

多變量常態的多重檢定實證研究

汪群超[†]

國立臺北大學統計學系

摘 要

由於多變量常態檢定並不存在一個具全方位檢定力 (uniformly powerful) 的統計量，因此當各種具檢定力的檢定方法做出不同判決時，不知道該相信哪一個？Tenreiro (2011) 根據 Fromont and Laurent (2006) 所提出改良型的 Bonferroni 校正法，建立一套檢定多變量常態的多重檢定程序。該程序包含了四個在面對不同特性的對立假說時，各具高檢定力的統計量，試圖藉群組力量，截長補短，使之具備全方位檢定力。而其蒙地卡羅實驗也驗證了這個想法。本文藉由這個多重比較的概念，並根據 Bonferroni 校正法的原理，認為較少數量的組合應該能提供較高檢定力，只要組成群組的統計量也具備對廣泛的對立假說的檢定力。由於本文作者曾提出一個深具競爭力的統計量 $W_{min,m}(5)$ (Wang and Hwang, 2011)，認為可以取代 Tenreiro (2011) 的四個組合成員之兩個或甚至三個，成為新的多重檢定的組合。透過與 Tenreiro (2011) 同樣的實證研究，發現新的組合中，以三個成員組合的多重檢定表最好，不但整體表現優於 Tenreiro (2011) 的四個統計量的組合，也同樣展現全方位的檢定力。

關鍵詞：多重檢定程序、Bonferroni 校正法、多變量常態分配、蒙地卡羅模擬。

JEL classification: C15, C63, C65.

[†]通訊作者: 汪群超
E-mail: ccw@gm.ntpu.edu.tw

1. 緒論

多變量分析是統計學上重要的應用工具，有些多變量分析的方法要求資料必須服從多變量常態分配的假設，譬如，多變量變異數分析 (MANOVA)、多變量迴歸、判別分析 (discriminant analysis) 等。多變量常態的檢定問題描述為：

令 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 代表一 p 維度變量 \mathbf{X} 的隨機樣本。假設 \mathbf{X} 服從一未知分配 f ，則多變量常態檢定的虛無假說寫為

$$H_0 : f \in \mathbf{N}_p$$

其中 \mathbf{N}_p 為 p 維度常態分配的機率密度函數，而對立假說包含一切非常態的分配機率密度函數。

目前的研究都得到一致的結論，即任何具檢定力的檢定方法，都只對某些特定型態的對立假說資料表現高的檢定力，而對某些資料形態卻幾乎不具檢定力 (Tenreiro, 2011; Wang, 2015)。譬如，許多研究，諸如 Henze and Zirkler (1990); Horswell and Looney (1992); Mecklin and Mundfrom (2005); Wang and Hwang (2011)，都高度推薦 Mardia (1970) 的多變量峰度統計量 (multivariate kurtosis)，當其面對多變量峰度不同於多變量常態的對立假說分配時 (例如 Pearson 家族系列)，能展現很高的檢定力；反之，若遇到與多變量常態峰度接近的分配時，例如 Khintchine (GEP marginals) 與 Burr-Pareto-Logistic (normal marginal)，檢定力則降到 0.1 以下。此外，如 Henze and Zirkler (1990) 提出一系列具不變一致性 (invariant consistent) 的統計量 BHEP 也得到廣泛的討論與高度評比，但其面對不同型態的對立假說所表現的檢定力與一平滑參數 (β) 的選擇相關，除非預先知道對立假說的分配型態，否則無從選擇 β 。同樣地，Wang and Hwang (2011) 試著從多變量常態的必要條件著手；即除了各邊際變量必須服從常態外，所有邊際變量的任意組合也必須是常態；但實證結果，其提出的函數型統計量 (functional statistic) 雖然表現很好，卻也不具全方位的檢定力。

因此當面對未知型態的對立假說時，多變量常態檢定的結果往往令人存疑。既然如此，是否可以將幾個針對不同對立假說表現好的統計量集合起來，發揮集體的力量，產生一個具全方位檢定力表現的檢定方法？

Fromont and Laurent (2006) 提出一個多重檢定的方法，對資料所服從的機率密度函數做適合度檢定。該多重檢定法改良自 Bonferroni 校正法，讓多個統計量共同決

策，其決策原則傾向支持敏感度高的統計量，而非多數決，以提升整體檢定力，不過該文之實證模擬只針對單變量資料的適合度表現 (goodness-of-fit)。Tenreiro (2011) 將這個改良的多重檢定方法，用來檢定多變量常態，並挑選四種在不同特色的對立假說下，具較高檢定力的多變量常態檢定統計量，以實證模擬方式驗證其組合為一具全方位檢定力的新檢定方法。這四種具仿射不變性 (affine invariance) 的檢定統計量包括前述 Mardia (1970) 的多變量峰度與多變量偏態係數，分別簡稱 MK (multivariate kurtosis) 與 MS (multivariate skewness)，及 Henze and Zirkler (1990) 提出的一系列兼具不變一致性 (invariant consistent) 的統計量 BHEP 中的 $B(h_S)$ 與 $B(h_L)$ 。這四個檢定統計量分別對峰度、偏態、薄尾 (light tail) 及厚尾 (heavy tail) 的資料具較高的敏感度，可說涵蓋相當廣泛的對立假說。

Tenreiro (2011) 組合四個各具特色的統計量 MK, MS, $B(h_S)$ 與 $B(h_L)$ 並建立一多重檢定程序，從蒙地卡羅模擬的實證中，面對廣泛的對立假說時，展現全面性的好表現，也就是與個別的統計量相比，雖不居首位，卻穩定地保持第二名且接近第一名。因此，當面對未知來源的高維度資料時，其檢定表現確實較令人放心。

