照結果跟其他分析方法，如決策樹等進行比對，希望可以找出其他的研究成果。

貳、相關分析分析結果與講解

下面的表格是本研究相關分析的結果，在相關分析中數值本研究以小於 0.005 數值作為相關性較大的，而所有變數中，大多數的數值相互的相關性都是小於 0.005 的，而除了每股盈餘的部分比較特別，因為除了每股盈餘大部分變數互相的相關性都很大，除了每股盈餘例外，所以本研究將相關分析較為有差異的部分作為表格，而從下表可以看出，所得稅費用和營業利益的相關性都小於 0.005，這代表這兩個變數對於每股盈餘有較大的相關性，而從每股盈餘的數學公式中看出每股盈餘是用稅後淨利除以普通股加權股數所獲得的所以所得稅費用與每股盈餘有相關性是正常的，而營業利益是有稅後淨利進行換算得來的所以也跟每股盈餘有相關性，但是除了這 2 個數值每股盈餘跟其他的變數相關性都很小，所以可以知道每股盈餘是相對獨立的數值。

表 4-9 2013 年相關分析分析結果

2013 年相關分析				2014 年相關分析				2015 年相關分析			
	所得稅 費用	營業利 益	每股盈 餘(元)		所得稅 費用	營業利 益	每股盈 餘(元)		所得稅 費用	營業利 益	每股盈 餘(元)
所得稅 費用	1	0.59607	0.48951	所得稅 費用	1	0.61799	0.41601	所得稅 費用	1	0.64844	0.57058
		<.0001	<.0001			<.0001	<.0001			<.0001	<.0001
營業利 益	0.4961	1	0.4267	營業利 益	0.3414	1	0.375	營業利 益	0.3293	1	0.4261
	0.0001		<.0001		0.001		0.0003		0.0014		<.0001
每股盈 餘(元)	0.1283	0.4267	1	每股盈 餘(元)	0.1042	0.375	1	每股盈 餘(元)	0.0952	0.4261	1
	0.2257	0.0001			0.3283	0.0003			0.3693	<.0001	

第六節 分群結果與說明

壹、分群分析的結果

在分群分析的結果中可以看出，不論是台達電或者是正崙、欣興、群光、華通、南電、國巨等都不是以決策樹分析結果為依據來進行分群，但從財務資料中可以看到分群結果是照資產總額進行分群，但這個關鍵數值卻不會影響各公司的經營能力，進一步分析中可以知道在 3 年的財務資料裡其他變數中，資產總額排名較為前面的 7 家公司在固定資產、負債及股東權益總額的額度上也是較大的，而在購置的固定資產的負債也是較大的，推斷出公司的固定資產越大，固定資產折舊就越多，另外營業收入淨額、應收帳款及票據、存貨、長期投資、流動負債、負債總額、股東權益總額也會因為公司的規模越大，額度會越大，但是只是代表公司的資產總額越大，公司進行流動的資金越大，並且從 3 年間的 DEA 來分析，資產總額較大的公司，獲利能力並不會比較好，所以公司的獲利能力跟公司的資產大小無關，並且跟獲利能力有關的應該很明確是公司的獲利中，盈餘的部分。

1. 營業成本、營業利益、每股盈餘，都不是分群的依據。
2. 分群結果經由分析是依照資產總額進行分群。
3. 資產總額較大的公司，獲利能力並不會比較好。
4. 與獲利能力有關的是公司的獲利中，盈餘的部分。

貳、分群分析的結果概況

2013 年的分群分析中可以看到台達電為唯一的第 3 群，而正崙、欣興、南電、群光、華通、國巨為第 1 群，2014 年的分群分析中可以看到台達電為唯一的第 1 群，而正崙、欣興、南電、群光、華通、國巨、健鼎為第 2 群，2015 年的分群分析中可以看到台達電為唯一的第 2 群，而正崙、欣興、群光、華通、南電、國巨、健鼎為第 1 群，以分群概況來說台達電連續 3 年都分為一群，而正崙、欣興、群光、華通、南電、國巨被分為一群，而健鼎在 2014、2015 年有被與正崙、欣興、群光、華通、南電、國巨分在一群，但是 2015 年就沒有了。

