

以資料探勘技術應用於時間因素與基金績效分析—以國內開放式

股票一般型基金為例

馬麗菁林美珍徐啟睿鍾宇翔蔡美惠

國立聯合大學資訊管理學系

lcma@nuu.edu.tw

摘要

隨著時代的演進，民眾理財的方式也隨著潮流逐漸多元化，其中專業性高且風險較低的共同基金，是受大眾青睞的投資工具之一，面對琳瑯滿目的基金，如何選擇報酬率高的基金是大多數投資者的共同問題，除此之外，不同投資者可能有不同投資期限偏好與需求，因此如何針對不同投資期限偏好的投資者，分析出影響報酬率的關鍵因素是一重要議題。

本研究目的主要就在於分析影響不同投資期限基金績效的重要因素，以提供不同投資期限偏好的基金投資人選擇基金之參考依據，以獲得較高的報酬率。由於影響基金績效之因素眾多，如何從中找出影響基金成效的主要因素就變的格外重要，因此本研究利用資料探勘中的決策樹演算法，分析出影響短期(三個月)、中期(六個月)或長期(一年)基金報酬率的關鍵因素。研究發現前期基金之報酬表現對未來三個月、六個月與一年的績效最具解釋力，而且當投資期限低於一年時，過去績效好(差)的基金未來之報酬表現亦較佳(差)，顯示報酬動能現象亦存在於台灣之共同基金市場。

關鍵詞：共同基金、報酬率、資料探勘、決策樹

1. 緒論

隨著時代的演進，民眾理財的方式也隨著潮流逐漸多元化，各式各樣的理財工具提供民眾不同的選擇，其中專業性高且風險較低的共同基金，是受大眾青睞的投資工具之一，共同基金主要是由大眾集資交給基金公司，由專業的基金經理人管理並運用其資金於適當之投資管道，如股票、債券等，提供給一般沒有多餘時間瞭解投資環境變化或不具投資專業知識的民眾一個投資的選擇。

基金有許多種分類方式，舉例來說，基金若依據發行方式，可分為開放及封閉型基金，主要的不同在於基金的規模是否固定；而若以發行地區來區分，則可分為國內、外基金，國內基金是指由國內投信公司發行的基金，在國內註冊，以國內投資人為銷售對象，受到國內相關法律監督的基金；而國外基金是指國外基金管理公司所發行的基金，由國外的投信擔任基金經理公司，以全球投資人為銷售對象，在國際間籌資後，再投入投資目標國家的證券市

場進行投資；另外，如果是以投資標的進行分類，則可分為股票、債券及貨幣基金，其中股票型基金還可細分為科技類、中小型、價值型、一般型、特殊類、中概股及指數型。

當投資人進行基金投資選擇時，面對琳瑯滿目的基金，如何選擇報酬率高的基金是大多數投資者的共同問題，許多研究(如李純瑩, 2004；詹振旻, 2005；黃綺年, 2004 等)即曾針對影響報酬率的因素分析，然而不同投資者可能有不同的投資期限偏好與需求，例如有人喜歡風險高的短期投資，有人偏好風險較低的長期投資¹，因此如何針對不同投資期限偏好的投資者，分析出影響報酬率的關鍵因素是一重要議題。

本研究即是針對不同投資者對投資期限的偏好，找出影響報酬率的重要因素，研究範圍設定在國內開放式股票一般型基金，由台灣經濟新報資料庫蒐集

¹ 如果資產報酬呈現平均數復回歸(mean-reverting) 現象，長期投資報酬將低於短期投資。

近五年(2001~2005)的資料共 5854 筆進行分析，以投資時間長短為三個月、六個月及一年的期限分為短期、中期及長期投資三種時間類型分析，以資料探勘中的決策樹分析法，分析與比較影響不同時間類型報酬率的關鍵因素。希望提供不同投資期限偏好的投資者選擇基金之參考依據，本研究模式亦可提供後續相關研究的參考。