本文依據 Tenreiro (2011) 的多重檢定的程序，提出兩個多重檢定的組合，分別包含兩個與三個統計量。背後的想法有二：

1. 如同 Bonferroni 校正對 FWER (familywise error rate) 整體型一誤 α 的維持，將個別檢定的型一誤調整為 α/m (m 為檢定數量)，其檢定力隨檢定方法的數量 m 增加而降低。因此，依據 Tenreiro (2011) 的多重檢定程序，想增加檢定力，可以朝降低檢定統計量的數量著手。
2. 降低檢定數量雖然能提高多重檢定的檢定力，但也必須對廣泛的對立假說有足夠的檢定力。本文根據 Wang (2015) 做過相當規模的實證檢定力模擬，得到 WK 與 Wang and Hwang (2011) 提出的 $W_{win,m}(5)$ 幾乎輪佔各項情境 (對立假說、樣本數、維度) 的首位。因此將 $W_{win,m}(5)$ 統計量納入組合，以取代 Tenreiro (2011) 組合中的 MS、 $B(h_S)$ 與 $B(h_L)$ 。

以上想法將透過蒙地卡羅模擬，比較 Tenreiro (2011) 的組合與本文提出的兩個新組合，在面對廣泛的對立假說時的檢定力表現。第 2 節先介紹這三個組合所採用的 5 個多變量檢定統計量及其特色，第 3 節敘述 Tenreiro (2011) 提出的多變量常態的多重檢定程序，第 4 節描述本文提出的組合建議與理由，接著第 5 節利用蒙地卡羅模擬比較三種組合的檢定力表現，最後第 6 節結論。

2. 本文採用的多變量常態檢定統計量

2.1 MS 與 MK 統計量

Mardia (1970) 提出 MS 與 MK 統計量，分述如下：

- MS 統計量：

$$b_{1,p} = \frac{1}{n^2} \sum_{i,j=1}^n g_{ij}^3, \quad (1)$$

其中 $g_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})'S^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$ ， S 為樣本共變異矩陣且 \mathbf{x}_i 代表第 i 個樣本， $1 \leq i \leq n$ 。假設資料來自多變量常態分配，則 $b_{1,p}/6$ 漸進於 χ^2 分配，自由度為 $p(p+1)(p+2)/6$ 。

- MK 統計量：

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n g_{ii}^2 \quad (2)$$

這裡 $\sqrt{n}(b_{2,p} - p(p+2)(n-1)/(n+1))/\sqrt{8p(p+2)}$ 漸近於標準常態。

如 MS 與 MK 所代表的多變量偏態係數與峰度，當對立假說的偏態與峰度偏離常態時，會展現較強的檢定力，反之，遇到偏態與峰度接近常態時，檢定力會下降許多。譬如，MS 面對多變量均等分配的資料時，簡直毫無檢定力可言，而 MK 遇到較厚尾分配的資料，譬如多變量偏斜常態，檢定力幾乎都降到 0.1 以下 (Wang, 2015)。即便如此，MK 仍是許多的研究推薦較具全方位的方法。如此看來，似乎 MS 與 MK 可以合成一個統計量以為互補之用，譬如 Mardia and Foster (1983) 試著結合 MS 與 MK 的優點，提出 S_W^2 統計量，而其表現正如預期的介於兩者之間，不會很糟，但也沒有太好。

2.2 $B(h_S)$ 與 $B(h_L)$ 統計量

Henze and Zirkler (1990) 提出檢定統計量 BHEP：

$$T_{n,\beta}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = n(4I\{S_n \text{ is singular}\} + D_{n,\beta}I\{S_n \text{ is nonsingular}\}) \quad (3)$$

其中

$$D_{n,\beta} = \int_{R^p} |\hat{P}(t) - \hat{Q}(t)|^2 \varphi_\beta(t) dt \quad (4)$$

$I\{\cdot\}$ 代表指示函數 (indicator function)， $\hat{P}(t)$ 與 $\hat{Q}(t)$ 分別是多變量常態分配的特徵函數與實證特徵函數 (empirical characteristic function)，最後 $\varphi_\beta(t)$ 代表常態分配 $N_p(\mathbf{0}, \beta^2 \mathbf{I}_p)$ 的密度函數， β 稱為平滑參數 (smoothing parameter)。為便於計算，式 (4) 經進一步推演，寫成

$$D_{n,\beta} = \frac{1}{n^2} \sum_{j,k=1}^n \exp\left(-\frac{\beta^2}{2} \|Y_j - Y_k\|^2\right) \quad (5)$$

$$-2(1 + \beta^2)^{-\frac{p}{2}} \frac{1}{n} \sum_{j=1}^n \exp\left(-\frac{\beta^2}{2(1 + \beta^2)} \|Y_j\|^2\right) + (1 + 2\beta^2)^{-\frac{p}{2}} \quad (6)$$

裡面的 $\|Y_j - Y_k\|^2 = (\mathbf{x}_j - \mathbf{x}_k)' S^{-1} (\mathbf{x}_j - \mathbf{x}_k)$ ， $\|Y_j\|^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$ ，而共變異矩陣 $S = n^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$ 。

