

共同基金卓越績效的認定與評估： 新逐步檢定法的應用

莊惠菁* 管中閔

本文採用能避免資料窺探偏誤的檢定方法來認定表現卓越的基金，並評估由這些卓越基金所建構之資產組合的樣本外績效。與莊惠菁與管中閔(2010)的實證方法相比，本文的檢定方式放寬了「全族錯誤率」的定義，因此能正確的認定出更多超越大盤指數的卓越基金，並建構表現優異的資產組合。實證結果發現，對於臺灣股票型與平衡型共同基金在2011年至2017年間，不論利用均數，夏普值，資訊比率，或是多因子異常報酬為績效指標，我們都能認定出相當數量顯著優於台股加權指數的卓越基金。而利用這些樣本內卓越基金建構的資產組合，其樣本外績效和持有期間報酬亦多能超越台股加權，臺灣50與MSCI臺灣指數。

關鍵詞：資料窺探偏誤、全族錯誤率、共同基金、逐步真實性檢定、逐步卓越預測力檢定。

「政策與管理意涵」

本研究發現，採用能避免資料窺探偏誤的適當檢定方式，可以幫助投資人正確的找出更多優於大盤之卓越基金，並依此來建構可獲利的投資組合。

* 通訊作者：莊惠菁。地址：320 臺灣桃園市中壢區遠東路 135 號。電話：886-34638800。作者感謝科技部研究計畫經費補助。計畫編號：MOST 107-2632-H-155-001(莊惠菁)。MOST 106-2410-H-002-014-MY2(管中閔)。電子郵件：hcchuang@saturn.yzu.edu.tw(莊惠菁)，ckuan@ntu.edu.tw(管中閔)。莊惠菁：元智大學管理學院，管中閔：臺灣大學財務金融學系。

壹、前言

市場上究竟是否有績效卓越的共同基金？這是學術研究的重要題目，也是實務界關切的議題。一般投資人衡量基金績效的方法是，將基金的報酬率或是調整風險後的報酬率(如資訊比率，夏普值等) 排序，並與大盤指數的表現相互比較；許多基金分析網站與新聞媒體亦採用類似的評估方法。(註1)除了報酬與風險調整後的報酬外，學者也採用多因子模型的異常報酬作為績效衡量指標。然而，基金的績效一旦超越大盤(或基準)指數，這個結果是否具有統計上的顯著意義，或者只是在特定樣本期間因抽樣誤差等因素導致的幸運結果，需要透過嚴謹的計量檢定來加以判別。

在檢定基金績效時，我們必須避免資料窺探偏誤 (data snooping bias)，以免造成錯誤推論。當我們需要同時檢定許多假設 (模型或基金績效)，這是一個多重檢定 (multiple testing) 的問題，此時若我們只檢定其中一個假設，卻忽略其他必須同時考慮的假設，就可能產生資料窺探偏誤；參見 White (2000)。為了避免資料窺探的問題，White (2000) 提出了真實性檢定 (Reality Check, RC)，利用自我重抽樣法 (bootstrap) 來估計檢定統計量的虛無分布；Hansen (2005) 的卓越預測力 (Superior Predictive Ability, SPA) 檢定則修正了 RC 檢定統計量，使其實際的分布更接近其理論上的分布。但 RC 和 SPA 這兩種檢定只能推論的確有顯著優於比較基準的模型(或基金)，卻無法確認究竟有多少這種模型(或基金)，Romano and Wolf (2005) 的逐步真實性 (Step-RC) 檢定和 Hsu et al. (2010) 的逐步卓越預測力 (Step-SPA) 檢定，將前述方法從單一步驟延伸為數個步驟，使研究者得以找到更多顯著優於比較基準的模型。

註1：例如：晨星基金的研究評級 (Morningstar Analyst Rating) 反映了調整長期風險後的基金報酬率勝過同組別基金及基準指數的能力；見 <http://www.morningstar.com/invGLOSSARY/morningstar-analyst-rating-for-funds.aspx>。基智網的基金評等方法則將長短期基金報酬，夏普值 與過去打敗指數的月份數，綜合為一指標；見 <https://www.moneydj.com/funddj/yp999998.djhtm>。新聞媒體報導也常將基金績效與大盤指數績效比較，如 <https://money.udn.com/money/story/5618/4068716> 等。

在財務或經濟研究中，無資料窺探偏誤的檢定方法已經有了許多應用。例如，White (2000)，Hsu and Kuan (2005), Qi and Wu (2006), Hsu et al. (2010), Shynkevich (2012), Kuang et al. (2014), 和 Hsu et al. (2016) 針對技術分析檢定其獲利性；Romano and Wolf (2005) 檢定基金的獲利能力；Neuhierl and Schlusche (2011), Genre et al. (2013), 和 Goyal and Wahal (2015) 則比較特定交易策略的績效。這些檢定方法也被用來評估財務或經濟計量模型的表現。例如，比較資產組合與資產定價模型：Patton (2004), Chen and Ludvigson (2009) 和 Patton and Timmermann (2010); 比較不同的波動性 (Volatility) 模型：Hansen and Lunde (2005), Koopman et al. (2005), Martin et al. (2009), Patton (2011) 和 Liu et al. (2015); 比較不同涉險值 (Value-at-Risk) 模型：Bao et al. (2006); 以及比較不同總體模型的表現：Hanck (2009), Molodtsova and Papell (2009) 和 Kao et al. (2013)。這些研究提供了許多新的結果，也使我們對這些議題有更清楚的認識。在臺灣基金績效的研究中，莊惠菁與管中閔 (2010) 考慮了資料窺探偏誤，並且採用 Step-RC 與 Step-SPA 檢定來評估 2002–2007 年間臺灣的股票型與平衡型共同基金。時隔十年，我們重新檢視當時所認定的基金 (編號 18, 53, 182) 在近年的績效；請見圖 1。由此圖可以發現，其中二檔基金(編號 18 與 182)，即使歷經 2008 年的金融風暴，其報酬仍穩健的超越臺股加權指數。(註2)這結果顯示，無資料窺探偏誤的檢定具有實務上的參考價值。

如果能認定出卓越的基金，我們就有可能建構出未來表現較好的投資組合或組合型基金 (fund of funds)。若每檔基金的風險程度相同，一個包括較多基金的平均加權資產組合，相較於僅由少數基金形成資產組合，其風險較低。由於前述莊惠菁與管中閔 (2010) 採用的檢定方法相對保守，在實證上通常僅能認定出少數基金，因此可能遺漏了其他也有卓越績效的基金；當找到的卓越基金太少，實務上也很難做適當的資產組合。本文於是採用較 Step-RC

註2：基金的表現與經理人有關。我們追蹤此三檔基金的經理人替換情形發現，在 2007 年至 2016 年間，編號 53 號基金頻繁更換經理人 (總共十位)，經理人最長任期不超過 15 個月。相較之下，編號 182 號基金從以前至今從未換過經理人；編號 18 號基金更換了 4 位經理人，其中一位任期長達 43 個月。

與 Step-SPA 檢定更具檢定力的方法，設法認定出更多臺灣共同基金中績效卓越的基金，以利投資者建構適當的資產組合。

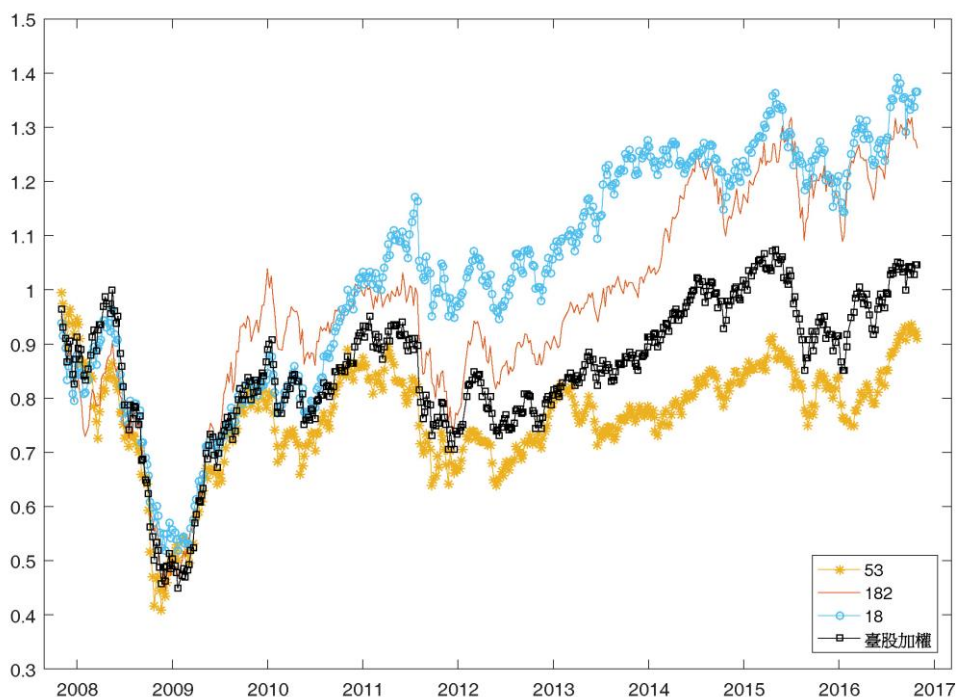


圖 1 莊惠菁與管中閔 (2010) 文中卓越基金的樣本外持有期間報酬

從方法上看，Step-RC 和 Step-SPA 檢定是在逐步檢定時持續控制“全族錯誤率”(familywise error rate, FWER)，即至少有一個錯誤拒絕的機率，不超過某個給定的顯著水準。Romano and Wolf (2007) 指出，當檢定的模型 (或基金) 數眾多時，以“至少有一個錯誤拒絕”來決定錯誤率是一個相對嚴格的概念，因此使檢定方法變得保守。文獻上已有許多檢定方法放寬了錯誤率的定義，以“至少有 k 個錯誤拒絕”來定義錯誤率 (k 是執行檢定者事先選取的一個整數)，如 Hommel and Hoffmann (1988) 和 Lehmann and Romano (2005)。Romano and Wolf (2007) 於是根據這種錯誤率概念提出了 Step-RC(k) 檢定，確保其“一般化全族錯誤率”(亦即至少有 k 個錯誤拒絕的全族錯誤率) 低於給定的顯著水準。由於此方法放寬了錯誤率的定義，其檢定力必然不低於 Step-RC 的檢定力。基於相同的想法，Hsu et al. (2014) 提出 Step-SPA(k)