參、分群分析的結果圖

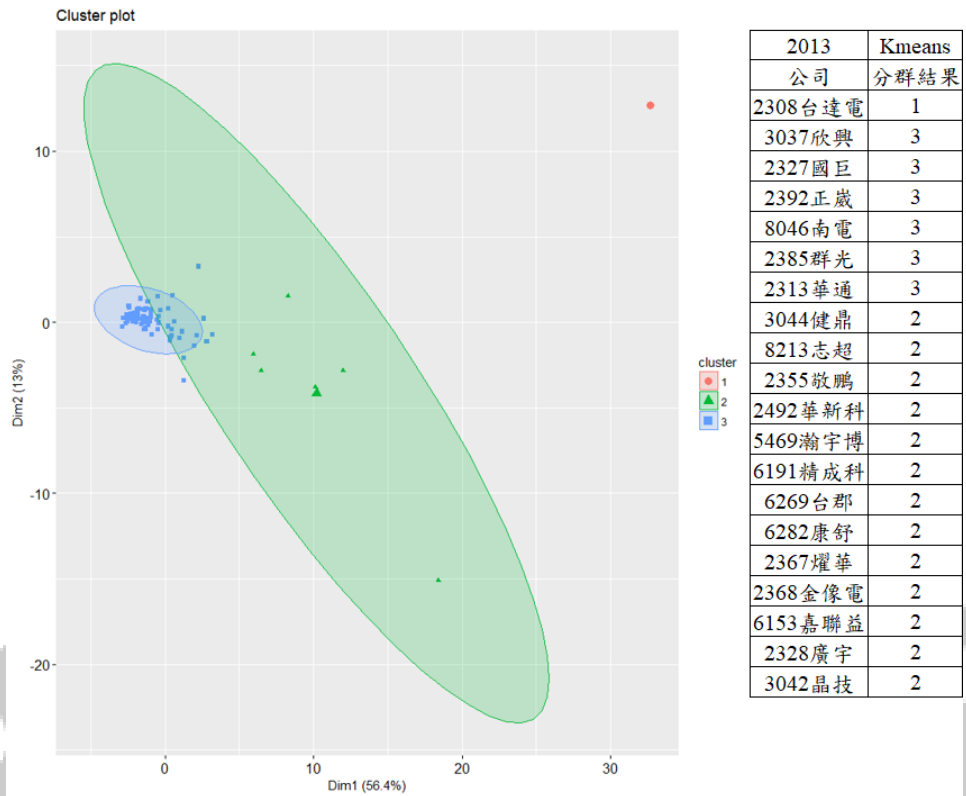


圖 4-10 2013 年分群結果

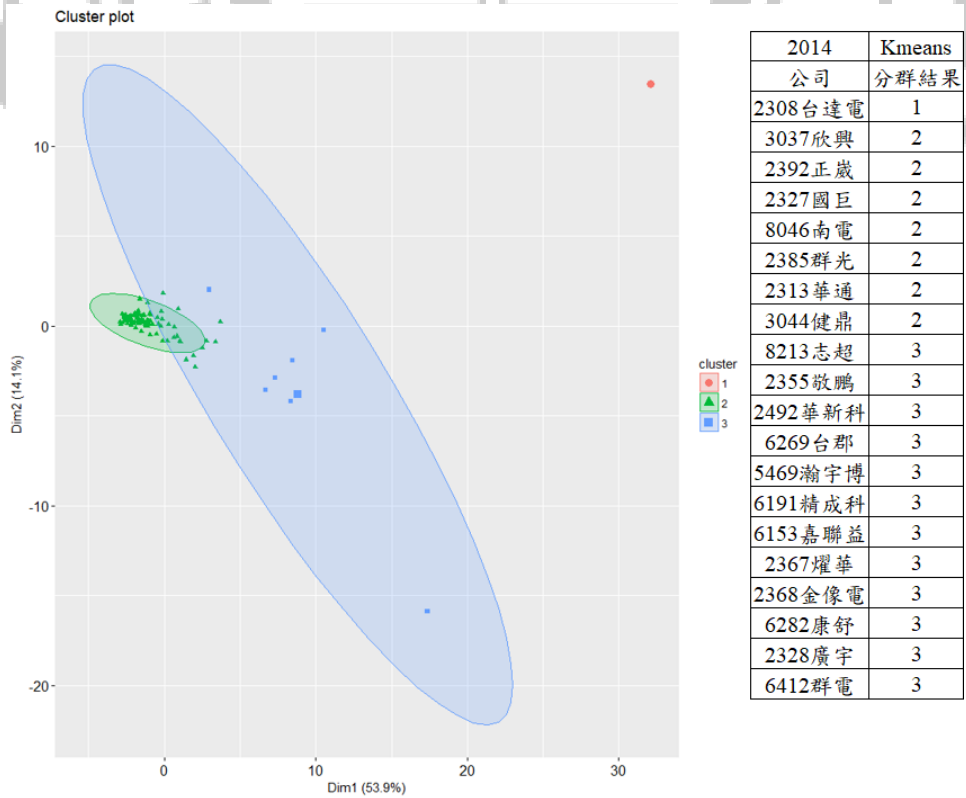


圖 4-11 2014 年分群結果

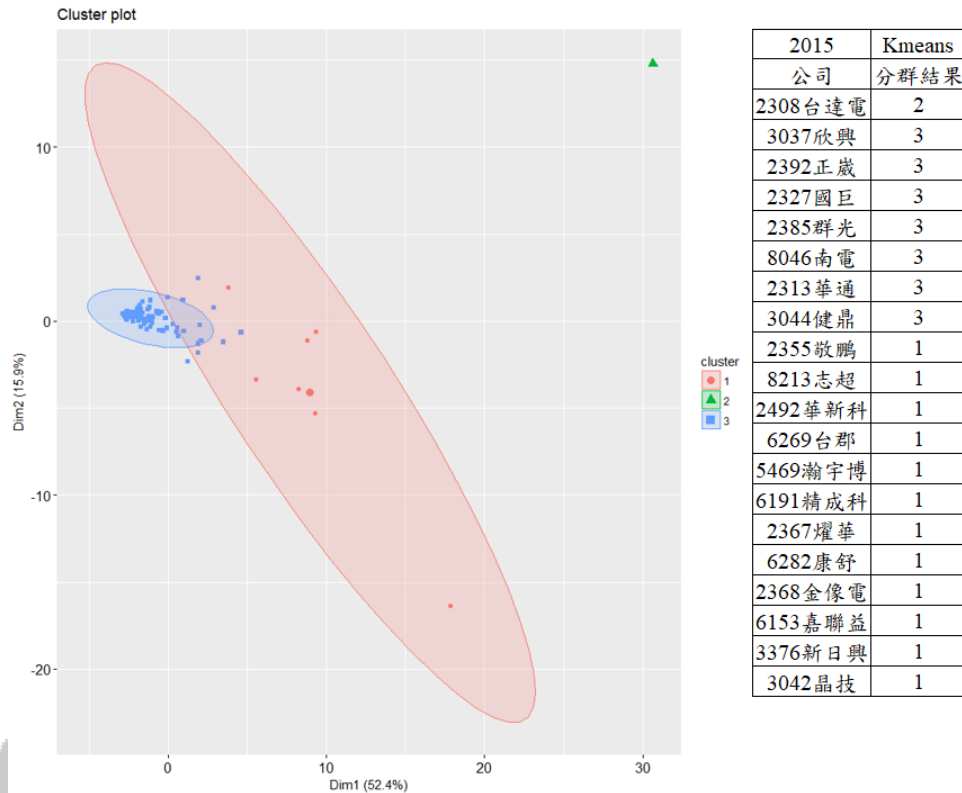


圖 4-12 2015 年分群結果

以上 3 張分群結果的分布圖，主要是用來查看各公司間依照所分析的財務數值中的某些變數進行分群的結果，這張圖的右方本研究放入的結果的表格方便可以知道分群的結果概況，其中因台達電都是被分為唯一一群，所以本研究將台達電放入第一順位，而正崙、欣興、群光、華通、南電、國巨等，通常被分為一類，所本研究將其放入第二順位，而起於的公司都在一類，所以本研究就沒將其全部表示出來，但是也足夠可以藉由表格了解分布狀況，而本研究分群的結果在前章節已經進行解說，本研究分群依照財務報表進行比對，是依照資產總額進行分群，代表各公司最明顯的差別就是經營規模的部分，而此結果代表分群不是主要可以知道各公司間財務狀況差別的依據，並不能推導出獲利能力的結果。