BHEP 統計量的檢定力表現與平滑參數 β 的選擇相關。Henze and Zirkler (1990) 在蒙地卡羅的模擬實證中發現，較小的 β 值，譬如， $\beta = 0.5$ ，對厚尾的對立假說有較高的檢定力。Tenreiro (2009) 從大規模的模擬研究中，建議兩組 β 值，分別針對短尾 (short tailed) 與長尾 (long tailed) 的對立假說有較高的檢定力，¹記為

$$\beta_S = \frac{1}{\sqrt{2}h_S} = \frac{1}{\sqrt{2}(0.448 + 0.026p)}$$

$$\beta_L = \frac{1}{\sqrt{2}h_L} = \frac{1}{\sqrt{2}(0.928 + 0.049p)}$$

Tenreiro (2011) 稱採用 β_S 與 β_L 的 BHEP 統計量 (3) 為 $B(h_S)$ 與 $B(h_L)$ 。總結 Tenreiro (2011) 組合的四個統計量 MS、MK、 $B(h_S)$ 與 $B(h_L)$ 分別對多變量偏態、峰度、薄尾 (短尾) 與厚尾 (長尾) 的特性具檢定力。

2.3 $W_{win,m}(5)$ 統計量

統計量 $W_{win,m}(5)$ 基於 Roy (1953) 的聯合交集原理 (UIP, union intersection principle) 認為多變量常態檢定可以表達為多個單變量常態檢定的交集。亦即針對每個單變量進行常態檢定，當所有單變量檢定都不拒絕時，才能不拒絕 H_0 。於是多變量常態檢定變成多重單變量常態檢定。Wang and Hwang (2011) 以 UIP 為多變量常態檢定的原則，並採用 Royston (1992) 延伸自 Shapiro-Wilk 的 W 統計量做為個別單變量

¹ 「短尾與長尾」常與「薄尾與厚尾」混淆不清。一般而言，短尾屬薄尾的一種，而長尾歸類於厚尾。這裡沿用短尾與長尾的名稱，來自 Tenreiro (2011) 命名的統計量 $B(h_S)$ 與 $B(h_L)$ 中的 S(hort) 與 L(ong)。

常態檢定，又因為 W 檢定統計量愈小，表示資料愈偏離常態，於是多變量常態檢定統計量的計算變成多變量函數的最小值問題，即統計量定義為

$$W_{\min} = \min_{\mathbf{c} \in R^p} W(\mathbf{c}), \quad (7)$$

其中 \mathbf{c} 代表多變量資料的組合係數，而 $W(\mathbf{c})$ 則是該組合的 W 檢定統計量。²當多變量的維度愈高，計算多變量函數的最小值將變得非常棘手。Fang and Wang (1994) 提出「投影尋蹤法」(projection pursuit (PP) approach) 來緩和這個在多維度空間的無限組合中，尋求最佳解的問題。其想法是藉由在資料空間上選取有限且均勻分佈的投射點當作組合係數 \mathbf{c} ，接著直接計算有限組合最小值即可，於是式 (7) 變成

$$W_{\min,m} = \min_{i=1,2,\dots,m} W(\mathbf{c}_i), \quad (8)$$

\mathbf{c}_i 稱為投影係數，取自 p 維度球體表面均勻分佈的 m 個點。請注意

$$\lim_{m \rightarrow \infty} W_{\min,m} = W_{\min}$$

亦即當 m 趨近無窮大時， $W_{\min,m}$ 將接近式 (7) 的解。

不過 Wang and Hwang (2011) 卻發現以式 (7) 的最小值做為檢定統計量的表現並不如預期，反而從蒙地卡羅實證發現，以距離最小值不遠的第 q 個百分位做為檢定統計量，表現更好。於是提出一函數型檢定統計量 $W_{\min}(q)$ ，定義為：

$$P(W(\mathbf{c}) \leq W_{\min}(q) | \mathbf{c} \in R^p, \mathbf{c}'\mathbf{c} = 1) = q\%. \quad (9)$$

在實務上，Wang and Hwang (2011) 採用式 (8) 的投影尋蹤法，並從對廣泛的對立假設的實證中，建議 $q = 5$ 為最佳的選擇並取名為 $W_{\min,m}(5)$ ，在本文簡稱 WH 統計量。

3. 多變量常態檢定的多重檢定程序

以 Bonferroni 校正法控制的多重檢定，要求每一個獨立檢定的型一誤都降低到 α/m 。但是當多重檢定的成員之間並非完全獨立，在維繫整體型一誤為 α 下，個別型一誤 α/m 便偏低了。不過，若不採取 α/m ，那麼那個較大（合適）的型一誤從何而來？Fromont and Laurent (2006) 提出以實證的計算方式，得到較 α/m 合理的型一

² W_{\min} 也具仿射不變性。

誤。Tenreiro (2011) 根據 Fromont and Laurent (2006) 提出的多重檢定的程序，並應用在多變量資料，其檢定程序敘述如后：

令檢定統計量

$$\mathbf{T}_n(u) = \max_{h \in H} (T_{n,h} - c_{n,h}(u)) \quad (10)$$

其中 $T_{n,h}, h \in H$ 代表有限統計量組合的個別統計量，而 $c_{n,h}(u)$ 等於 $T_{n,h}$ 在虛無假設為真的條件下，第 $(1 - u) \times 100\%$ （百分位）的臨界值。當 $\mathbf{T}_n(u_{n,\alpha}) > 0$ 時，拒絕虛無假設。³其中，

$$u_{n,\alpha} = \sup I_{n,\alpha}, \quad I_{n,\alpha} = \{u \in (0, 1) : P_\phi(\mathbf{T}_n(u) > 0) \leq \alpha\} \quad (11)$$