2. 文獻探討

由於基金市場的蓬勃發展，許多學者都開始著手於基金各方面的研究，舉例來說，楊凱婷(2002)等提到共同基金流量會受市場報酬率影響，總體經濟變數的加入也不會影響共同基金流量與整體市場報酬間的關聯性；李純瑩(2004)對於基金的研究除了使用傳統的指標外，還加上了基金淨值報酬率為考量因素，指出基金週轉率對於基金績效的影響並無顯著性，而基金流量，市場報酬、系統風險、標準差對於基金績效有顯著性影響；Miller et al.(2003)提出了每週的不同時間點投資和基金績效表現的相關性。

除了傳統影響基金報酬率的研究外，近期有許多研究著重於應用電腦科技於基金研究上，隨著資訊科技的成熟，越來越多產業應用資訊科技工具來做各種應用，由於現在資料庫技術越來越成熟，資料的儲存成本降低且儲存方便，各產業開始將大量的資料存入資料庫，但由於資料量過於龐大雜亂，往往無法將資料作最有效的應用，因此發展出了資料探勘技術，來輔助管理者對於大量資料的分析處理，資料探勘能夠找出隱藏於資料背後有用的知識，越來越多人將資料探勘應用於不同產業，王耀德(2005)使用資料探勘於手機產業，找出使用者通路路徑的週期性；廖述賢(2005)等應用在自行車業，找出消費者對於產品功能真正的需求，設計出迎合消費者的車款，王秀育(2005)等用資料探勘於汽車業，分析舊客戶車輛回廠保養的預測。除了上述產業，也有人應用於醫療方面(例如：唐壽生，2005；董明峰，2005)。

目前也有許多研究將資料探勘技術應用在基金績效的評估，例如：黃綺年(2004)以傳統統計方法和

類神經網路對股票型基金建立分類模式，並用這兩種方法來做分析並做預測；詹振旻(2005)用資料探勘技術探討景氣循環下影響共同基金績效關鍵因素，找出股票型基金共通的因素大多為各產業的報酬率。然而大多數的研究並未針對投資期限長短和基金報酬的關係分析，本文即針對短期、中期、長期不同的投資期限之基金報酬率的影響因素進行比較分析。

3. 研究架構與步驟

本研究的研究架構主要依據 Fayyad(1996)提出的資料探勘流程，如圖 1 所示，共分為資料收集、資料前置處理、建立整合性資料庫、資料探勘和結果評估與解釋五個步驟。首先，第一步驟為資料收集，研究範圍為 2001 年 1 月至 2005 年 12 月間的開放式股票一般型基金，共 5854 筆資料，資料內容包含基金基本屬性、績效評等及投資之專業化程度三大類，主要資料來源取自台灣經濟新報資料庫，本研究先從資料庫篩選出探勘欲使用之指標，如基金週轉率、年化標準差、前一月報酬率(P1MR)、前三月報酬率(P3MR)、前半年報酬率(P6MR)和前一年報酬率(P1YR)等等十四個指標，加上基金所屬投信公司投資之基金型態集中度(focus, 反應其投資之專業化程度)²，共十五個指標(如表一)。第二步驟為資料前置處理，本研究主要進行資料清理和資料精簡，為避免資料撫平造成探勘結果有所偏差，本研究針對屬性值有遺缺的部分進行清理，即刪除該筆資料，經過資料清理之後，共 4438 筆資料；在資料精簡部分，除指標系統風險(β)以大於市場風險($\beta > 1$)、小於市場風險($\beta < 1$)分割為二區間之外，其餘指標皆利用裝箱法中的等深分割法(即是將資料劃分為數個近乎相同數量的資料區間)進行分析，接著精簡裝箱資料，以利決策樹之分析，舉例來說，本研究將半年報酬率、一年報酬率等指標分為很高、高、低和很低四區間，將前一月報酬率、三月報酬率等指標分為高、中和低三區間，

²本研究採 Herfindahl 指數代表基金所屬投信公司投資之基金型態集中度。

³ From University of Waikato New Zealand, 1999
<http://www.cs.waikato.ac.nz/ml/weka/>