肆、標準化後分群分析的結果概況

以下是本研究將資料進行標準化分析後，再進行決策數分析的結果，會跟未做標準化的分群分析最為比較，並比較出相互的差異，以下就是本研究標準化的結果圖。

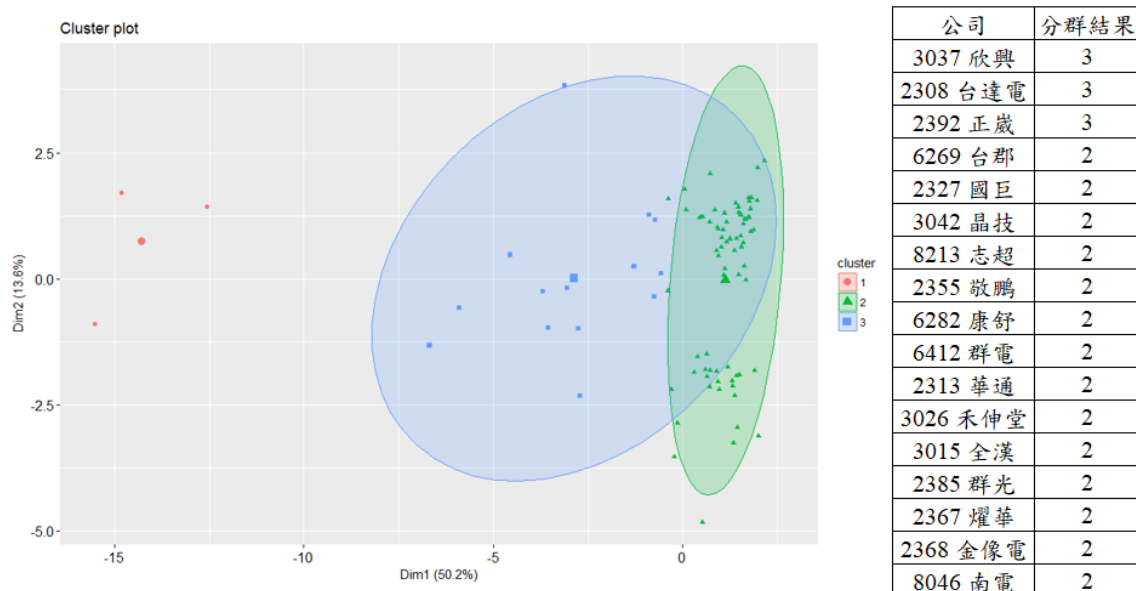


圖 4-13 2013 資料標準化後分群結果

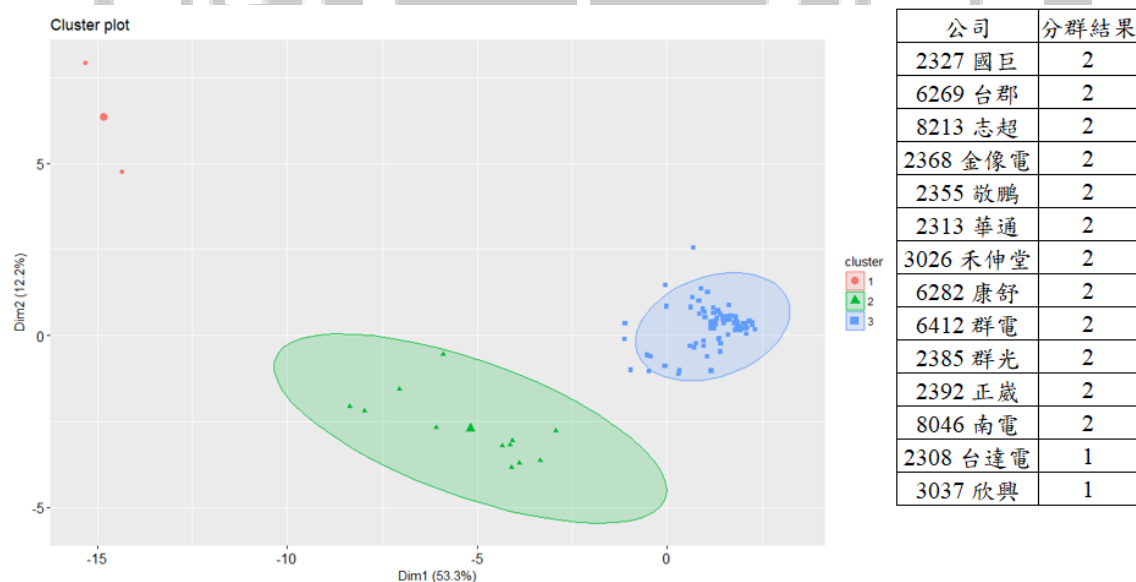


圖 4-14 2014 資料標準化後分群結果

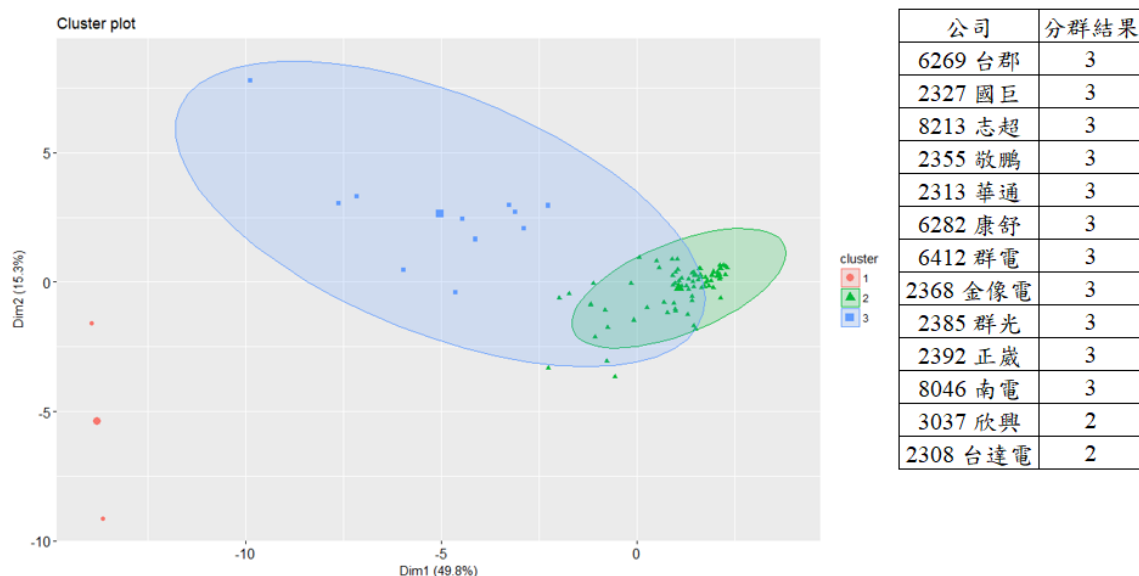


圖 4-15 2015 資料標準化後分群結果

從上圖標準化的的分群結果大致上是以欣興和台達電為一群，南電、群光、正崙、華通等為一群，只有 2014 年正崙和欣興、台達電分為一群，本研究比照了分析結果中 91 家公司，共 39 種財務指標進行分析，39 種財務指標中並沒有明顯是哪一種變數是主要的分析依據，但可以知道所有數字大小較大的大多分在一群，其中以欣興、台達電為主要分析目標中，各財務指標數字較大的，其他 2.3 