檢定，其檢定力不僅不低於 Step-SPA 的檢定力，也不低於 Step-RC(k) 的檢定力。本文的研究即利用 Hsu et al. (2014) 所提出之 Step-SPA(k) 檢定，評估臺灣共同基金績效表現。

在本文的實證中，我們除認定樣本內具有顯著績效之基金外，並利用這些卓越基金建構資產組合，然後評估其樣本外的績效指標和持有期間報酬。我們採用的績效衡量指標包括：基金報酬均數與大盤的差異，夏普值與大盤的差異，因子異常報酬，以及資訊比率。(註3) 實證結果顯示，臺灣股票型與平衡型共同基金在 2011–2017 年間，不論採用那一種績效指標，不論是否考慮手續費，我們都可認定多檔基金在樣本內具有顯著優於台股加權指數之表現。利用這些樣本內認定的卓越基金，我們建構不同的資產組合，並與台股加權，臺灣 50，和 MSCI 臺灣指數進行樣本外表現的比較。我們發現絕大多數的資產組合，在樣本外均能超過這些指數的均數與夏普值，且有正的資訊比率與多因子異常報酬。這些實證結果顯示，本文採用的檢定方法的確有助於選取卓越基金，並建構表現優異的資產組合。

本文結構如下。第二節介紹 Step-RC(k) 和 Step-SPA(k) 檢定方法，並簡述其理論性質。我們在第三節介紹本研究的資料，然後討論樣本內的檢定結果，以及認定的卓越基金的資產組合在樣本外的表現。最後一節為本文結論。

貳、方法

一、標準化 Step-SPA(k)檢定法

我們首先介紹本研究採用的檢定方法及執行步驟；此方法的理論性質可參見 Romano and Wolf (2007) 和 Hsu et al. (2014) 文中的討論。我們將基金報酬與大盤報酬的資料表示為矩陣 $X_{T \times (S+1)}$ ，前 S 欄為基金報酬（或基金之超額報酬），最後一欄為大盤之報酬(或大盤之超額報酬)其元素 $x_{t,j}$, $1 \leq t \leq$

註3：我們感謝匿名評審的建議，在本研究中考慮了資訊比率和多因子異常報酬作為績效衡量指標。此外，我們在績效評估的討論中亦考慮了交易成本。

$T, 1 \leq j \leq S + 1$ ，代表第 j 檔基金在第 t 期的報酬。我們將績效指標表示為 $\theta := (\theta_1, \theta_2, \dots, \theta_S)'$ 。例如，令 μ_j 和 μ_ϕ 分別代表第 j 檔基金報酬與大盤指數報酬的母體均數，一種績效指標是 $\theta_j = \mu_j - \mu_\phi$ ，亦即基金報酬均數與大盤報酬均數的差距。本研究將採用不同的績效指標，各指標和其估計方式將在下一小節說明。

我們將檢定 S 檔基金的績效，虛無假設為 S 個不等式假設：

$$H_{0,j}: \theta_j \leq 0, j = 1, \dots, S,$$

亦即沒有任何一檔基金的績效超越大盤指數。這種多重不等式檢定和傳統檢定方法不同之處有二。第一，由於虛無假設是不等式，必需選擇一個方式來決定檢定統計量在虛無假設下的分布；最常見的方式就是根據最不利狀態 (least favorable configuration, LFC)，亦即計算 $\theta_j = 0, j = 1, \dots, S$ 之下的分布。第二，檢定方法必須控制所有假設的全族錯誤率 FWER，而非單一假設下的型一誤差機率。令 I_0 為所有正確的虛無假設的集合，則 FWER 就是 S 個假設中至少錯誤拒絕一個正確的虛無假設的機率：

$$\text{FWER} = P\{\text{拒絕至少一個 } H_{0,j} : j \in I_0\}.$$

正確的檢定方法必須控制 FWER 不超過某個給定的顯著水準 α (如 5%)，或是大樣本下， $\limsup_{T \rightarrow \infty} \text{FWER} \leq \alpha$ 。文獻上已有許多控制 FWER 的方法，例如 Bonferroni p 值調整法，White (2000) 的 RC 檢定，和 Hansen (2005) 的 SPA 檢定。其中 RC 檢定就是根據 LFC 得到檢定量的極限分布，而 SPA 檢定則修正了 RC 檢定的統計量，使其分布更接近理論分布。

前述的 RC 和 SPA 檢定若拒絕虛無假設，我們只能推論的確有績效顯著優於大盤的基金，但不知道究竟有多少這樣的基金。將這些檢定延伸為逐步 (stepwise) 檢定，則可以透過多次檢定步驟來找出更多優於大盤的基金，同時確保 FWER 得到適當控制；例如，Holm (1979) 和 Benjamini and Hochberg (1995) 的逐步 p 值調整法，Romano and Wolf (2005) 的 Step-RC 檢定，和 Hsu et al. (2010) 的 Step-SPA 檢定。在實證上，當檢定的基金數 S 很大時，只拒

絕一個真實假設就被視為錯誤是相對嚴苛的標準，控制 FWER 則使得檢定方式變得非常保守。因此，文獻上也考慮另一個標準：拒絕至少 k 個真實假設 (如 $k = 3$ 或 5) 才被視為錯誤，而檢定方法也可將全族錯誤率放寬到 $\text{FWER}(k)$:

$$\text{FWER}(k) = P\{\text{拒絕至少 } k \text{ 個 } H_{0,j} : j \in I_0\},$$

並控制 $\text{FWER}(k)$ 不超過給定的顯著水準 α ，或是大樣本下， $\limsup_{T \rightarrow \infty} \text{FWER}(k) \leq \alpha$ 。控制 $\text{FWER}(k)$ 的檢定方法包括 Hommel and Hoffmann (1988) 的延伸 Bonferroni p 值調整法，Lehmann and Romano (2005) 多階段延伸 Bonferroni p 值調整法，Romano and Wolf (2007) 延伸 Romano and Wolf (2005) 的 Step-RC(k) 檢定，和 Hsu et al. (2014) 延伸 Hsu et al. (2010) 的 Step-SPA(k) 檢定。(註4)

本文採用標準化的 Step-SPA(k) 檢定。由於 SPA 檢定來自對 RC 檢定的修正，以下我們先介紹 Romano and Wolf (2007) 的標準化 Step-RC(k) 檢定，然後討論 Hsu et al. (2014) 的標準化 Step-SPA(k) 檢定。對於每一個假設，定義相對應的檢定統計量為 w_j ，標準化的檢定統計量為 $z_j := \sqrt{T}w_j/\hat{\sigma}_j$ 其中 $\hat{\sigma}_j$ 為 w_j 的標準差的估計式。將所有標準化的檢定統計量由大到小排列，即

$$z_{r_1} \geq z_{r_2} \geq \dots \geq z_{r_S}.$$

其中下標 r_1 代表最大的標準化檢定統計量的基金指標，而 r_S 代表最小的標準化檢定統計量的基金指標。而相對應的虛無假設為 $H_{0,r_1}, \dots, H_{0,r_S}$ 。假定有一個集合 $\{y_j : j \in K\}$ ，包含了 $|K|$ 個實數，且 y_j 由大至小排序： $y_{(1)} \geq \dots \geq y_{(|K|)}$ ，我們將此集合第 k 大的數字表示為 $k \max_{j \in K} (y_j) = y_{(k)}$ 。因此，對 $K \subset \{1, 2, \dots, S\}$ ，我們定義：

註4：文獻上還有其他一般化的錯誤率概念，如錯誤拒絕比 (false discovery proportion, FDP) 和錯誤拒絕率 (false discovery rate, FDR)，前者是指錯誤拒絕的假設(模型)數佔被拒絕的假設(模型)數的比值，後者則為此比值的期望值。Hsu et al. (2014) 曾延伸 Step-SPA(k)，得到控制 FDP 的逐步檢定；Benjamini and Hochberg (1995) 和 Benjamini and Yekutieli (2001) 則提出控制 FDR 的逐步檢定法。文獻上相關應用包括：Romano et al. (2008) 和 Barras et al. (2010) 檢定避險基金與共同基金的績效；Bajgrowicand and Scaillet (2012) 和 Harvey and Liu (2014) 評估交易策略的獲利性；Harvey et al. (2016) 討論了資產定價因子的真實解釋性。

$$c_K(1 - \alpha, k, P) := \inf \left\{ x: P \left\{ k \max_{r_j \in K} \left(\frac{\sqrt{T}(w_{r_j} - \theta_{r_j})}{\hat{\sigma}_{r_j}} \right) \leq x \right\} \geq 1 - \alpha \right\},$$

此為在機率分布 P 下， $k \max_{r_j \in K} \left(\sqrt{T} w_{r_j} - \theta_{r_j} / \hat{\sigma}_{r_j} \right)$ 的 $1 - \alpha$ 分位數。理論上，如果我們知道所有真實假設的集合 \tilde{K} ，即 $r_j \in \tilde{K}$ 若且為若 $j \in I_0$ ，則可設定拒絕虛無假設的規則為：若 $z_{r_j} > c_{\tilde{K}}(1 - \alpha, k, P)$ ，則拒絕 H_{0,r_j} 。此一作法則可控制 $\text{FWER}(k)$ 在顯著水準 α 之下：

$$\begin{aligned} \text{FWER}(k) &= P \left\{ \text{拒絕至少 } k \text{ 個 } H_{0,r_j} : r_j \in \tilde{K} \right\} \\ &= P \left\{ k \max_{r_j \in \tilde{K}} \left(\frac{\sqrt{T} w_{r_j}}{\hat{\sigma}_{r_j}} \right) > c_{\tilde{K}}(1 - \alpha, k, P) \right\} \\ &\leq P \left\{ k \max_{r_j \in \tilde{K}} \left(\frac{\sqrt{T}(w_{r_j} - \theta_{r_j})}{\hat{\sigma}_{r_j}} \right) > c_{\tilde{K}}(1 - \alpha, k, P) \right\} \\ &= \alpha, \end{aligned}$$

前面的不等式成立是因為，如果 $r_j \in \tilde{K}$ ，則虛無假設為真，此時 θ_{r_j} 為一個小於或等於零的數字，最後的等式則是根據 $c_{\tilde{K}}(1 - \alpha, k, P)$ 的定義。