其餘指標皆分為高和低二區間，詳如表 1。第三步驟為建立整合性資料庫，將前置處理過後之基金基本屬性資料(例如基金週轉率、基金淨資產等)、績效評等及投資之專業化程度之各項指標資料，建立整合性資料庫。第四步驟為資料探勘，從整合性資料庫中選取所需資料，利用資料探勘軟體 weka 3

中的 J48 決策樹進行資料探勘，分析出三月、半年和一年之決策樹，分別如圖 2、圖 3 和圖 4 所示，藉此找出影響短期、中期和長期投資的關鍵因素。最後，針對探勘出來的結果進行分析與篩選，並且解釋之，以得到有意義的知識。

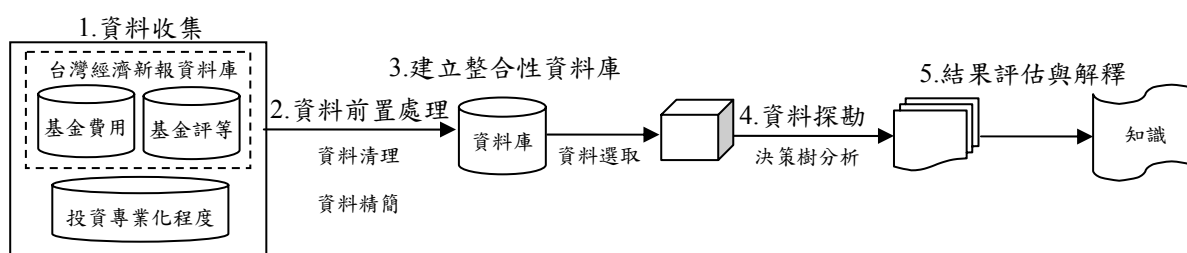


圖 1 本研究之資料探勘流程

表 1 本研究進行資料探勘所使用之指標

編號	指標名稱	指標定義	區間數
1	基金週轉率	在一段時間內，基金組合持股的更新次數。	高、低(二區間)
2	基金淨資產	基金總資產價值減去總負債的值。	高、低(二區間)
3	手續費率	為證券經紀商受託買賣股票或受益憑證成交後，向委託人收取之費用。	高、低(二區間)
4	費用率	申購共同基金所需支付的經理費、保管費、保證費及其他費用之總計。	高、低(二區間)
5	成立時間	基金的成立日至本研究期間的長短。	高、低(二區間)
6	基金型態集中度(Focus)	基金所屬投信公司投資之專業化程度。	高、低(二區間)
7	前一月報酬率(P1MR)	基金前一期報酬率。	高、中、低(三區間)
8	三月報酬率	基金最近三個月報酬率。	高、中、低(三區間)
9	前三月報酬率(P3MR)	基金前一期三個月報酬率。	高、中、低(三區間)
10	半年報酬率	基金最近半年報酬率。	很高、高、低、很低(四區間)
11	前半年報酬率(P6MR)	基金前一期半年報酬率。	很高、高、低、很低(四區間)
12	一年報酬率	基金最近一年報酬率。	很高、高、低、很低(四區間)
13	前一年報酬率(P1YR)	基金前一期一年報酬率。	很高、高、低、很低(四區間)
14	年化標準差	用以衡量共同基金投資風險，評估基金資產淨值的變化及基金穩定性。	高、低(二區間)
15	貝他值(β)	為一風險指數，用以衡量基金報酬率對市場漲跌的敏感程度。	高、低(二區間)

4. 資料探勘結果分析

經資料探勘後，分別產生三月、半年和一年報酬率之決策樹如圖 2、3、4，因篇幅限制本研究僅呈現最佳、最差兩個極端的結果來表示決策樹，並將個數小於 30 與正確率小於 60%(三月報酬率)或 70%(半年及一年報酬率)的葉節點刪除，因這些節點較不具代表性，在圖 2~4 中，被省略之節點均以“ ”表示，另外，在圖中葉節點內的第一個數值為符合條件的總個數，第二個數值代表正確率。