群也是有此狀況，所以我們可以知道做過標準化後，分群結果跟未做過標準化主要的差異是標準化後無法在財務指標中找到很明確進行分群的依據，而標準化可以，但是經由總體數字的判斷，公司規模是分群分析的主要依據是不變的，所以公司規模不會對 DEA 有所影響的結果是不變的。

第五章 結論與研究限制

本研究利用了資料包絡分析法、決策樹、隨機森林等多種分析法來尋找資產負債表、現金流量表、損益表中影響 91 家電子零組件公司的財務指標，先是在資料包絡分析法(DEA)分析出目標變數，並用於接下來的分析，在決策樹分析法中，分析出營業成本、營業利益、每股盈餘這 3 種變數是影響目標變數最大的數值，而在隨機森林方面也驗證了這 3 種數值的影響程度是顯著的，在各種變數中影響目標變數是在排名靠前的，在相關分析中，可以知道大部分的變數相關性都很大除了每股盈餘與大多數的變數沒有相關性，其餘變數之間都具有一定的相關性，也因此可以知道在決策樹中所分析出得結果營業成本、營業利益、每股盈餘中每股盈餘是較為獨立的一項數值所以在這三種數值中最具有明顯的分析價值，在在分群分析中，可以了解各公司經營規模跟公司的獲利能力沒有太大的關聯性，從中可以看出在分群分析的結果與決策樹所分析出的 3 項變數之間沒有關聯性。