在實證分析中，我們不知道集合 \tilde{K} ，也不知道真實的機率分布 P ，所以無法得知 $c_{\tilde{K}}(1 - \alpha, k, P)$ 。Romano and Wolf (2007) 提出了一個保守的作法：利用所有指標的集合 $\{1, 2, \dots, S\}$ 取代 \tilde{K} 。因為 \tilde{K} 是 $\{1, 2, \dots, S\}$ 的子集合， $c_{\tilde{K}}(1 - \alpha, k, P) \leq c_{\{1, 2, \dots, S\}}(1 - \alpha, k, P)$ 。統計量若大於 $c_{\{1, 2, \dots, S\}}(1 - \alpha, k, P)$ 也必然大於 $c_{\tilde{K}}(1 - \alpha, k, P)$ (所以這是保守的作法)。此外，Romano and Wolf (2007) 建議利用重複自我抽樣法所得的 \hat{P}_T 來估計 P 。綜合前述，拒絕一個虛無假設的規則為：若 $z_{r_s} > \hat{c}_{\{1, 2, \dots, S\}}(1 - \alpha, k, \hat{P}_T)$ ，則拒絕 H_{0,r_s} 。這種作法在大樣本下仍可控制 $\text{FWER}(k)$ 小於顯著水準 α 。

在逐步檢定法中，讓 R_1 代表前一階段所拒絕的虛無假設總數。如果 $R_1 < k$ ，則此檢定到此結束，因為很有可能第一階段被拒絕的假設都是真的假設。但若 $R_1 \geq k$ ，則可能有些假設被錯誤的拒絕，此時我們進行下一階段的檢定。在第二階段中，若是可以得到最理想的 $c_{\tilde{K}}(1 - \alpha, k, P)$ ，則可控制

$\text{FWER}(k)$ 小於顯著水準。但 \tilde{K} 和 P 仍是未知，我們除了以自我重複抽樣所得的 \hat{P}_T 來估計 P 外，Romano and Wolf (2007) 建議利用前一階段的資訊，來獲得一個較 $\hat{c}_{\{1,2,\dots,S\}}(1-\alpha, k, \hat{P}_T)$ 小的臨界值(使得檢定不會太過保守)，來當作執行第二階段時的拒絕標準。令

$$\hat{c}_2 = \max_{N \subset \{1,2,\dots,R_1\}, |N|=k-1} \left\{ \hat{c}_K(1-\alpha, k, \hat{P}_T) : K = N \cup \{R_1 + 1, \dots, S\} \right\}.$$

此臨界值為在 $\binom{R_1}{k-1}$ 的指標組合 N 和前一階段未被拒絕的假設指標的聯集下，取最大值。舉例而言，若依照檢定統計量大小排序後之虛無假設編號為 $\{1, 2, 3, 4, 5, 6, 7\}$ ， $R_1 = 4$ ， $k = 3$ ，表示前四個虛無假設 $j = 1, 2, 3, 4$ 已在第一階段中被拒絕。在第二階段時，我們將被拒絕的四個假設中任選出 $k - 1 = 2$ 個，如 $\{1, 2\}, \{1, 3\}, \{1, 4\}, \dots, \{3, 4\}$ ，這些集合就是 N 。我們再根據 N 以及尚未被拒絕的假設，建構 $\hat{c}_K(1-\alpha, k, \hat{P}_T)$ 。我們計算所有可能的 $\hat{c}_K(1-\alpha, k, \hat{P}_T)$ ，再取這些 \hat{c}_K 中的極大值作為第二階段的拒絕標準 \hat{c}_2 。這方法的直觀想法為，若我們在第一階段中已經控制了至少有 k 個錯誤拒絕的機率，則至多只有 $k - 1$ 個真實的假設會在第一階段被拒絕。因為理想的拒絕標準 $c_{\tilde{K}}(1-\alpha, k, P)$ 中的 \tilde{K} 集合為真實假設的集合，我們於是可以在這一階段將 $\{1, 2, \dots, S\}$ 範圍縮小到 $\{N \cup \{R_1 + 1, \dots, S\}\}$ 。但我們並不清楚那些在第一階段被拒絕的假設為真實假設，保守的作法就是就根據所有可能的組合都計算臨界值，再選取其中最大值為第二階段的拒絕標準。第三階段和以後階段都可用類似的方法找到各階段的拒絕標準。

我們將 Romano and Wolf (2007) 的標準化 Step-RC(k) 的執行步驟整理如下：

- 步驟一：將標準化的檢定統計量 z_j ，由大到小排序，定義 r_1 為標準化檢定統計量最大的假設，而 r_S 為標準化檢定統計量最小的假設。
- 步驟二：對所有的假設 $1 \leq j \leq S$ ，若是 $z_{r_j} > \hat{c}_1$ 拒絕虛無假設 H_{0,r_j} ，其中

$$\hat{c}_1 := \hat{c}_{\{1,2,\dots,S\}}^{RC}(1 - \alpha, k, \hat{P}_T)$$

步驟三：定義 R_1 為第一階段拒絕的假設數目。如果 $R_1 < k$ 則檢定程序停止。若

$$R_1 \geq k, \text{ 則令 } h = 2。$$

步驟四：對於 $R_{h-1} + 1 \leq j \leq S$ 的虛無假設而言，若 $z_{r_j} > \hat{c}_h$ ，拒絕虛無假設，其中

$$\hat{c}_h = \max_{N \subset \{1,2,\dots,R_{h-1}\}, |K|=k-1} \left\{ \hat{c}_K^{RC}(1 - \alpha, k, \hat{P}_T) : K = N \cup \{R_{h-1} + 1, \dots, S\} \right\}. \quad (1)$$

步驟五：若沒有任何假設被拒絕，則檢定停止。若有假設被拒絕，則令 R_h 為到目前為止被拒絕的虛無假設總數；設 $h = h + 1$ ，回到步驟四。

為獲得檢定步驟所需的臨界值，我們利用 Politis and Romano (1994) 的定態自我重抽樣 (stationary bootstrap) 法建立原始資料的 B 個重製樣本 (replicated samples)， $\{X_{T \times (S+1)}^b\}_{b=1}^B$ 。(註5) 我們利用下面步驟來獲得各階段的臨界值：

臨界值步驟一：對每一自我重抽樣樣本 $X_{T \times (S+1)}^b$ 計算檢定統計量及其標準差： $w_{r_1}^b, \dots, w_{r_S}^b$ 和 $\hat{\sigma}_{r_1}^b, \dots, \hat{\sigma}_{r_S}^b$ 。

臨界值步驟二：對每一自我重抽樣樣本 $X_{T \times (S+1)}^b$ 和任意的指標集合 $K, K \subset \{1, \dots, S\}$ ，定義

$$kmax_K^{b,RC} = kmax_{j \in K} \left(\frac{\sqrt{T}(w_j^b - \hat{\theta}_{r_j})}{\hat{\sigma}_{r_j}^b} \right).$$

則可定義 $\hat{c}_K^{RC}(1 - \alpha, k, \hat{P}_T)$ 為 $\{kmax_K^{b,RC}\}_{b=1}^B$ 的 $1 - \alpha$ 的分位數。

註5：詳細的定態自我重抽樣法的執行方式請參見莊惠菁與管中閔 (2010) 文章附錄或 Lahiri (2013) 書中討論。

臨界值步驟三： $h = 1$ 則 $\hat{c}_1 = \hat{c}_{\{1,2,\dots,S\}}^{RC}(1 - \alpha, k, \hat{P}_T)$ 。若 $h > 1$ ，則 \hat{c}_h 為式(1)。

Hsu et al. (2014) 的標準化 Step-SPA(k) 法的執行步驟與 Romano and Wolf (2007) 的標準化 Step-RC(k) 的執行步驟類似。唯一的差別在於，當利用自我重抽樣法估計拒絕標準時，將檢定統計量的中心位置改變，以得到較正確的分布，從而可以增加檢定效力。明確而言，標準化 Step-SPA(k) 將上述的臨界值步驟二中的 $kmax_K^{b,RC}$ 取代為

$$kmax_K^{b,SPA} = kmax_{j \in K} \left(\frac{\sqrt{T}(w_{r_j}^b - \hat{\theta}_{r_j} \mathbf{1}_A)}{\hat{\sigma}_{r_j}^b} \right),$$

其中 $A := \left\{ \sqrt{T}\hat{\theta}_{r_j} \geq -\sqrt{2\hat{\sigma}_{r_j}^2 \log \log T} \right\}$ 。 $\mathbf{1}_A$ 為指標函數：若 A 事件成立，則 $\mathbf{1}_A = 1$ ；反之為 0。定義 $\hat{d}_K(1 - \alpha, k, \hat{P}_T)$ 為上述 $\{kmax_K^{b,SPA}\}_{b=1}^B$ 的 $1 - \alpha$ 分位數。則臨界值步驟二中的 $\hat{c}_K^{RC}(1 - \alpha, k, \hat{P}_T)$ 改為

$$\hat{c}_K^{SPA}(1 - \alpha, k, \hat{P}_T) = \max(\hat{d}_K(1 - \alpha, k, \hat{P}_T), 0).$$

其餘檢定執行步驟和臨界值步驟則與 Romano and Wolf (2007) 的標準化 Step-RC(k) 的執行步驟相同，只是將 $\hat{c}_{\{1,2,\dots,S\}}^{RC}$ 和 \hat{c}_K^{RC} 分別替換為 $\hat{c}_{\{1,2,\dots,S\}}^{SPA}$ 和 \hat{c}_K^{SPA} 。

Hsu et al. (2014) 證明標準化 Step-SPA(k) 檢定法具有以下性質：(一) 大樣本下，標準化 Step-SPA(k) 法可以控制 $FWER(k)$ 小於預設的顯著水準 α 之下；(二) 若虛無假設為假，樣本數趨近於無限大時，標準化 Step-SPA(k) 拒絕此假設的機率趨近於 1；(三) 在一些正規條件之下，標準化 Step-SPA(k) 的檢定力在任何一種 Romano and Wolf (2005) 所定義的檢定力之下，均高於 Romano and Wolf (2007) 的標準化 Step-RC(k) 檢定。這些性質的討論與證明請參見 Hsu et al. (2014)。

二、基金績效指標

我們的實證研究中有四類基金績效指標，分別為基金平均報酬與大盤平均報酬的差異，基金夏普值與大盤夏普值差異，資訊比率，與多因子模型下基金的異常報酬。令 $r_{t,j}$ 為基金 j 的第 t 週週報酬， $r_{t,\phi}$ 為比較基準的大盤指數週報酬，無風險利率為 $r_{t,f}$ 。令第 j 檔基金，大盤報酬與無風險利率之母體均數為 $\mu_j := E(r_{t,j})$ ， $\mu_\phi := E(r_{t,\phi})$ 與 $\mu_f := E(r_{t,f})$ ；對應的樣本平均數為 $\bar{r}_j := T^{-1} \sum_{t=1}^T r_{t,j}$ ， $\bar{r}_\phi := T^{-1} \sum_{t=1}^T r_{t,\phi}$ 與 $\bar{r}_f := T^{-1} \sum_{t=1}^T r_{t,f}$ 。而第 j 檔基金和大盤報酬之母體變異數為 $\text{Var}(r_{t,j})$ 和 $\text{Var}(r_{t,\phi})$ ；對應的樣本變異數為

$$s_j^2 := \frac{1}{T-1} \sum_{t=1}^T (r_{t,j} - \bar{r}_j)^2; \quad s_\phi^2 := \frac{1}{T-1} \sum_{t=1}^T (r_{t,\phi} - \bar{r}_\phi)^2.$$