(1)短期三個月報酬率分析

首先我們針對短期三個月報酬率分析，由圖 2 可明顯看出，影響未來三個月報酬率之關鍵因素為前三個月報酬率，其次為費用率和前一個月報酬率，舉例來說，當前三個月報酬率表現維持在前三分之一(高)且費用率較低時，三個月報酬率表現是前三分之一(高)的正確率為 78%；當前三月報酬率表現維持在前三分之一(高)，費用率卻較高時，只有前一個月報酬率排名位於前三分之一(高)而且前一年報酬率排名位於前四分之三(很高、高、低)時，未來三個月報酬率才可能位於前三分之一(高)。由決策樹之中間樹狀圖，發現當前三個月報酬率表現平平(報酬率排名位於中間)時，其未來三個月之報酬率表現將不易單純由少數因子來解釋。此時，如果前一月報酬率表現不在後三分之一(中、高)，次一個觀察指標為前一年報酬率，如果前一年報酬率排名中下水準(低)，年化標準差又偏高時，三月報酬率通常為高；如果前一年報酬率排名中下水準(低)，費用率偏低，且前六個月報酬率排名位於中上(高)時，三月報酬率通常較高。如果前一月報酬率表現在後三分之一(低)，次一個觀察指標為前半年報酬率，如果前半年報酬率排名屬於前二分之一(高、很高)，年化標準差偏高或前一年報酬率排名屬極佳或極差之群組時，三月報酬率通常較低。相反地，當前三個月報酬率表現維持在後三分之一(低)、無論費用率高低，其未來三個月報酬率大多為低，此決策樹的平均正確率為 70.51%。

(2)中期六個月報酬率分析

中期六個月報酬的分析結果如圖 3 所示，最主要的影響因素就為前期報酬率。如果基金前半年報酬率

現維持在前四分之一間(很高)、前三個月報酬率表現又在前三分之一(高)，此基金將有 90%的機率在未來半年報酬率表現排名前三分之一；如果基金前半年報酬率表現維持在前四分之一間(很高)，但前三個月報酬率表現平平(中)，則需進一步觀察其過去一年之報酬率，如果此基金前一個月的報酬率不在後三分之一(低)，則此支基金將有 81%的機率在未來半年報酬率表現排名前三分之一。

如果基金前半年報酬率表現維持在前二分之一至四分之三間(高)、前三個月報酬率有前三分之一(高)的表現，並且承擔的投資風險高(年化標準差高)，前一年報酬率的表現也在前二分之一(高、很高)，則此支基金較有可能會有很高的半年報酬率。反之，如果某基金的前半年報酬率表現落在後四分之一(很低)，或前半年報酬率表現平平(排名 25%~75%，低、高)且前三個月報酬率落在後三分之一(很低)，則即使其承擔了高的投資風險(年化標準差或貝他值)，此基金的半年報酬率仍有很高的機率會落在「最差的 25%之內(很低)」，此決策樹的平均正確率為 72.10%。

(3)長期一年報酬率分析

長期一年報酬率分析由圖 4 可發現一年報酬率高低的首要判斷因素是前一年報酬率，之後可藉由前一年報酬率是否很高、很低來找出規則。假設前一年報酬率為後 25%(很低)，可發現無論費用率高低或前三個月報酬率高低，有很大的比例其未來一年的報酬率將會很低。假設前一年報酬率績效為前四分之一(很高)時，接下來可以觀察前半年報酬率績效，當前半年報酬率為前 25%(很高)時，無論年化標準差為何，一年的報酬率有相當高的可能性會有不錯的表現；此時當前半年報酬率為介於 25%~50%之間(高)時，只要前三個月報酬率不在後三分之一(低)，則此支基金將有很大的機會在未來一年報酬率表現排名前四分之一。前一年報酬率績效為前 25%(很高)時，即使當前半年報酬率在後 25%(很低)時，只要前一個月報酬率仍屬於較差之群組(中、低)，其未來一年之報酬率仍有很高的可能會排名前四分之一，由此可觀察到報酬率會有波動的情形出現，前