統整結論：

1. 影響公司獲利能力的財務指標為營業成本、營業利益、每股盈餘。
2. 公司的經營規模的對獲利能力的影響並不明顯。
3. 每股盈餘和營業利益在財務報表中越高獲利能力越好，營業成本則越低越好。
4. 每股盈餘的數值較為獨立不受到大部分財務數值的影響。

研究限制：

1. 雖然台灣電子業投資交易額較市場其他投資類別大，但是以上市電子業公司為研究對象，還是不能代表全台灣所有上市公司的營運狀況。
2. 每年上市公司的總數不一定相同，要以歷年所有財務資料作為研究對象進行全面性研究，較為困難，但是本研究以 13 年 15 家公司的資料進行驗證了解本研究的研究方向與目的是可行的。

可改善方向：

本研究以 3 年的數據做為研究資料可能造成研究成果較不明顯，所以本研究利用 13 年共 15 家公司的資料做另外的分析，驗證 DEA 的數值的確可以判斷股票的漲跌，所以如過利用較長年度區間的資料作為分析基本上可以得到更好的成果。



參考文獻

中文文獻

1. 王建智、陳銘樹、徐正容 (民 97)。應用資料探勘方法中的分群分析技術來探究高階健檢客戶之型態組成以進行客戶關係管理-以台北某醫學中心之高階健檢客戶為例。健康管理學刊，6，51-58。
2. 王淑燕、曹正鳳、陳銘芷 (民 105)。隨機森林在量化選股中的應用研究。運籌與管理，3，164-168。
3. 王艷輝、郭亞軍、王麗在 (民 93)。上市公司短期信貸資金使用效率的 DEA 分析。中國經濟評論，4，100-109。
4. 吳允皓 (民 103)。台灣財經變數相關結構分析：景氣、金融與痛苦指標構面。真理大學，統計與精算學系，19-32。
5. 吳孟紋、林怡茹 (民 105)。資產負債表外業務活動對銀行風險的研究。兩岸金融季刊，4，26-35。
6. 吳怡瑾、李睿傑、陳子立 (民 103)。以資料探勘技術建立宅配業之車輛維修及預警決策支援系統。管理與系統，21，115-117。
7. 吳植森、廖柏森、余忠祐 (民 99)。結合特徵選取與 DEA 方法建立基金選購決策樹模型。「數位教學暨資訊實務研討會」，86-93。
8. 吳翠治、齊德彰、李科翰 (民 100)。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。真理大學運動知識學報，12，87-97。
9. 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，6，2-28。
10. 李建億、蔡芳遠 (民 93)。應用資料探勘技術於網路專題學習活動之分析。南師學報，38，1-23。
11. 李綱 (民 101)。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。國立臺北科技大學，工程科技研究所，15-56。

12. 林苡甄、李怡慶、黃志璿、吳信宏 (民 105)。利用資料探勘於病人安全文化的復原力研究以中部某區域教學醫院為例。國立彰化師範大學，企業管理學系，37-46。
13. 邱一卉、米紅 (民 99)。基於隨機森林和轉導推理的特徵提取方法。廈門大學學報自然科學版，信息科學與技術學院，49，1-6。
14. 胡志堅、黎漢林 (民 90)。以資料包絡法與投資報酬法評量產業績效—以台灣 IC 設計業為例。國立交通大學，管理學院資訊管理學程。
15. 徐志明、彭育秀 (民 102)。應用資料包絡分析法、蜂群演算法與類神經網路以建構股票投資組合最佳策略。明新學報，40，263-271。
16. 翁政雄、洪令莊、呂培豪、陳學瀚、郭家佑、施博惟、謝孟哲 (民 102)。應用決策樹於心臟病預測之研究。資訊管理學報，19，2-12。
17. 張仲岳 (民 90)。損益表之分類與獲利能力之預測。當代會計，2，4-15。
18. 張俊郎、陳志豪、蔡玉純 (民 96)。應用決策樹與類神經網路技術以提升皮膚病變鑑別診斷之品質。中華民國品質學會，13，3-8。
19. 張麗娟、許佳豪、張耀元 (民 101)。建構臺灣電子業財務預警 —以資料探勘技術分析。臺灣銀行季刊，63，182-215。
20. 曹正鳳、紀宏、謝邦昌 (民 103)。使用隨機森林算法實現優質股票的選擇。首都經濟貿易大學學報，2，21-27。
21. 陳南福、陳明宏、湯湧 (民 102)。大專校院柔道教練領導行為與柔道選手組織承諾之分析。臺中科大體育學刊，9，151-157。
22. 程守雄、陳亦伶 (民 98)。資料探勘於台灣股市之擇股分析。資訊科技國際研討會論文集，1-7。
23. 黃尹佐、陳筑螢 (民 96)。影響產業傳統財務績效指標之研究-以台灣資訊電子產業為例。稻江學報，5，3-17。
24. 黃加賜 (民 90)。企業的動態性財務規劃--以系統動力學方法初探。國立中山大學，財務管理研究所，9-123。