第一個績效指標為基金報酬均數與大盤報酬均數的差，所要檢定的參數 θ_j 為

$$\theta_j = \mu_j - \mu_\phi.$$

其樣本估計式和檢定統計量為： $w_j = \hat{\theta}_j = \bar{r}_j - \bar{r}_\phi$ 。依據 Hansen (2005) 的建議，採用定態自我重抽樣法時，可利用下式估計 $\hat{\theta}_j$ 的變異數：

$$\hat{\sigma}_j^2 = \frac{1}{T} \left[\hat{\gamma}_{0,j} + 2 \sum_{i=1}^{T-1} \kappa(T, i) \hat{\gamma}_{i,j} \right],$$

其中 $\hat{\gamma}_{i,j} = T^{-1} \sum_{t=1}^{T-i} (d_{t,j} - \bar{d}_j)(d_{t+i,j} - \bar{d}_j)$ ，而 $d_{t,j} := r_{t,j} - r_{t,\phi}$ ， $\bar{d}_j := T^{-1} \sum_t d_{t,j}$ ，

$$\kappa(T, i) = \frac{T-i}{T} (1-q)^i + \frac{i}{T} (1-q)^{T-i},$$

q 為定態自我重抽樣法下幾何分布的參數。

第二個績效指標為基金夏普值與大盤夏普值差異，檢定參數為基金與大盤的夏普值差距：

$$\theta_j = \frac{\mu_j - \mu_f}{\sqrt{\text{Var}(r_{t,j})}} - \frac{\mu_\phi - \mu_f}{\sqrt{\text{Var}(r_{t,\phi})}}.$$

θ_j 的估計式和檢定統計量 $w_j = \theta_j$ 可以利用對應的樣本平均數與樣本標準差來計算

$$\hat{\theta}_j = \frac{\bar{r}_j - \bar{r}_f}{s_j} - \frac{\bar{r}_\phi - \bar{r}_f}{s_\phi},$$

其中，為了計算標準化的檢定統計量，必須先得到夏普值差距 $\hat{\theta}_j$ 的變異數估計式 $\hat{\sigma}_j^2$ ；此估計式的詳細計算方式請見莊惠菁與管中閔 (2010) 文中的討論。

第三個績效指標為資訊比率，檢定參數為：

$$\theta_j = \frac{\mu_j - \mu_\phi}{\sqrt{\text{Var}(r_{t,j} - r_{t,\phi})}}$$

θ_j 的估計式和檢定統計量 $w_j = \theta_j$ 可以利用對應的樣本平均數與樣本標準差來計算

$$\hat{\theta}_j = \frac{\bar{r}_j - \bar{r}_\phi}{s_{j-\phi}}$$

其中 $s_{j-\phi}$ 為基金報酬與大盤報酬差異的樣本標準差。即，定義 $r_{t,j-\phi} = r_{t,j} - r_{t,\phi}$ ，則

$$s_{j-\phi}^2 := \frac{1}{T-1} \sum_{t=1}^T (r_{t,j-\phi} - \bar{r}_{j-\phi})^2.$$

為了計算檢定統計量，類似於 Ledoit and Wolf (2008) 之方法，我們利用定態自我抽樣的方式來獲得資訊比率估計式 $\hat{\theta}_j$ 的變異數估計式 $\hat{\sigma}_j^2$ 。資訊比率的算法，是將個別基金平均報酬減去大盤平均報酬，再除以相減後差額的標準差。資訊比率越高，代表風險調整後，基金擊敗大盤之能力越強。(註6)

第四個績效指標為因子模型的異常報酬。定義基金和大盤的之超額報酬分別為 $r_{t,j}^e := r_{t,j} - r_{t,f}$ ，則對個別 j 檔基金而言，多因子模型為複迴歸模型：

$$r_{t,j}^e = \alpha_j + \beta_j' \mathbf{f}_t^{(p)} + \varepsilon_{t,j}; \quad t = 1, \dots, T.$$

註6：資訊比率與夏普值差距的定義非常接近。夏普值差距為先計算個別基金風險調整後報酬(即夏普值)再與大盤之夏普值相減。而資訊比率為基金與大盤平均報酬相減後，再調整其風險。

我們在實證研究中考慮一因子，三因子，與四子模型，並定義個別模型之因子報酬為：

$$\mathbf{f}_t^{(1)} = [r_{t,\phi}^e];$$

$$\mathbf{f}_t^{(3)} = [r_{t,\phi}^e; smb_t; hml_t];$$

$$\mathbf{f}_t^{(4)} = [r_{t,\phi}^e; smb_t; hml_t; mom_t].$$

其中， $r_{t,\phi}^e := r_{t,\phi} - r_{t,f}$ 為市場超額報酬； smb_t 為規模溢酬； hml_t 為淨值市價比溢酬；而 mom_t 為動能因子。欲檢定的參數 θ_j 為迴歸截距項 α_j 即因子模型的異常報酬。若 α_j 大於 0，表示基金報酬有高於因子報酬的異常報酬存在。檢定統計量和參數估計式為利用最小平方法估計的截距項， $w_j = \hat{\alpha}_j$ 。為了計算標準化的檢定統計量，我們採用 HAC (heteroscedasticity and autocorrelation consistent) 的變異數估計式來估計 $\hat{\sigma}_j^2$ ，以避免對誤差項做過多假設。在後續的實證中，我們的 HAC 估計式採用 Bartlett 核函數 (kernel function)，並依照 Newey and West (1994) 的算法自動選擇頻寬 (bandwidth)；這些選擇和許多文獻的實證作法相同。

參、實證結果

一、資料與基本分析

本文分析國內股票與平衡型共同基金，總共 273 檔基金的績效。我們的實證資料為股利調整後的基金淨值報酬，考慮資料頻率為週資料，因此，從 2011 年 5 月至 2017 年 8 月，每檔基金有 327 筆週資料；資料來源為臺灣經濟新報資料庫 (以下簡稱新報資料庫)。與基金績效比較的大盤基準指數有三種：台股加權指數，臺灣 50 指數，MSCI 臺灣指數。本文之無風險利率為可轉讓定期存單初級市場 1-90 天週息。我們將各大盤週報酬的基本敘述統計量綜合於表 1。以均數和夏普值而言，最低的是台股加權指數，最高的為臺灣

50 指數。以偏峰度係數而言，三種大盤指數皆為左偏且峰度係數皆大於 3。在 1% 的顯著水準之下，Jarque-Bera 常態檢定結果皆為拒絕常態分布的假設。Augmented Dickey-Fuller 檢定則拒絕有單根之假設。

我們亦將基金超越大盤的情況整理於表 2。在此表的上半部，“總數”代表在總樣本期間績效指標超越大盤的基金檔數，“百分比”為“總數”佔所有基金數的比例。舉例而言，總共有 175 檔 (64.10%) 基金的平均週報酬超越台股加權指數；分別有 143 檔和 148 檔基金 (52.38% 和 54.21%) 的夏普值超越臺灣 50 和 MSCI 臺灣指數；當利用 MSCI 指數為市場因子時，共有 186 檔基金 (68.13%) 的一因子異常報酬為正值，177 檔和 138 檔基金的三因子與四因子異常報酬為正值。綜合而言，若只以績效指標的數字大小來判斷，國內共同基金中很大一部分的績效會超越大盤指數。

由於上述的比較並未考慮統計顯著性，我們也對個別基金執行右尾的 t 檢定；檢定結果綜合於表 2 的下半部。當顯著水準為 5%，表中“顯著數”代表 t 檢定統計量顯著的基金數，“百分比”為“顯著數”佔總基金數的比例。由表中我們可以看到，若以 t 檢定的結果作為判準，在顯著水準為 5% 之下，顯著優於大盤的基金數大幅減少；例如，有 13.92% 的基金平均報酬統計顯著優於台股加權指數；只有 4.40% 和 5.13% 的基金報酬夏普值顯著超越臺灣 50 和 MSCI 臺灣指數；而以 MSCI 指數為市場因子時，只有 11.72% 的基金一因子異常報酬為顯著的正值，10.26% 與 4.76% 的基金三因子與四因子異常報酬顯著為正值。

值得注意的是，不論用那一個績效指標，能超越臺灣 50 和 MSCI 臺灣指數的基金檔數都較超越台股加權指數的基金檔數少很多；這顯示臺灣 50 和 MSCI 臺灣指數指數的績效相對優異，這結果也可以從表 1 的敘述統計量可知。因此，在我們實證的設計上，為了能夠得到較多的樣本內卓越基金來建構樣本外的資產組合，並簡化分析，我們在建構資產組合時，將只採取相較於台股加權指數卓越的基金來形成資產組合，然後再進行此資產組合與台股加權，臺灣 50 與 MSCI 臺灣指數的樣本外績效比較。

表 1 大盤週報酬的敘述統計量

| | 臺股加權 | 臺灣 50 | MSCI |
|------|--------|--------|--------|
| 均數 | 0.092 | 0.126 | 0.121 |
| 標準差 | 2.054 | 2.190 | 2.185 |
| 夏普值 | 0.045 | 0.057 | 0.055 |
| 最大值 | 7.016 | 5.600 | 5.731 |
| 最小值 | -9.151 | -9.357 | -8.967 |
| 偏態係數 | -0.471 | -0.327 | -0.369 |
| 峰度係數 | 4.791 | 4.205 | 4.061 |
| JB | 0.001 | 0.001 | 0.001 |
| ADF | 0.001 | 0.001 | 0.001 |
| 樣本數 | 327 | 327 | 327 |

註：表中 JB 值為 Jarque-Bera 常態檢定結果之 p 值。ADF 值為 augmented Dickey-Fuller 單根檢定結果之 p 值。本表樣本期間為 2011/05/06 至 2017/08/25。