其中 ϕ 代表標準多變量常態的密度函數，而 $P_\phi(\cdot)$ 為其機率測度函數， $I_{n,\alpha}$ 是所有滿足 $P_\phi(\mathbf{T}_n(u) > 0) \leq \alpha$ 的 u 的集合；在資料來自多變量常態（虛無假設為真）的條件下，集合內的 u 都能讓檢定統計量大於 0 的機率被控制在型一誤以下。而 $u_{n,\alpha}$ 是集合內的最小上界，所有比 $u_{n,\alpha}$ 小的 u 值都滿足 $P_\phi(\mathbf{T}_n(u) > 0) \leq \alpha$ 。Tenreiro (2011) 證明 $\alpha/|H| \leq u_{n,\alpha}$ ，⁴即當整體型一誤為 α 時，採用 $u_{n,\alpha}$ 做為個別檢定的型一誤，至少大於等於 Bonferroni 校正的 $\alpha/|H|$ ，於是提升了個別檢定力，當然也提升了總體檢定力（至少等於採用 $\alpha/|H|$ 為個別型一誤的情況）。

本文採用 Fromont and Laurent (2006) 建議的蒙地卡羅實驗來估計 $u_{n,\alpha}$ 。在資料來自 H_0 ，且整體型一誤設為 α ， $u_{n,\alpha}$ 的估計採逐一探索的方式，從最小的 $u = 0.0001$ 開始往上計算整體的型一誤。為了涵蓋較大的 α 值，本文將 u 探索的上限訂在 0.3，間隔為 0.0001，針對不同的樣本數 n 與變量維度 p ，分別以蒙地卡羅隨機抽樣 $N = 50000$ 次來估計，探索程序如下：

1. 令 $u_i = 0.0001i, i = 1, 2, \dots, 3000$ ，計算整體型一誤 $\alpha_{u_i} = |\mathbf{T}_n(u_i) > 0|/N$ ，這裡 $|\mathbf{T}_n(u_i) > 0|$ 代表以 u_i 為臨界點，拒絕虛無假設的次數。
2. 選擇 $u_{n,\alpha} = \max_{u_i} (\alpha_{u_i} < \alpha)$ 。

根據上述程序，表 1 呈現 Tenreiro (2011) 的多重檢定組合在不同樣本數 n 與維度 p 下的 $u_{n,\alpha}$ 。表內呈現的 $u_{n,\alpha}$ 值正如預期的大於 Bonferroni 校正值 $\alpha/4$ ，且與 Tenreiro

³此時，對個別統計量皆視為右尾檢定（upper tailed test），因此當 $\mathbf{T}_n(u_{n,\alpha}) > 0$ 時，表示至少有一個統計量大於其臨界值 $c_{n,h}(u_{n,\alpha})$ ，而拒絕虛無假設。換句話說，只要有一個統計量拒絕虛無假設，等於全體拒絕之意。

⁴ $|H|$ 為組合內的檢定個數。

表 1: $u_{n,\alpha}$ 的估計：針對 Tenreiro (2011) 的多重檢定組合。

n	p					
	2	3	4	5	7	10
$\alpha = 0.01$						
20	0.0038	0.0036	0.0033	0.0032	0.0032	0.0031
60	0.0030	0.0031	0.0032	0.0031	0.0028	0.0032
100	0.0034	0.0030	0.0028	0.0029	0.0028	0.0027
200	0.0034	0.0027	0.0030	0.0027	0.0030	0.0028
400	0.0028	0.0028	0.0025	0.0028	0.0027	0.0025
$\alpha = 0.05$						
20	0.0194	0.0184	0.0181	0.0180	0.0175	0.0168
60	0.0177	0.0170	0.0165	0.0164	0.0159	0.0160
100	0.0174	0.0166	0.0162	0.0164	0.0152	0.0158
200	0.0175	0.0158	0.0163	0.0152	0.0160	0.0152
400	0.0165	0.0158	0.0152	0.0150	0.0148	0.0144

(2011) 的估計大致吻合，並在中高維度與大樣本數的條件下， $u_{n,\alpha}$ 趨近 Bonferroni 校正值： $0.01/4 = 0.0025$ 及 $0.05/4 = 0.0125$ 。

本文在進行 $u_{n,\alpha}$ 估計時，針對組合內的每一個單獨的檢定統計量皆採取關鍵值 (critical value) $c_{n,h}(u)$ 的比對，以確認統計量是否落在拒絕域。此關鍵值 $c_{n,h}(u)$ 均依賴高達 30 萬次的蒙地卡羅實驗所得的平均數，⁵而非依該統計量的理論分配去計算。這麼做的理由是，絕大多數的統計量都是漸進趨近於某個分配，也就是樣本數愈少，得到的關鍵值愈不準確，這便限制了檢定的使用範圍。另一方面，採用 30 萬次的蒙地卡羅實驗的平均，也是為了降低關鍵值估計的變異，使之愈趨近真實值。

4. 新的多重檢定的組合建議

Tenreiro (2011) 提出的多重檢定程序中，組合 MK, MS, $B(h_S)$ 與 $B(h_L)$ 四種檢定統計量（以下以四個字母 **MMBB** 簡稱之），並在實證中面對不同的對立假說，其檢定力呈現全方位的好表現。不過多重檢定雖採各家之長，卻也付出檢定力降低的缺點。因此若思索其他組合時，會先考慮較少的組合數量，其次才是置換更合適的檢定統計量。

⁵在 Tenreiro (2011) 裡，每一個檢定的關鍵值估計來自 2 萬次蒙地卡羅實驗。

4.1 兩個新組合的建議

本文的研究將以較少的檢定統計量做為組合的考量，期望達到與 Tenreiro (2011) 的四個統計量組合的全方位表現，甚至具更高的檢定力。以下提出兩個新組合，分別包含兩個與三個統計量：