25. 黃焜煌、卓統佑(民 87)。模糊平等趨勢對台灣上市電子公司財務危機之預測。朝陽學報，5，241-259。
26. 劉文斌、左晉璋、彭克仲(民 100)。應用資料包絡分析選擇股票投資標所之研究—以上市食品產業為例。台灣農學會報，11，409-428。
27. 劉彥甫(民 101)。股權結構、信用風險與盈餘管理之關聯性。中國文化大學，會計研究所，2-39。
28. 蔡育佑(民 95)。運動教練工作壓力、工作價值觀與組織承諾之相關分析。輔仁大學體育學刊，5，106-117。
29. 蔡佳玲、洪新原、袁繼銓(民 96)。以決策樹模型探討未開立慢性病連續處方之影響因子。資訊管理學報，18，141-162。
30. 蔡榮發、張原彰(民 2004)。運用資料包絡分析法與集群分析做績效評估。中華國際機率統計與計量管理學術研討會，1，3-58。
31. 蔣德威、王佩琪、簡春雄、阮定憲(民 99)。應用決策樹技術與財務分析比率探討台灣電子股漲跌。致理技術學院，會計資訊系，1-7。
32. 鄭佩芸(民 103)。決策樹預測台灣資通訊產業技術效率-以資料包絡分析與財務危機為例。東海大學，資訊管理研究所，1-30。
33. 鄭茂松(民 105)。利用資料探勘技術建立破產預測模型。國立中央大學，資訊管理研究所，3-50。
34. 魯濤、周晶(民 99)。基於 DEA 和決策樹的服務流程標竿框架研究。統計與決策，20，50-52。
35. 蕭文姪、顏慧明、謝昌隆、周德佳(民 103)。上市櫃電子公司購併與分割宣告效果之研究。管理資訊計算，3，293-301。
36. 簡禎富、林昀萱、鄭仁傑(民 96)。建構模糊決策樹及其在有交互作用之半導體資料之資料挖礦以提昇良率之研究。品質學報，15，193-208。
37. 羅瑞霖、廖冠傑(民 85)。台灣鋼鐵產業經營績效之研究—DEA 分析法之應用。績效與策略研究，3，57-69。

38. 邊曉磊、陳學彬 (民 98)。基於 DEA 方法的我國證券公司經營績效分析。新金融，4，50-55。
39. 蘇高玄 (民 106)。利用特徵選取與增加樣本來提高決策樹在信用卡預測的準確度。管理資訊計算，6，165-145。

英文文獻

40. Abarbanell, J. S. , & B. J. Bushee ,(1997). Fundamental Analysis, Future Earnings and Stock Prices. *Journal of Accounting Research*. pp 1-24.
41. Avijan,Dutta, Gautam, Bandopadhyay, Suchismita, & Sengupta , (2012). Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. *International Journal of Business and Information*, Vol. 7, No. 1. pp 105-136.
42. Banker, R. D., A., Charnes & W. W. Cooper , (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, Vol. 30. pp 1078-1092.
43. Breiman Leo ,(2001). Statistical Modeling : The Two Cultures. *Statistical Science*, 16. pp 199-215.
44. Cabena, Hadjnian, Stadler, Verhees, & Zanasi ,(1997).*Discovering Data Mining : From Concept to Implementation* by Peter Cabena. Prentice Hall.
45. Camila, Epprecht, Álvaro, & Veiga ,(2012). Evaluating the Predictability of Stock Market Returns via STARX-Tree Models. *China-USA Business Review*, 11. pp 3-23.
46. Charnes, A., W. W. Cooper & E. Rhodes ,(1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), pp. 429-444.
47. Chehata, N. , Guo, L. , & Mallet, C. ,(2009). Airborne LiDAR Feature Selection for Urban Classification using Random Forests. *IntArchPhRS*, 38. pp 207-212.

48. Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Vol. 120, pp.253-290.
49. Fayyad, U. , Piatetsky Shapiro, G. , & Smyth, P. ,(1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17, pp.3.
50. Frawley, W. , Piatetsky Shapiro, G. , & Matheus, C. , (1991). Knowledge Discovery in Databases : an overview. *AI magazine*, 14(3). pp 57-70.
51. Grupe, F. H. , & M. M. Owrang ,(1995). Data Base Mining Discovering New Knowledge and Cooperative Advantage. *Information Systems Management*, 12(4), pp 26-31.
52. Han, J. & M. Kamber ,(2000). *Data Mining : Concepts and Techniques*. Morgan Kaufmann.
53. Hiral, R. , Patel, Satyen, M. , Parikh, Dhara, N. , & Darji ,(2016). Prediction model for stock market using news based different Classification, Regression and Statistical Techniques. *ICT in Business Industry & Government , International Conference on*. pp 3-5.
54. Hsin-Hung Chen ,(2008). Stock selection using data envelopment analysis, *Industrial Management & Data Systems*, Vol. 108 Issue : 9. pp 1255-1268.
55. Kamran Raza(2017). Prediction of Stock Market performance by using machine learning techniques. *Innovations in Electrical Engineering and Computational Technologies , International Conference on*. pp 1.
56. Lakshmi Tharun Ponnamp, K. , Srinivas, V. , Srinivasa Rao. , & Vamsi Raavi ,(2016). A comparative study on techniques used for prediction of stock market. *Automatic Control and Dynamic Optimization Techniques , International Conference on*. pp 2-6.

57. Li Xiong. , & Yue Lu. ,(2017). Hybrid ARIMA-BPNN model for time series prediction of the Chinese stock market. Information Management International Conference on. pp 94-96.
58. MacQueen, J. B. ,(1967). Some Methods for Classification and Analysis of Multivariate Observations. Fifth Berkeley Symp. on Math. Statist. and Prob, 1. pp 281-297.
59. Michael, J. A. Berry , & Gordon, S. Linoff ,(1997)Data Mining Techniques : For Marketing, Sales, and Customer Support. London. New York, NY : John Wiley & Sons.
60. Monica Tirea. , & Viorel Negru. ,(2014). Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach. Symbolic and Numeric Algorithms for Scientific Computing. International Symposium on. pp 519-526.
61. Monica Tirea. , & Viorel Negru. ,(2015). Stock Market Trading Strategies Applying Risk and Decision Analysis Models for Detecting Financial Turbulence. Symbolic and Numeric Algorithms for Scientific Computing. International Symposium on. pp 217-222.
62. Odom, M. D. , & R. Sharda ,(1990). A neural network model for bankruptcy prediction. IJCNN International Joint Conference on, San Diego, USA. pp 164-167.
63. Risul Islam Rasel, Nasrin Sultana , & Nasimul Hasan ,(2016). Financial instability analysis using ANN and feature selection technique : Application to stock market price prediction. Innovations in Science, Engineering and Technology. International Conference on. pp 1-4.
64. S. S. Panigrahi , & J. K. Mantri ,(2015). A text based Decision Tree model for stock market forecasting. Green Computing and Internet of Things. International Conference on. pp 405-411.

65. STEVEN, L. SALZBERG ,(1994). C4.5 : Programs for Machine Learning. Morgan Kaufmann Publishers. Ross Quinlan. Morgan Kaufmann Publishers. pp 235-240.
66. Vikash Singh Rajput , & Shirish Mohan Dubey ,(2016). Stock market sentiment analysis based on machine learning. Next Generation Computing Technologies. International Conference on. pp 506-510.