一年表現好，接下來的半年報酬率偏低，然而一年的報酬率會反彈回升最後仍會有不錯的表現，此決策樹的平均正確率為 74.49%。

(4)綜合分析

綜合分析所有決策樹，可以發現不論是短期、中期或長期的報酬率，主要的影響因素都是對應之前一短、中與長期的報酬率，次要的影響因素則是時間較短的前期報酬率，依序類推。例如在一年報酬率的決策樹中，首先考慮的影響因素就是前一年報酬率，接著是前半年報酬率，最後納入考慮的則是前三個月報酬率。

此外，基金屬性中的費用率對於報酬率的影響程度，會因分析不同期限的報酬率而有所改變，以期限愈短的報酬率影響愈大，但由前三個月報酬率高(低)時，無論費用率高低，其後續三個月報酬率皆較有可能為高(低)，顯示其影響程度亦是有限。此外

，其它基金特徵因子，如基金週轉率、基金淨資產、手續費率、成立時間與基金所屬投信公司投資之基金型態集中度，似乎並非基金績效的主要決定因素。

決策樹中還有一項特性就是，針對愈長期的報酬率預測，其主要的影響因素影響的程度就愈強，光從主要的影響因素推估出的結果，其準確度就愈高。以一年報酬率的預測為例，如果前一年報酬率的表現在前 25%以上(很高)，則未來一年報酬率將很高，相反地，如果前一年報酬率的表現在後 25%(很低)，未來一年報酬率將很低。此現象亦存在於以三個月報酬率和六個月報酬率為衡量指標時，反應當投資期限低於一年時，此種過去績效好(差)的基金未來之報酬表現亦較佳(差)的報酬動能現象亦存在於台灣之共同基金市場。