1. MK + WH (以兩個字母 MW 簡稱之)
2. MK + B(h_S) + WH (以三個字母 MBW 簡稱之)

選擇 MW 的理由來自 Wang (2015) 的實證研究，就其對多種知名的檢定量所作的評比，發現 MK 與 WH 對於各類形態的對立假設資料的檢定力互有領先，兩者的表現似成互補之勢。若僅考慮選擇兩個統計量的組合，則以 MK 與 WH 為最佳組合。此外，在 MK 與 WH 兩者所擅長的資料形態之外，BHEP 系列的 B(h_S) 統計量可以填補一塊缺角（薄尾分配）。因此加入 MW 的組合，做為三個統計量組合的代表，簡稱 MBW。表 2 與表 3 呈現 MW 與 MBW 組合內個別統計量的型一誤 $u_{n,\alpha}$ （見式 (11)），因組合的數量較少， $u_{n,\alpha}$ 值相較於表 1 來自 MMBB 組合的 $u_{n,\alpha}$ ，確實提高一些。而組合內個別檢定統計量的型一誤 $u_{n,\alpha}$ 提升，是否有助於提升整體檢定力，還需要模擬實證，以確定 MW 與 MBW 是否表現出全方位的好表現，甚至比 MMBB 更好的檢定力。

4.2 顯著水準的維持與檢定力的初步評估

利用表 1 至 3 所估計的個別型一誤，接著要確認在虛無假設條件下，這三個多重檢定組合的顯著水準是否能維持在設定的 α 值。表 4 至 6 是經 50000 次的蒙地卡羅模擬所估計的顯著水準，可以看出三者大約都能維持在設定的整體型一誤 $\alpha = 0.01$ 與 $\alpha = 0.05$ 。

在確定新組合的檢定統計量能維持顯著水準後，接著進行新組合的檢定力評估，看看是否如事先預期的具備相當的檢定力，同時觀察其檢定力受組內成員檢定力的影響程度。在此選擇了四種對立假設分配，分別呈現對稱、偏斜、低擴峰、高峽峰、薄尾與厚尾的特性，圖 1 以雙變量的立體圖與等高線圖大略呈現其分配的特性。圖 2 則是由三種組合所涵蓋的五個統計量，在面對四種不同特性的對立假說時的檢定力表現。表 7 整理了這項比較的結果。

綜合表 7 的個別統計量在面對不同特色的對立假說的表現，僅 B(h_L) 沒有居於領先的地位，其表現與 WH 類似而稍遜色。另一方面，MS 雖面對 Burr-Pareto-Logistic

表 2: $u_{n,\alpha}$ 的估計：針對 MW 的多重檢定組合。

n	p					
	2	3	4	5	7	10
$\alpha = 0.01$						
20	0.0056	0.0058	0.0054	0.0055	0.0057	0.0057
60	0.0057	0.0058	0.0056	0.0058	0.0056	0.0058
100	0.0060	0.0056	0.0057	0.0059	0.0058	0.0058
200	0.0058	0.0050	0.0058	0.0052	0.0057	0.0058
400	0.0054	0.0054	0.0055	0.0052	0.0054	0.0050
$\alpha = 0.05$						
20	0.0290	0.0296	0.0290	0.0296	0.0300	0.0288
60	0.0289	0.0286	0.0289	0.0284	0.0290	0.0294
100	0.0292	0.0279	0.0294	0.0288	0.0284	0.0292
200	0.0294	0.0272	0.0296	0.0268	0.0288	0.0285
400	0.0272	0.0273	0.0283	0.0270	0.0276	0.0268

表 3: $u_{n,\alpha}$ 的估計：針對 MBW 的多重檢定組合。

n	p					
	2	3	4	5	7	10
$\alpha = 0.01$						
20	0.0042	0.0042	0.0036	0.0038	0.0037	0.0038
60	0.0039	0.0038	0.0038	0.0038	0.0036	0.0039
100	0.0040	0.0037	0.0034	0.0036	0.0037	0.0038
200	0.0042	0.0034	0.0038	0.0036	0.0040	0.0036
400	0.0036	0.0036	0.0034	0.0035	0.0034	0.0031
$\alpha = 0.05$						
20	0.0216	0.0204	0.0198	0.0204	0.0202	0.0193
60	0.0211	0.0201	0.0197	0.0189	0.0192	0.0194
100	0.0210	0.0198	0.0196	0.0196	0.0188	0.0193
200	0.0210	0.0190	0.0202	0.0186	0.0198	0.0190
400	0.0202	0.0194	0.0184	0.0184	0.0185	0.0183

表 4: 多重檢定組合 MW 的顯著水準估計。

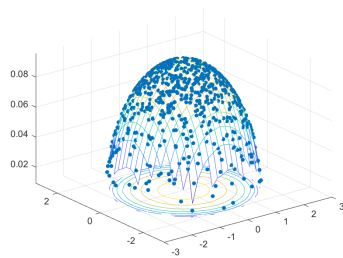
n	p					
	2	3	4	5	7	10
$\alpha = 0.01$						
20	0.0089	0.0102	0.0096	0.0092	0.0101	0.0104
60	0.0092	0.0097	0.0098	0.0109	0.0094	0.0102
100	0.0102	0.0098	0.0095	0.0109	0.0104	0.0100
200	0.0103	0.0096	0.0110	0.0101	0.0099	0.0103
400	0.0100	0.0099	0.0099	0.0085	0.0098	0.0097
$\alpha = 0.05$						
20	0.0491	0.0497	0.0507	0.0506	0.0498	0.0486
60	0.0510	0.0507	0.0505	0.0474	0.0499	0.0473
100	0.0535	0.0494	0.0514	0.0503	0.0489	0.0510
200	0.0523	0.0492	0.0527	0.0478	0.0517	0.0509
400	0.0486	0.0512	0.0528	0.0504	0.0505	0.0484