表 2 基金績效指標超越大盤概況

| | 臺股加權 | | 臺灣 50 | | MSCI | |
|---------|------|--------|-------|--------|------|--------|
| | 總數 | 百分比 | 總數 | 百分比 | 總數 | 百分比 |
| 均數差 | 175 | 64.10% | 139 | 50.92% | 145 | 53.11% |
| 夏普值差 | 180 | 65.93% | 143 | 52.38% | 148 | 54.21% |
| 資訊比率 | 175 | 64.10% | 139 | 50.92% | 145 | 53.11% |
| 一因子異常報酬 | 210 | 76.92% | 185 | 67.77% | 186 | 68.13% |
| 三因子異常報酬 | 204 | 74.73% | 170 | 62.27% | 177 | 64.84% |
| 四因子異常報酬 | 164 | 60.07% | 130 | 47.62% | 138 | 50.55% |
| | 顯著數 | 百分比 | 顯著數 | 百分比 | 顯著數 | 百分比 |
| 均數差 | 38 | 13.92% | 11 | 4.03% | 15 | 5.49% |
| 夏普值差 | 28 | 10.26% | 12 | 4.40% | 14 | 5.13% |
| 資訊比率 | 34 | 12.45% | 13 | 4.76% | 15 | 5.49% |
| 一因子異常報酬 | 48 | 17.58% | 29 | 10.62% | 32 | 11.72% |
| 三因子異常報酬 | 40 | 14.65% | 26 | 9.52% | 28 | 10.26% |
| 四因子異常報酬 | 22 | 8.06% | 12 | 4.40% | 13 | 4.76% |

註：上表中總數代表均數與夏普值超越對應大盤數目，或資訊比率，因子異常報酬為正的基金數目；百分比為總數佔總基金數之比率。下表中，顯著數為顯著水準為 5%，個別基金執行右尾 t 檢定顯著的基金數目。本表樣本期間為 2011/05/06 至 2017/08/25。

二、無資料窺探偏誤的基金績效檢定

如同前言中所討論，如果只針對個別基金執行 t 檢定，則檢定結果可能會產生資料窺探的偏誤。本研究於是以標準化後的 Step-SPA(k) 檢定法來檢定所有基金的績效。我們以 2011/05/06 至 2015/05/08，總共四年的週報酬作為樣本內資料，計有 208 筆週報酬。在 Step-SPA(k) 檢定中，定態自我重抽樣的重製樣本數為 $B = 1,000$ ；為了避免定態自我重抽樣法中的參數 q 值選定的任意性，依據 Politis and White (2004) 與 Patton et al. (2009) 的方式，我們為每一檔基金報酬資料分別估計最適的 q 值，然後利用所有基金 q 值的中位數當作定態自我重抽樣時的參數 q 。

標準化的 Step-SPA(k) 之檢定結果列於表 3。表中“顯著數”為控制 FWER(k) 不超過 5% 之下，拒絕虛無假設的基金數目，“百分比”則為“顯著數”占基金總數的比例。除了 $k = 1$ 的結果外，我們亦列出 $k = 3, 8, 14$ 的結果（這些 k 值約相當於總基金數的 1%，3% 和 5%）。當 $k = 1$ 時，標準化 Step-SPA(k) 方法即為莊惠菁與管中閔 (2010) 文中所利採用之標準化 Step-SPA 檢定。從表 3 可看到，若只控制 FWER(1) 小於顯著水準 5%，我們最多僅能找出 1 檔基金顯著優於大盤表現；特別是以夏普值和異常報酬為指標時，我們甚至無法找出超越臺灣 50，MSCI 指數之基金。若放寬到 $k = 3$ ，我們能找到的基金數和 $k = 1$ 的結果相差無幾。

若考慮較高的 k 值，我們可以認定較多的卓越基金。當 $k = 8$ ，而且以均數為績效指標時，我們可以認定出 6 檔基金超越台股加權指數，但只有 1 檔基金超越臺灣 50 與 MSCI 臺灣指數。以夏普值為績效指標時，我們可以認定出 6 檔基金超越台股加權指數，1 檔基金超越 MSCI 臺灣指數，但沒有基金可以超越臺灣 50 指數。若以一因子異常報酬為績效指標，我們可以認定出 7 檔基金超越台股加權指數，3 檔基金超越 MSCI 臺灣指數，但只有 1 檔基金超越臺灣 50 指數。若再放寬到 $k = 14$ 時，我們可以認定的卓越基金數就更多。例如，以夏普值為績效指標時，我們可以認定出 6 檔基金超越台股加權指數，1 檔基金超越臺灣 50，MSCI 臺灣指數。若以一因子異常報酬為績

效指標，總共有 11 檔基金超越台股加權指數，有 3 檔基金超越臺灣 50 和 MSCI 臺灣指數。若以三因子異常報酬為績效指標，總共有 15 檔基金超越台股加權指數，有 4 檔基金超越臺灣 50 和 MSCI 臺灣指數。

表 3 基金績效 Step-SPA(k) 的檢定結果

| | 台股加權 | | 臺灣 50 | | MSCI | |
|-------------------------|------|-------|-------|-------|------|-------|
| | 顯著數 | 百分比 | 顯著數 | 百分比 | 顯著數 | 百分比 |
| Panel A: 均數差 | | | | | | |
| $k = 1$ | 1 | 0.37% | 1 | 0.37% | 1 | 0.37% |
| $k = 3$ | 2 | 0.73% | 1 | 0.37% | 1 | 0.37% |
| $k = 8$ | 6 | 2.20% | 1 | 0.37% | 1 | 0.37% |
| $k = 14$ | 10 | 3.66% | 1 | 0.37% | 2 | 0.73% |
| Panel B: 夏普值差 | | | | | | |
| $k = 1$ | 1 | 0.37% | 0 | 0.00% | 0 | 0.00% |
| $k = 3$ | 1 | 0.37% | 0 | 0.00% | 0 | 0.00% |
| $k = 8$ | 6 | 2.20% | 0 | 0.00% | 1 | 0.37% |
| $k = 14$ | 6 | 2.20% | 1 | 0.37% | 1 | 0.37% |
| Panel C: 資訊比率 | | | | | | |
| $k = 1$ | 1 | 0.37% | 0 | 0.00% | 1 | 0.37% |
| $k = 3$ | 2 | 0.73% | 1 | 0.37% | 1 | 0.37% |
| $k = 8$ | 5 | 1.83% | 1 | 0.37% | 2 | 0.73% |
| $k = 14$ | 11 | 4.03% | 2 | 0.73% | 2 | 0.73% |
| Panel D: 一因子異常報酬 | | | | | | |
| $k = 1$ | 1 | 0.37% | 0 | 0.00% | 0 | 0.00% |
| $k = 3$ | 1 | 0.37% | 1 | 0.37% | 1 | 0.37% |
| $k = 8$ | 7 | 2.56% | 1 | 0.37% | 3 | 1.10% |
| $k = 14$ | 11 | 4.03% | 3 | 1.10% | 3 | 1.10% |
| Panel E: 三因子異常報酬 | | | | | | |
| $k = 1$ | 1 | 0.37% | 0 | 0.00% | 0 | 0.00% |
| $k = 3$ | 2 | 0.73% | 1 | 0.37% | 1 | 0.37% |
| $k = 8$ | 9 | 3.30% | 2 | 0.73% | 3 | 1.10% |
| $k = 14$ | 15 | 5.49% | 4 | 1.47% | 4 | 1.47% |
| Panel F: 四因子異常報酬 | | | | | | |
| $k = 1$ | 1 | 0.37% | 0 | 0.00% | 0 | 0.00% |
| $k = 3$ | 1 | 0.37% | 0 | 0.00% | 1 | 0.37% |
| $k = 8$ | 2 | 0.73% | 1 | 0.37% | 1 | 0.37% |
| $k = 14$ | 8 | 2.93% | 1 | 0.37% | 1 | 0.37% |

註：表中顯著數為利用標準化之 Step-SPA(k) 檢定法，顯著水準為 5%，利用不同績效指標，在不同的錯誤拒絕數 k 的設定，顯著超越大盤的基金數目。百分比為顯著基金數佔總基金數之比率。本表樣本期間：2011/05/06 至 2015/05/08。

以上的分析並未考慮交易手續等交易成本。為考慮交易成本，我們利用新報基金資料庫中基金費用及債券型基金收入模組的資料進行分析。因為新報基金淨值已經是扣除各種相關費用後的總金額，除以當日的總受益單位數的結果，因此，我們僅考慮手續費與交易稅這二項直接成本對報酬的影響並假定在樣本內四年間僅進出一次。我們考慮表 3 中，不論比較績效指標與比較大盤指數為何，所有（未考慮交易成本下）卓越的基金。在剔除重複的基金後，上述績效卓越的基金，共計有 19 檔基金，扣除直接成本後，最高的樣本內報酬為 120.83%，最低的為 18.19%，皆遠大於同時期台股加權，臺灣 50 與 MSCI 臺灣的報酬，分別為 7.59%，13.32% 和 11.74%。

前述結果顯示，我們比較容易認定出顯著超越台股加權指數的基金，但較難找到超越臺灣 50 或是 MSCI 臺灣指數的基金。從基金的績效標準來看，對於特定的大盤指數，若以夏普值或是四因子異常報酬為標準，我們能認定出來的卓越基金數較少，而採用其他標準則能認定出相對較多的卓越基金。值得一提的是，雖然放寬 k 值可能會使我們認定出其實並未超越大盤的基金（錯誤拒絕增加），但根據我們過去（未發表的）模擬結果，以及 Hsu et al. (2014) 的模擬結果，這些被錯誤認定的基金通常是相對接近大盤表現的基金，而不是表現很差的基金。換言之，實務上放寬 k 所付的代價並不大，且可以增加選到好基金的可能性。

三、基金績效的持續性

根據前面認定的基金，我們在本小節內進一步評估其績效是否具有持續性。由於比較的基準為台股加權指數時，樣本內所能找到的卓越基金數最多，故我們利用這些樣本內卓越顯著優於台股加權的基金，來建構平均加權資產組合，並且計算此資產組合在樣本外期間，相對應於台股加權，臺灣 50 與 MSCI 臺灣指數下的各種績效指標值。我們的樣本內認定卓越基金的期間為 2011/05/06 至 2015/05/08 (4 年)，而樣本外計算績效的期間為 2015/05/15 至 2017/8/25，共計 119 筆週資料 (2.28 年)。