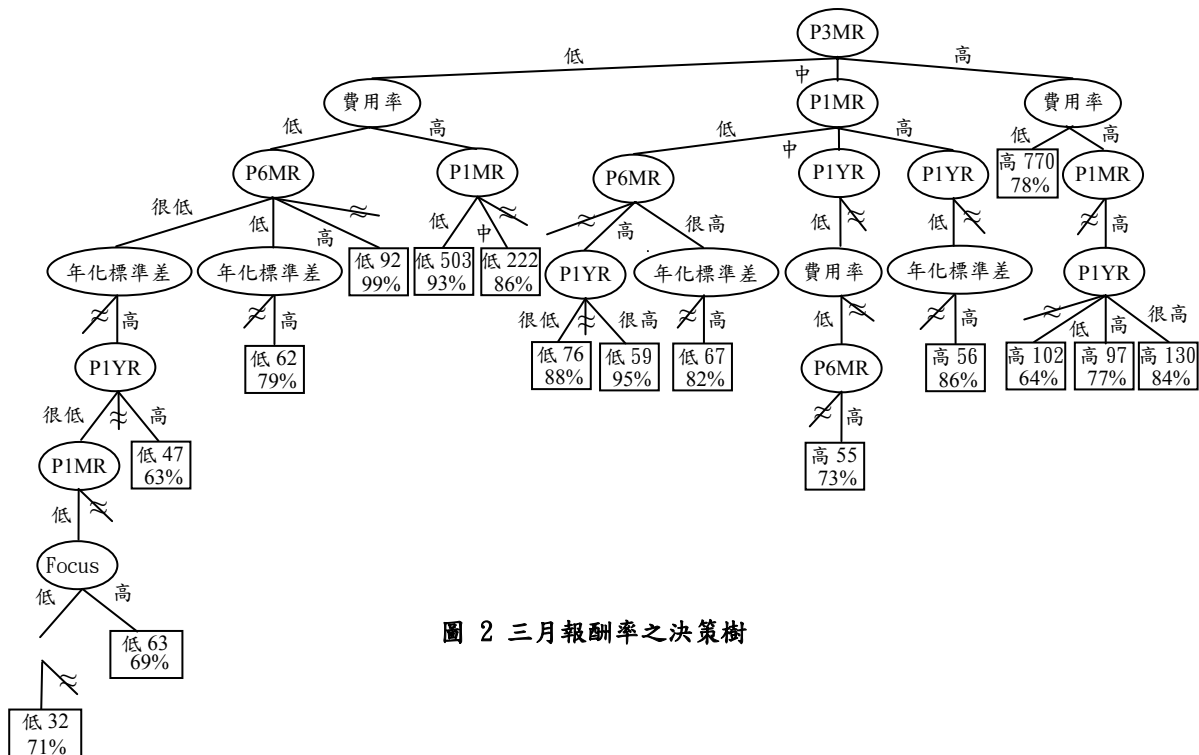


圖 2 三月報酬率之決策樹

5. 結論

本研究依據投資者對不同投資期限的偏好，利用資料探勘中的決策樹演算法分別分析出影響短期(三個月)、中期(半年)和長期(一年)報酬率之決策樹，主要目的在於分析與比較影響不同投資期限基金績效的關鍵因素，希望提供不同投資期限偏好的投資者選擇基金之參考。

本研究結果發現，前期基金之報酬表現對未來三個月、六個月與一年的績效最具影響力，且當投資期限低於一年時，過去績效好(差)的基金未來之報酬表現亦較佳(差)，顯示報酬動能現象亦存在台灣之共同基金市場，而關於其它基金特徵因子，如基金週轉率、基金淨資產、手續費率、成立時間與基金所屬投信公司投資之基金型態集中度，對基金績效的影響力不大。其次，除前期報酬率之外，基金屬性中的費用率對於報酬率的影響程度，會因分析不同期限的報酬率而有所改變，以期限愈短的報酬率影響愈大。

本研究結果除可提供基金投資人參考依據外，本研究模式也可提供後續相關研究參考，由於目前研究僅針對開放式股票一般型的報酬率分析，後續研究可朝開放式股票其他類型，如中概股、科技股等的影響因素比較。此外，本研究的績效指標以報酬率為主，後續研究亦可採用其他績效評估指標，如 Sharp、Treyner 等指標，做比較分析。

參考文獻

- [1] 王秀育、羅家蓉、楊玲惠，2005，『資料探勘於汽車業的應用—車輛不回廠保養預測』，國際資訊管理學術研討會. 第十六屆。
- [2] 王耀德，2005，『以抽樣技術探勘行動電話使用者之週期性移動樣式』，國際資訊管理學術研討會. 第十六屆。
- [3] 李純瑩，2004，『國內股票型基金績效之影響因素研究』，世新大學經濟學系碩士論文。
- [4] 唐壽生、林怡君、韓懷恩、簡鈺玫，2005，『資料探勘技術應用於結核病完治的預測—以嘉義某區域教學醫院為例』，國際資訊管理學術研討會. 第十六屆。
- [5] 梁恩輝、董明峰、林英潔、黎濤、陳嘉浩、陳怡珊、楊心吟、王水深、周迺寬，2005，『應用資料探勘技術在心臟外科手術風險分析之研究』，國際資訊管理學術研討會. 第十六屆。
- [6] 理財工具及其應用編撰委員會，2005，理財工具及其應用，台北：財團法人台灣金融研究院。
- [7] 黃綺年，2004，『統計方法與類神經網路應用於國內開放式股票型基金投資績效分類及投資報酬率預測之研究』，國立成功大學統計研究所碩士論文。
- [8] 楊凱婷、劉美纓，2002，『共同基金流量與證券市場報酬率關聯性之研究』，東吳大學企業管理學系碩士論文。
- [9] 詹振旻，2005，『以資料探勘技術探討景氣循環下影響共同基金績效關鍵因素』，銘傳大學資訊管理研究所碩士論文。
- [10] 廖述賢、劉基金、溫志皓，2005，『資料探勘應用於中小企業新產品發展之研究-以自行車產業為例』，國際資訊管理學術研討會. 第十六屆。
- [11] U. Fayyad , G. Piatetsky-Shapiro and P. Smyth, From Data Mining to Knowledge Discovery: An overview. "In advances in Knowledge Discovery and Data Mining", pp.471-493, 1996.
- [12] E. M. Miller and L. J. Prather and M. I. Mazumder, "Day-of-the-Week Effect Among Mutual Funds," Quarterly Journal of business & Economics, Vol. 42, Nos. 3 and 4, pp. 113-128, 2003。