表 5: 多重檢定組合 MBW 的顯著水準估計。

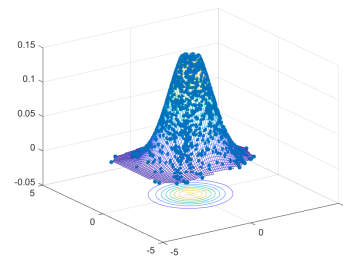
n	p					
	2	3	4	5	7	10
$\alpha = 0.01$						
20	0.0101	0.0109	0.0094	0.0094	0.0092	0.0100
60	0.0094	0.0094	0.0100	0.0105	0.0087	0.0100
100	0.0098	0.0094	0.0096	0.0101	0.0100	0.0098
200	0.0111	0.0092	0.0102	0.0105	0.0109	0.0096
400	0.0096	0.0100	0.0092	0.0095	0.0098	0.0095
$\alpha = 0.05$						
20	0.0496	0.0486	0.0509	0.0503	0.0516	0.0492
60	0.0508	0.0508	0.0501	0.0463	0.0494	0.0477
100	0.0536	0.0502	0.0503	0.0496	0.0474	0.0499
200	0.0509	0.0484	0.0521	0.0481	0.0526	0.0503
400	0.0492	0.0514	0.0493	0.0494	0.0495	0.0489

表 6: 多重檢定組合 MMBB 的顯著水準估計。

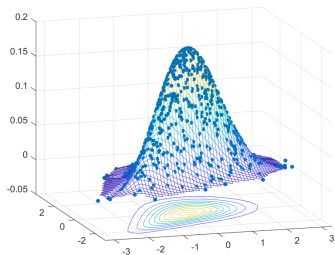
n	p					
	2	3	4	5	7	10
$\alpha = 0.01$						
20	0.0102	0.0096	0.0094	0.0099	0.0087	0.0096
60	0.0087	0.0098	0.0104	0.0104	0.0083	0.0101
100	0.0101	0.0095	0.0091	0.0102	0.0093	0.0092
200	0.0106	0.0094	0.0096	0.0101	0.0101	0.0098
400	0.0088	0.0093	0.0088	0.0095	0.0099	0.0102
$\alpha = 0.05$						
20	0.0491	0.0486	0.0512	0.0495	0.0507	0.0489
60	0.0499	0.0494	0.0487	0.0479	0.0484	0.0481
100	0.0523	0.0506	0.0495	0.0500	0.0475	0.0502
200	0.0505	0.0483	0.0518	0.0484	0.0510	0.0512
400	0.0485	0.0500	0.0500	0.0502	0.0509	0.0489



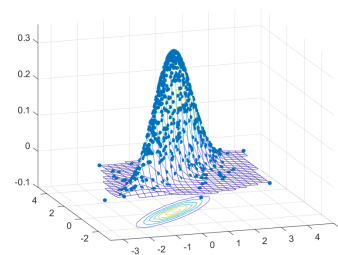
(a) Pearson Type II ($m=0.5$)



(b) Khintchine (GEP marginals)