定義 $r_{ew,T+m}^{\theta}$ 為利用 θ 為績效指標，在時間點 $T + m$ 時，所認定的樣

本內顯著優於台股加權指數的基金所組成的平均加權資產組合報酬，亦即

$$r_{ew,T+m}^{\theta} = \frac{1}{N_{\theta}} \sum_j r_{j,T+m}^{\theta},$$

其中 $r_{ew,T+m}^{\theta}$ 為 j 卓越基金的單期報酬，即新報資料庫中之週報酬， N_{θ} 為依據 θ 指標所選出的卓越基金的個數。舉例而言，若我們樣本內考慮利用夏普值當作績效標準，而錯誤拒絕程度設定為 $k = 8$ 時，依據表 3，樣本內認定有 6 檔基金優於台股加權報酬，我們便利用此 6 檔基金建構平均加權資產組合，並計算此資產組合樣本外報酬。

表 4 為資產組合在相對應市場報酬為台股加權指數時，樣本外的績效指標值。舉例而言，表中第二列的數字，為在樣本內利用均數差異當作績效指標，在 $k = 3$ 之下，所認定的 2 檔超越台股加權的基金所組成的平均加權資產組合，高於台股加權平均週報酬與夏普值分別為 12.55% 與 6.03%；資訊比率為 13.24%；一因子，三因子，四因子異常報酬分別為 13.50%，12.34% 與 7.86%。表最後一列的數字為在樣本內利用四因子異常報酬當作績效指標，在 $k = 14$ 之下，所認定的 8 檔超越台股加權的基金所組成的平均加權資產組合，其樣本外期間高於台股加權的平均週報酬與夏普值分別為 4.77% 與 4.11%；資訊比率為 5.54%；一因子，三因子，四因子異常報酬分別為 7.48% 6.87% 與 4.18%。表 4 中，我們可看到，除了在 $k = 8$ ，樣本內利用一因子所建構的資產組合，其四因子異常報酬為 -0.19% 外，其餘的績效指標皆為正值。這顯示利用標準化的 Step-SPA(k) 檢定來認定的基金，其資產組合的樣本外表現，不論以哪一種績效指標來衡量，皆較台股加權指數為佳。

表 5 為資產組合在相對應市場報酬為臺灣 50 指數時，樣本外的績效指標值，資產組合的建構方式與表 4 完全相同。由表 5 可發現，若樣本內利用均數差異，夏普值差異，資訊比率或是四因子報酬來決定卓越基金，其資產組合的樣本外績效表現較臺灣 50 來的好，但若是樣本內建構資產組合時的衡量指標為一因子時，則在 $k = 8$ ，會出現樣本外低於臺灣 50 指數的平均報酬與夏普值，負的資訊比率與四因子異常報酬。在 $k = 14$ 時，會出現低於臺灣 50 指數的平均報酬與負的資訊比率。而在樣本內衡量的指標為三因子

時，則在 $k = 8$ 時，會出現低於臺灣 50 指數的平均報酬與負的資訊比率。

表 4 卓越基金資產組合的樣本外績效與臺股加權比較值

| | 均數差 | 夏普值差 | 資訊比率 | 一因子異常 報酬 | 三因子異常 報酬 | 四因子異常 報酬 |
|-------------------------|--------|-------|--------|-------------|-------------|-------------|
| Panel A: 均數差 | | | | | | |
| $k = 1$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 3$ | 12.55% | 6.03% | 13.24% | 13.50% | 12.34% | 7.86% |
| $k = 8$ | 12.47% | 6.75% | 14.65% | 13.87% | 12.90% | 8.48% |
| $k = 14$ | 11.86% | 6.21% | 13.89% | 13.07% | 12.00% | 7.36% |
| Panel B: 夏普值差 | | | | | | |
| $k = 1$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 3$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 8$ | 10.94% | 6.32% | 13.43% | 12.55% | 11.68% | 7.28% |
| $k = 14$ | 10.94% | 6.32% | 13.43% | 12.55% | 11.68% | 7.28% |
| Panel C: 資訊比率 | | | | | | |
| $k = 1$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 3$ | 12.55% | 6.03% | 13.24% | 13.50% | 12.34% | 7.86% |
| $k = 8$ | 11.40% | 6.27% | 13.25% | 12.88% | 11.97% | 7.56% |
| $k = 14$ | 12.81% | 6.58% | 14.52% | 13.98% | 12.86% | 8.08% |
| Panel D: 一因子異常報酬 | | | | | | |
| $k = 1$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 3$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 8$ | 1.10% | 1.37% | 0.97% | 3.97% | 3.29% | -0.19% |
| $k = 14$ | 2.32% | 2.51% | 2.17% | 5.45% | 4.75% | 1.35% |
| Panel E: 三因子異常報酬 | | | | | | |
| $k = 1$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 3$ | 12.55% | 6.03% | 13.24% | 13.50% | 12.34% | 7.86% |
| $k = 8$ | 3.79% | 2.95% | 3.72% | 6.39% | 5.65% | 1.96% |
| $k = 14$ | 4.01% | 3.42% | 3.70% | 7.11% | 6.29% | 2.99% |
| Panel F: 四因子異常報酬 | | | | | | |
| $k = 1$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 3$ | 10.98% | 5.95% | 14.82% | 12.12% | 11.36% | 9.39% |
| $k = 8$ | 8.39% | 5.28% | 11.65% | 10.10% | 9.48% | 6.61% |
| $k = 14$ | 4.77% | 4.11% | 5.54% | 7.48% | 6.87% | 4.18% |

註：本表為利用標準化 Step-SPA(k) 檢定，顯著水準 5% 之下，樣本內認定超越臺股加權指數之基金的樣本外資產組合的績效表現。績效值為相對應於市場報酬為臺股加權指數，所計算出來的數值。舉例而言，表第二列的數字，為在樣本內利用均數差異當作績效指標，在 $k = 3$ 之下，所認定的 2 檔超越臺股加權的基金所形成的平均加權資產組合，高於臺股加權平均週報酬與夏普值分別為 12.55% 與 6.03%，資訊比率為 13.24%。一因子，三因子，四因子異常報酬分別為 13.50%，12.34% 與 7.86%。本表樣本內認定卓越基金的期間為 2011/05/06 至 2015/05/08。樣本外的績效指標計算期間為 2015/05/15 至 2017/8/25。

表 5 卓越基金資產組合的樣本外績效與臺灣 50 比較值

| | 均數差 | 夏普值差 | 資訊比率 | 一因子異常報酬 | 三因子異常報酬 | 四因子異常報酬 |
|------------------|--------|--------|--------|---------|---------|---------|
| Panel A: 均數差 | | | | | | |
| $k = 1$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 3$ | 8.73% | 4.62% | 7.52% | 11.48% | 9.63% | 6.91% |
| $k = 8$ | 8.65% | 5.34% | 8.07% | 11.90% | 10.31% | 7.53% |
| $k = 14$ | 8.04% | 4.80% | 7.46% | 11.08% | 9.34% | 6.37% |
| Panel B: 夏普值差 | | | | | | |
| $k = 1$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 3$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 8$ | 7.12% | 4.91% | 6.86% | 10.61% | 9.15% | 6.33% |
| $k = 14$ | 7.12% | 4.91% | 6.86% | 10.61% | 9.15% | 6.33% |
| Panel C: 資訊比率 | | | | | | |
| $k = 1$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 3$ | 8.73% | 4.62% | 7.52% | 11.48% | 9.63% | 6.91% |
| $k = 8$ | 7.58% | 4.86% | 7.00% | 10.94% | 9.40% | 6.61% |
| $k = 14$ | 8.99% | 5.17% | 8.12% | 12.01% | 10.18% | 7.10% |
| Panel D: 一因子異常報酬 | | | | | | |
| $k = 1$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 3$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 8$ | -2.72% | -0.04% | -2.06% | 2.33% | 1.09% | -1.05% |
| $k = 14$ | -1.50% | 1.10% | -1.19% | 3.84% | 2.63% | 0.54% |
| Panel E: 三因子異常報酬 | | | | | | |
| $k = 1$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 3$ | 8.73% | 4.62% | 7.52% | 11.48% | 9.63% | 6.91% |
| $k = 8$ | -0.03% | 1.54% | -0.02% | 4.66% | 3.38% | 1.08% |
| $k = 14$ | 0.19% | 2.01% | 0.15% | 5.59% | 4.14% | 2.17% |
| Panel F: 四因子異常報酬 | | | | | | |
| $k = 1$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 3$ | 7.15% | 4.54% | 7.99% | 9.85% | 8.79% | 8.41% |
| $k = 8$ | 4.57% | 3.87% | 5.00% | 8.02% | 7.01% | 5.62% |
| $k = 14$ | 0.95% | 2.70% | 0.87% | 5.79% | 4.67% | 3.37% |

註：本表為利用標準化 Step-SPA(k) 檢定法，顯著水準 5% 之下，樣本內認定超越台股加權指數之基金的樣本外資產組合的績效表現。績效值為相對應於市場報酬為臺灣 50 指數，所計算出來的數值。舉例而言，表第二列的數字，為在樣本內利用均數差異當作績效指標，在 $k=3$ 之下，所認定的 2 檔超越台股加權的基金所形成的平均加權資產組合，高於臺灣 50 指數平均週報酬與夏普值分別為 8.73% 與 4.62%，資訊比率為 7.52%。一因子，三因子，四因子異常報酬分別為 11.48%，9.63% 與 6.91%。本表樣本內認定卓越基金的期間為 2011/05/06 至 2015/05/08。樣本外的績效指標計算期間為 2015/05/15 至 2017/8/25。

表 6 資產組合在相對應市場報酬為 MSCI 臺灣指數時，樣本外的績效指標值。由表中可發現，與表 5 相同，若樣本內利用均數差異，夏普值差異，資訊比率或是四因子報酬來決定卓越基金，其資產組合的樣本外績效表現較 MSCI 臺灣指數來的好，但若是樣本內建構資產組合時的衡量指標為一因子時，則在 $k =$

8，會出現低於 MSCI 臺灣指數的平均報酬，夏普值，負的資訊比率與四因子異常報酬。在 $k = 14$ 時，會出現低於 MSCI 臺灣指數的平均報酬與負的資訊比率。而樣本內衡量的指標為三因子時，則在 $k = 8$ 或 $k = 14$ 時，會出現低於 MSCI 臺灣指數的平均報酬與負的資訊比率。綜觀表 5 與表 6，我們可以發現當選取的樣本內評估績效為一因子或三因子時，其樣本外的績效表現相對較差。

表 6 卓越基金資產組合的樣本外績效與 MSCI 臺灣比較值

| | 均數差 | 夏普值差 | 資訊比率 | 一因子異常報酬 | 三因子異常報酬 | 四因子異常報酬 |
|------------------|--------|--------|--------|---------|---------|---------|
| Panel A: 均數差 | | | | | | |
| $k = 1$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 3$ | 8.08% | 4.42% | 7.15% | 10.98% | 9.36% | 6.83% |
| $k = 8$ | 8.00% | 5.14% | 7.63% | 11.44% | 10.06% | 7.47% |
| $k = 14$ | 7.39% | 4.60% | 7.04% | 10.60% | 9.08% | 6.30% |
| Panel B: 夏普值差 | | | | | | |
| $k = 1$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 3$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 8$ | 6.47% | 4.71% | 6.35% | 10.16% | 8.91% | 6.29% |
| $k = 14$ | 6.47% | 4.71% | 6.35% | 10.16% | 8.91% | 6.29% |
| Panel C: 資訊比率 | | | | | | |
| $k = 1$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 3$ | 8.08% | 4.42% | 7.15% | 10.98% | 9.36% | 6.83% |
| $k = 8$ | 6.93% | 4.66% | 6.53% | 10.48% | 9.16% | 6.56% |
| $k = 14$ | 8.34% | 4.97% | 7.74% | 11.52% | 9.92% | 7.04% |
| Panel D: 一因子異常報酬 | | | | | | |
| $k = 1$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 3$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 8$ | -3.38% | -0.24% | -2.59% | 1.90% | 0.85% | -1.17% |
| $k = 14$ | -2.15% | 0.90% | -1.73% | 3.44% | 2.40% | 0.42% |
| Panel E: 三因子異常報酬 | | | | | | |
| $k = 1$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 3$ | 8.08% | 4.42% | 7.15% | 10.98% | 9.36% | 6.83% |
| $k = 8$ | -0.68% | 1.34% | -0.57% | 4.23% | 3.14% | 0.97% |
| $k = 14$ | -0.46% | 1.81% | -0.36% | 5.16% | 3.90% | 2.06% |
| Panel F: 四因子異常報酬 | | | | | | |
| $k = 1$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 3$ | 6.50% | 4.34% | 7.16% | 9.49% | 8.59% | 8.45% |
| $k = 8$ | 3.92% | 3.67% | 4.27% | 7.65% | 6.81% | 5.64% |
| $k = 14$ | 0.30% | 2.50% | 0.27% | 5.40% | 4.46% | 3.31% |

註：本表為利用標準化 Step-SPA(k) 檢定法，顯著水準 5% 之下，樣本內認定超越台股加權指數之基金的樣本外資產組合的績效表現。績效值為相對應於市場報酬為 MSCI 臺灣指數，所計算出來的數值。舉例而言，表第二列的數字，為在樣本內利用均數差異當作績效指標，在 $k = 3$ 之下，所認定的 2 檔超越台股加權的基金所形成的平均加權資產組合，高於 MSCI 臺灣指數平均週報酬與夏普值分別為 8.08% 與 4.42%，資訊比率為 7.15%。一因子，三因子，四因子異常報酬分別為 10.98%，9.36% 與 6.83%。本表樣本內認定卓越基金的期間為 2011/05/06 至 2015/05/08。樣本外的績效指標計算期間為 2015/05/15 至 2017/8/25。

除了資產組合績效指標值的樣本外表現外，我們亦關心資產組合的樣本外持有期間報酬。定義從 T 期持有至 $T+J$ 期的持有期間報酬 (J -period gross return) 為：

$$r_T^\theta(J) + 1 := \prod_{m=1}^J (1 + r_{ew, T+m}^\theta).$$

令 T 為樣本內最後一天 (2015/05/08)，基金持有期間為 1 至 119 週 ($J = 1, 2, \dots, 119$)。圖 2 為利用 Step-SPA(k) 檢定， $k = 14$ ，在樣本內比較的市場報酬為台股加權指數，依據不同績效指標認定的基金所形成的資產組合的樣本外持有期間報酬。我們將資產組合與台股加權指數樣本外的持有期間報酬相互比較。例如，標記為均數的持有期間報酬，其資產組成份為表 3 中所認定的 10 檔基金所建構的資產組合。標記為資訊比率的持有期間報酬，其資產組成份為表 3 中所認定的 11 檔基金所建構。由此圖我們可以發現，所有的樣本外時間點的持有期間報酬，皆大於台股加權指數的持有期間報酬 2017 年後，雖然資產組合的報酬仍大過台股加權，但幅度降低。例如以因子異常報酬建構的資產組合，在 2017 年六月的持有期間報酬和台股加權報酬幾乎一致。圖 3 和圖 4 的資產組合定義，與圖 2 的定義完全相同，只是我們將樣本外的比較基準，由台股加權指數變為臺灣 50 與 MSCI 台灣指數。由這二張圖中我們發現，持有期間報酬在 2017 年前，不論利用哪一個績效指標所形成的資產組合，都高於比較大盤樣本外持有期間報酬。但 2017 年後，利用因子異常報酬為績效指標選出的基金所組成的資產組合，其績效會低於臺灣 50 與 MSCI 臺灣指數。

綜合前述討論，我們認為本文所採用的 Step-SPA(k) 檢定具有實務操作上的價值。利用此一檢定方法在樣本內所認定出來的卓越基金，其投資組合的績效在樣本外都有相當程度的持續性，且多能超越大盤表現。需要注意的是，根據因子異常報酬所認定出來的基金，其投資組合的樣本外績效持續性會因持有期間的長短而有較大差異；但根據均數，夏普值或是資訊比率所認定的基金組成的資產組合，其樣本外績效持續性則沒有此一問題。