(c) Burr-Pareto-Logistic



(d) 多變量尺度混合常態: 厚尾

圖 1: 四種對立假說分配的立體圖、等高線圖與樣本點的散佈。

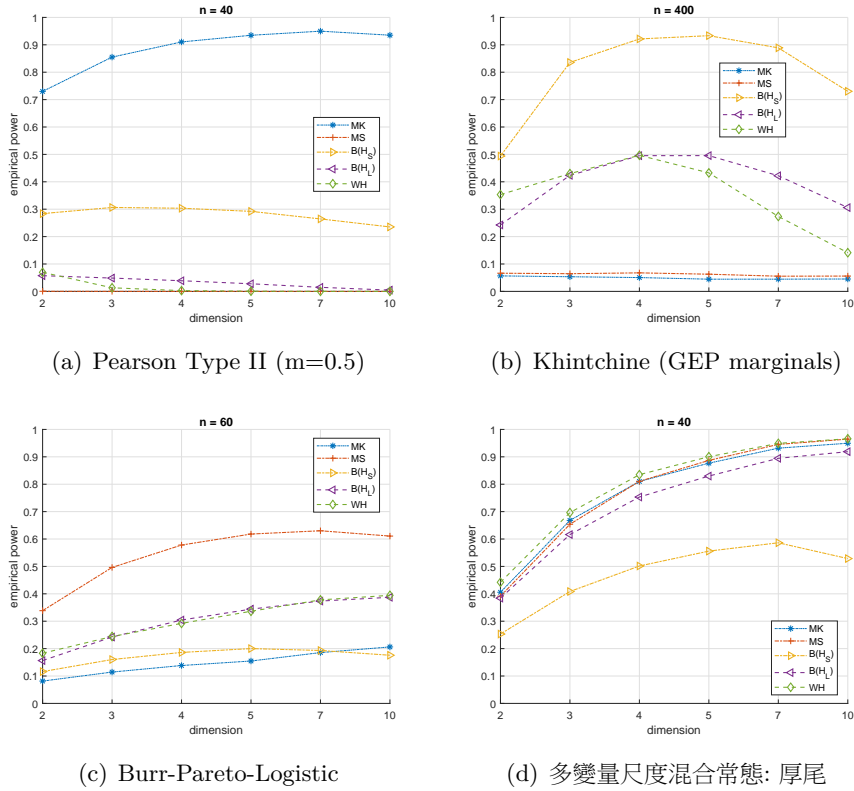


圖 2: 在顯著水準 $\alpha = 0.05$ ，個別檢定統計量面對四種對立假說的檢定力表現。

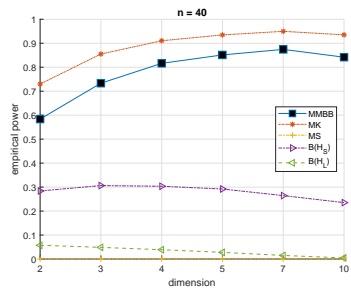
表 7: 四種對立假說的特色與五個檢定統計量的表現。

對立假說分配	特色	檢定力最好	檢定力最差
Pearson Type II($m=0.5$)	對稱、低闊峰及薄尾	MK	$B(h_L)$ 、WH、 MS^*
Khintchine(GEP marginals)	對稱、薄尾	$B(h_S)$	MS 、 MK^*
Burr-Pareto-Logistic (normal marginals)	略偏斜、厚尾	MS	MK
多變量尺度混合常態**	偏斜、高峽峰及厚尾	WH	$B(h_S)$

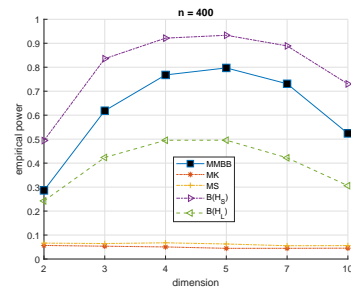
註 * 幾乎沒有檢定力。

** $0.9N(0, B) + 0.1N(0, I)$, $B = [1 \ 0.9; 0.9 \ 1]$ 。

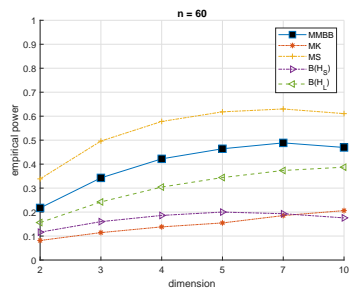
(normal marginals) 分配有最高的敏感度、面對混合常態比 WH 略遜，面對其他兩個分配幾無檢定力可言。因此，在考慮降低組合數時，便以 WH 取代 MS 與 $B(h_L)$ ，形成本文提出 MBW 的組合。



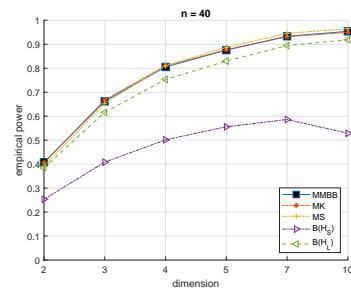
(a) Pearson Type II ($m=0.5$)



(b) Khintchine (GEP marginals)

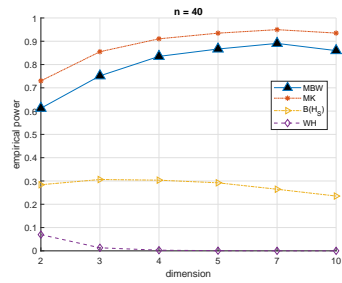


(c) Burr-Pareto-Logistic

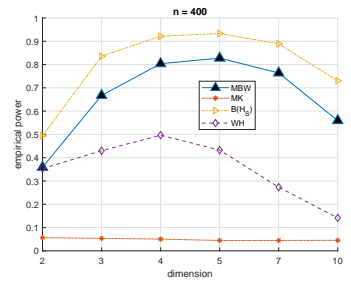


(d) 多變量尺度混合常態: 厚尾

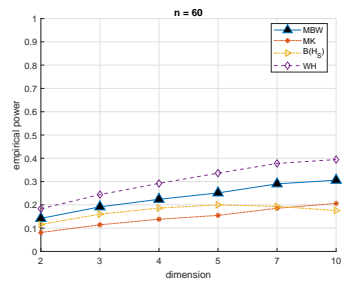
圖 3: 在顯著水準 $\alpha = 0.05$, MMBB 面對四種對立假說的檢定力表現。



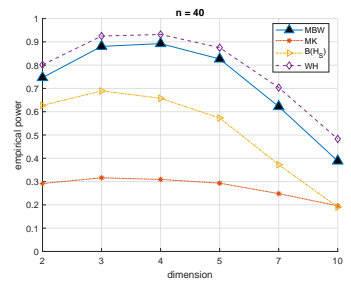
(a) Pearson Type II ($m=0.5$)



(b) Khintchine (GEP marginals)

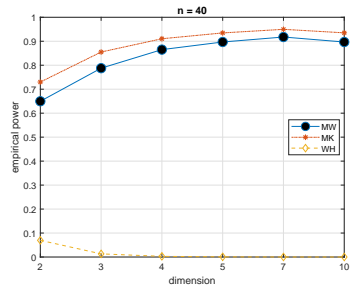


(c) Burr-Pareto-Logistic

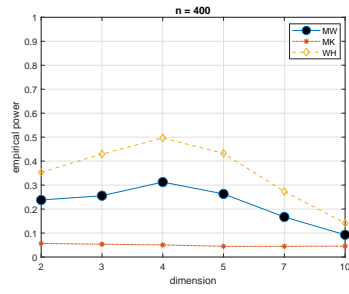


(d) 多變量尺度混合常態: 厚尾

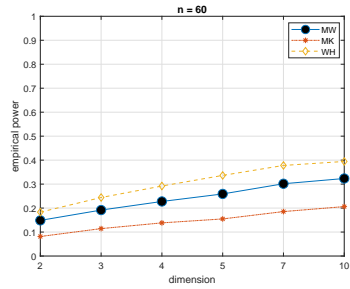
圖 4: 在顯著水準 $\alpha = 0.05$, MBW 面對四種對立假說的檢定力表現。