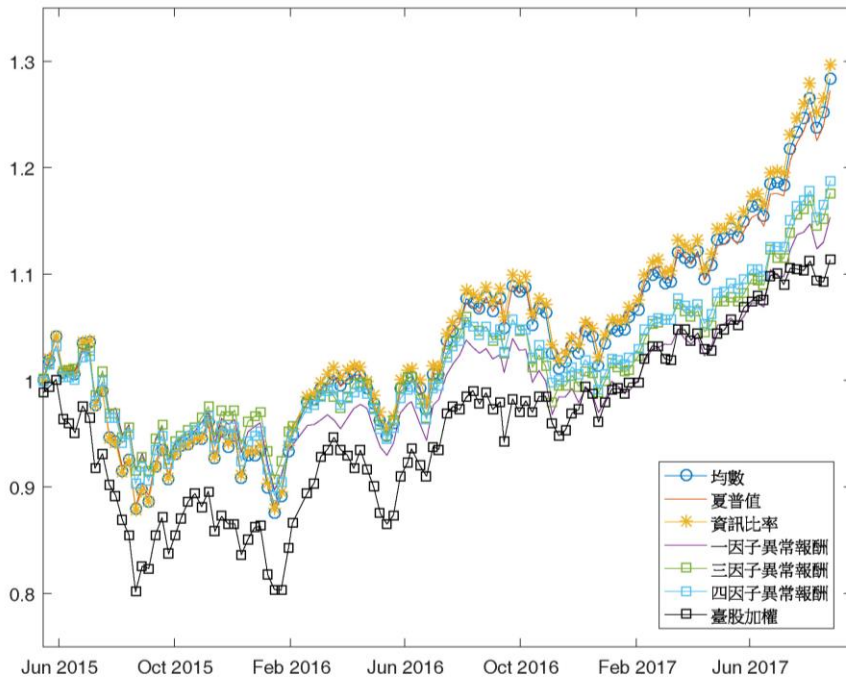


圖 2 資產組合與臺股加權之樣本外持有期間報酬

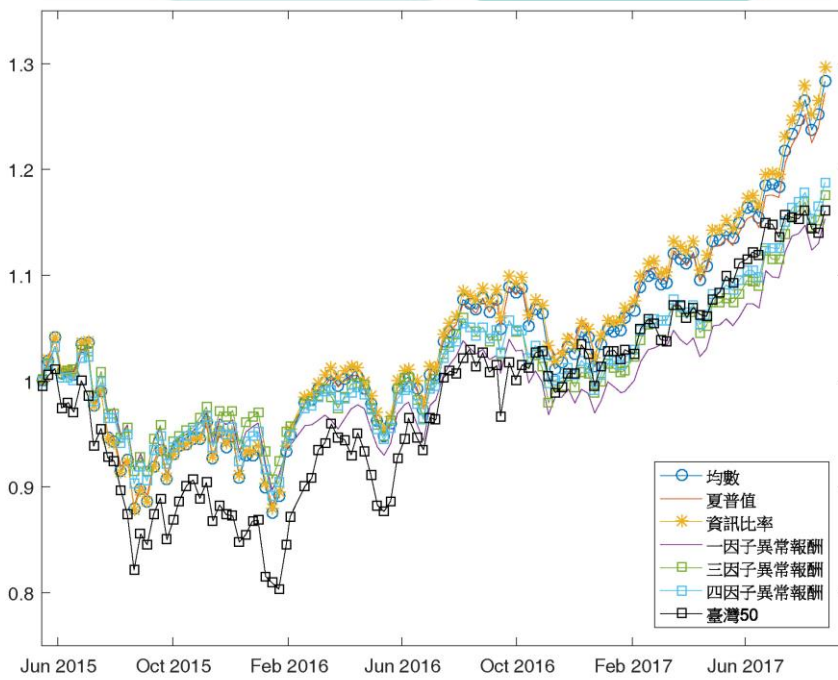


圖 3 資產組合與臺灣 50 指數之樣本外持有期間報酬

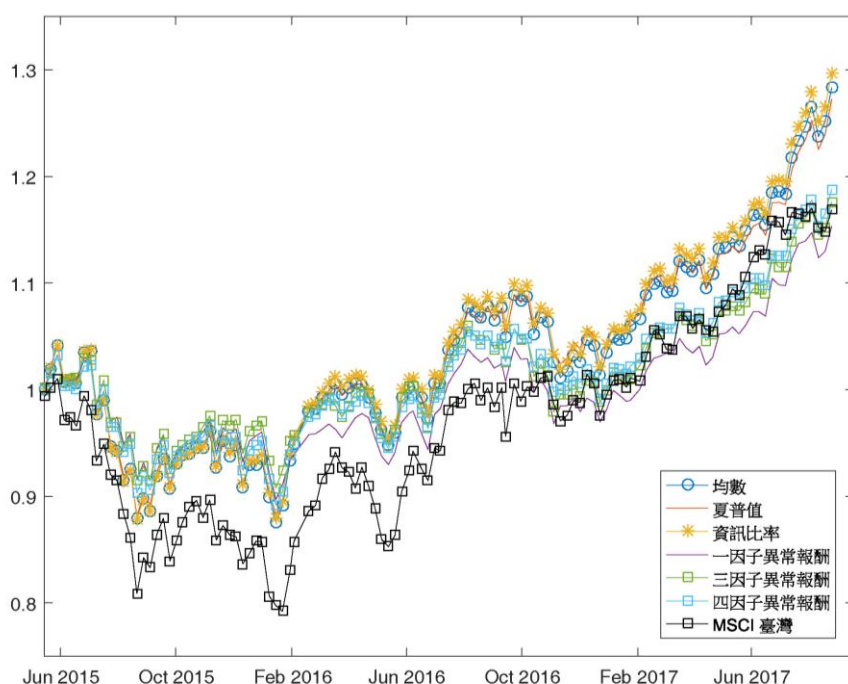


圖 4 資產組合與 MSCI 臺灣指數之樣本外持有期間報酬

肆、結論

本文以嚴謹的逐步檢定法來認定表現的卓越基金，並評估由這些卓越基金所建構之資產組合的樣本外績效。在方法上，我們採用了 Hsu et al. (2014) 的 Step-SPA(k) 檢定法來驗證共同基金的績效，其中 k 為我們願意容忍的錯誤拒絕次數。相較於莊惠菁與管中閔 (2010) 的檢定方法，本文採用的檢定放寬了錯誤容忍的程度，進而提高了能認定出更多卓越基金的可能性。我們對臺灣股票型與平衡型共同基金的實證研究發現，在 2011 年至 2017 年間，若利用 Step-SPA(8) 或 Step-SPA(14) 來檢定這些基金績效，我們可以找到相當數量的，超越大盤指數的卓越基金；這些卓越基金所架構的資產組合，其樣本外持有期間報酬亦多能超越比較基準的大盤指數。本文結果顯示，Step-SPA(k) 檢定可被視為認定卓越基金的有效方法，而這些基金組成的資產組合大多具有優異的樣本外表現。因此，Step-SPA(k) 檢定很值得業者在實務操作

(特別是建構基金組合) 時參考採用。

參考文獻

- 莊惠菁、管中閔 (2010), 「以無資料窺探偏誤的檢定評估共同基金績效」, 《證券市場發展季刊》, 第 22 卷第 3 期, 181-206。
- Bajgrowicz, P., Scaillet, O. (2012), “Technical trading revisited: False discoveries, persistence tests, and transaction costs.” *Journal of Financial Economics*, 106(3), 473-491.
- Bao, Y., Lee, T.-H., Saltoglu, B. (2006), “Evaluating predictive performance of Value-at-Risk models in emerging markets: A reality check.” *Journal of Forecasting*, 25(2), 101-128.
- Barras, L., Scaillet, O., Wermers, R. (2010), “False discoveries in mutual fund performance: Measuring luck in estimated alphas.” *The Journal of Finance*, 65(1), 179-216.
- Benjamini, Y., Hochberg, Y. (1995), “Controlling the false discovery rate: A practical and powerful approach to multiple testing.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300.
- Benjamini, Y., Yekutieli, D. (2001), “The control of the false discovery rate in multiple testing under dependency.” *The Annals of Statistics*, 29(4), 1165-1188.
- Chen, X., Ludvigson, S. C. (2009), “Land of addicts? An empirical investigation of habit-based asset pricing models.” *Journal of Applied Econometrics*, 24(7), 1057-1093.
- Genre, V., Kenny, G., Meyler, A., Timmermann, A. (2013), “Combining expert forecasts: Can anything beat the simple average?” *International Journal of Forecasting*, 29(1), 108-121.
- Goyal, A., Wahal, S. (2015), “Is momentum an echo?” *Journal of Financial and*

Quantitative Analysis, 50(6), 1237-1267.

Hanck, C. (2009), "For which countries did PPP hold?" *A multiple testing approach. Empirical Economics*, 37(1), 93-103.

Hansen, P. R. (2005), "A test for superior predictive ability." *Journal of Business & Economic Statistics*, 23(4), 365-380.

Hansen, P. R., Lunde, A. (2005), "A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)?" *Journal of Applied Econometrics*, 20(7), 873-889.

Harvey, C. R., Liu, Y. (2014), "Evaluating trading strategies." *The Journal of Portfolio Management*, 40(5), 108-118.

Harvey, C. R., Liu, Y., Zhu, H. (2016), "... and the cross-section of expected returns." *Review of Financial Studies*, 29(1), 5-68.

Holm, S. (1979), "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics*, 6(2), 65-70.

Hommel, G., Hoffmann, T. (1988), *Controlled Uncertainty*. Springer.

Hsu, P.-H., Hsu, Y.-C., Kuan, C.-M. (2010), "Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias." *Journal of Empirical Finance*, 17(3), 471-484.

Hsu, P.-H., Kuan C.-M. (2005), "Reexamining the profitability of technical analysis with data snooping checks." *Journal of Financial Econometrics*, 3(4), 606-628.

Hsu, P.-H., Taylor, M.P., Wang, Z. (2016), "Technical trading: Is it still beating the foreign exchange market?" *Journal of International Economics*, 102, 188-208.

Hsu, Y.-C., Kuan, C.-M., Yen, M.-F. (2014), "A generalized stepwise procedure with improved power for multiple inequalities testing." *Journal of Financial Econometrics*, 12(4), 730-755.

Kao, Y.-C., Kuan, C.-M., Chen, S. (2013), "Testing the predictive power of the

- term structure without data snooping bias.” *Economics Letters*, 121(3), 546-549.
- Koopman, S.J., Jungbacker, B., Hol, E. (2005), “Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements.” *Journal of Empirical Finance*, 12(3), 445-475.
- Kuang, P., Schröder, M., Wang, Q. (2014), “Illusory profitability of technical analysis in emerging foreign exchange markets.” *International Journal of Forecasting*, 30(2), 192-205.
- Lahiri, S. N. (2013), *Resampling Methods for Dependent Data*. Springer Science & Business Media.
- Ledoit, O., Wolf, M. (2008), “Robust performance hypothesis testing with the Sharpe ratio.” *Journal of Empirical Finance*, 15(5), 850-859.
- Lehmann, E. L., Romano, J. P. (2005), “Generalizations of the familywise error rate.” *The Annals of Statistics*, 33(3), 1138-1154.
- Liu, L. Y., Patton, A. J., Sheppard, K. (2015), “Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes.” *Journal of Econometrics*, 187(1), 293-311.
- Martin, G., Reidy, A., Wright, J. (2009), “Does the option market produce superior forecasts of noise corrected volatility measures?” *Journal of Applied Econometrics*, 24(1), 77-104.
- Molodtsova, T., Papell, D. H. (2009), “Out-of-sample exchange rate predictability with Taylor rule fundamentals.” *Journal of International Economics*, 77(2), 167-180.
- Neuhierl, A., Schlusche, B. (2011), “Data snooping and market-timing rule performance.” *Journal of Financial Econometrics*, 9(3), 550-587.
- Newey, W. K., West, K. D. (1994), “Automatic lag selection in covariance matrix estimation.” *The Review of Economic Studies*, 61(4), 631-653.
- Patton, A. J. (2004), “On the out-of-sample importance of skewness and asymmetric

- dependence for asset allocation.” *Journal of Financial Econometrics*, 2(1), 130-168.
- Patton, A. J. (2011), “Data-based ranking of realised volatility estimators.” *Journal of Econometrics*, 161(2), 284-303.
- Patton, A. J., Politis, D. N., White, H. (2009), “Correction to “Automatic block-length selection for the dependent bootstrap” by D. Politis and H. White.” *Econometric Reviews*, 28(4), 372-375.
- Patton, A. J., Timmermann, A. (2010), “Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts.” *Journal of Financial Economics*, 98(3), 605-625.
- Politis, D.N., Romano, J.P. (1994), “The stationary bootstrap.” *Journal of the American Statistical Association*, 89(428), 1303-1313.
- Politis, D. N., White, H. (2004), “Automatic block-length selection for the dependent bootstrap.” *Econometric Reviews*, 23(1), 53-70.
- Qi, M., Wu, Y. (2006), “Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market.” *Journal of Money, Credit, and Banking*, 38(8), 2135-2158.
- Romano, J. P., Shaikh, A. M., Wolf, M. (2008), “Formalized data snooping based on generalized error rates.” *Econometric Theory*, 24(2), 404-447.
- Romano, J. P., Wolf, M. (2005), “Stepwise multiple testing as formalized data snooping.” *Econometrica*, 73(4), 1237-1282.
- Romano, J.P., Wolf, M. (2007), “Control of generalized error rates in multiple testing.” *The Annals of Statistics*, 35(4), 1378-1408.
- Shynkevich, A. (2012), “Performance of technical analysis in growth and small cap segments of the US equity market.” *Journal of Banking & Finance*, 36(1), 193-208.
- White, H. (2000), “A reality check for data snooping.” *Econometrica*, 6(5), 1097-1126.

Identifying and Assessing Superior Mutual Funds: An Application of the New Stepwise Data-Snooping-Bias Free Test

Hui-Ching Chuang*

College of Management, Yuan Ze University

Chung-Ming Kuan

Department of Finance, National Taiwan University

This article adopts a hypothesis testing method that avoids the data snooping bias to identify the outstanding funds and evaluate the out-of-sample performance of the - portfolio constructed by these outstanding funds. Compared with the empirical methods of Chuang and Kuan (2010), the testing method used in this paper relaxes the “familywise error rate” concept and therefore, can correctly identify more funds that outperform broader market indices. Empirical results show that from 2011 to 2017, regardless of the use of average return, Sharp ratios, information ratios, or multi-factor abnormal returns as performance indicators, we can identify a considerable amount of Taiwan equity and balanced mutual funds, which are significantly better than the return of the Taiwan capitalization-weighted stock index (TAIEX). Also, the portfolio constructed by these superior funds remains outperforms the TAIEX, Taiwan 50, and MSCI Taiwan index in the out-of-sample period.

Key Words : Data Snooping Bias, Familywise Error Rate, Mutual Fund, Stepwise Reality Check, Stepwise Superior Predictive Ability Test.

* Correspondence: Hui-Ching Chuang. Address: No. 135 Yuan-Tung Road, Tao-Yuan, 320, Taiwan. Tel: 886-34638800. The financial support from the Ministry of Science and Technology, Taiwan, is gratefully acknowledged (MOST 107-2632-H-155-001 for H.-C. Chuang, and MOST 106-2410-H-002-014-MY2 for C.-M. Kuan). Email: hcchuang@saturn.yzu.edu.tw (H.-C. Chuang), ckuan@ntu.edu.tw (C.-M. Kuan).

莊惠菁 管中閔