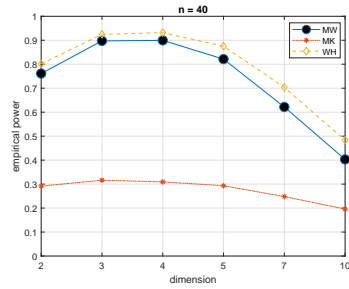
(a) Pearson Type II ($m=0.5$)



(b) Khintchine (GEP marginals)

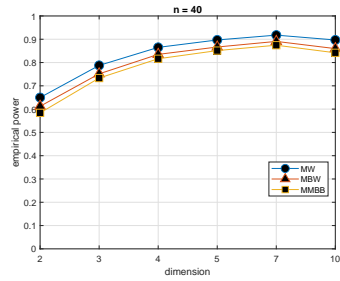


(c) Burr-Pareto-Logistic

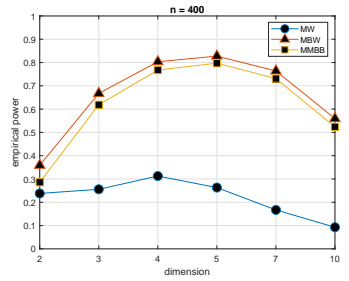


(d) 多變量尺度混合常態: 厚尾

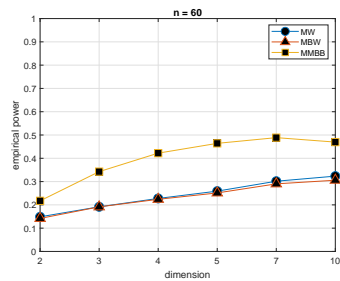
圖 5: 在顯著水準 $\alpha = 0.05$, MW 面對四種對立假說的檢定力表現。



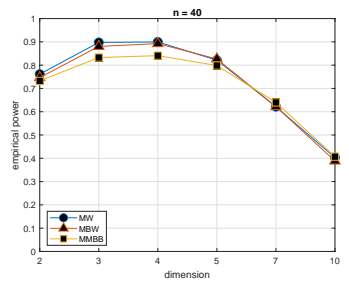
(a) Pearson Type II ($m=0.5$)



(b) Khintchine (GEP marginals)



(c) Burr-Pareto-Logistic



(d) 多變量尺度混合常態: 厚尾

圖 6: 在顯著水準 $\alpha = 0.05$, 三種不同組合面對四種對立假說的檢定力表現。

圖 3 至圖 5 則分別呈現 MMBB、MBW 及 MW 三種多重檢定的組合與其個別成員之間的檢定力比較。如原先的預期，這三個多重檢定的組合，在面對四種對立假設的檢定力都居第二位，且離第一位近，離其他較遠。圖 6 直接比較三種組合的檢定力，其中以 MW 的表現最不穩定，在面對薄尾的混合常態分配時，檢定力遠遠落後其他兩種組合。但也在 MK 與 WH 表現最好的兩個對立假說微幅領先。下一節，將從更廣泛的對立假說資料中，比較這三個組合的檢定力表現。

5. 實證檢定力的結果

本節將以蒙地卡羅實驗來檢視新組合 MW、MBW 與 [Tenreiro \(2011\)](#) 的組合 MMBB，比較三種多重檢定組合的檢定力。對立假說的分配則比照 [Tenreiro \(2011\)](#) 的選擇，採用與多變量常態接近且易混淆的分配。分配函數與其特性簡述如下 ([Johnson, 1987](#))：

1. Pearson Type II ($m=0, 0.5, 10$)：對稱、低擴峰、薄尾。
2. Pearson Type VII ($m=10$)：對稱、高峽峰、厚尾。
3. Khintchine(normal marginals)：單變量近似常態，而多變量峰度值較常態大，屬對稱、高峽峰與厚尾。
4. Khintchine(GEP marginals)⁶：單變量屬低擴峰，但多變量峰度值與偏態值皆與常態接近，屬對稱、薄尾。
5. Burr-Preato-Logistic(norml marginals)：單變量近似常態，而多變量的峰度接近多變量常態，偏態較常態略大。屬略偏斜、厚尾。
6. 多變量位置混合常態： $\pi\mathbf{N}(0, I) + (1 - \pi)\mathbf{N}(\mu, I), \mu = [3, 3, \dots, 3]'$ ，令 $\pi = 0.5, 0.79, 0.9$ 。當 $\pi = 0.5$ 時，其分配對稱，比常態薄尾。當 $\pi = 0.79, 0.9$ 時，其分配偏斜，較之常態厚尾，但是具常態的峰度。
7. 多變量尺度混合常態： $\pi\mathbf{N}(0, B) + (1 - \pi)\mathbf{N}(0, I), \pi = 0.5, 0.9$ ，其中 B 是相關矩陣，非對角線代表變數間的相關係數，設為 0.9。其分配對稱，較之常態厚尾。
8. 多變量偏斜常態 ($\alpha = 4$)：多變量的峰度接近多變量常態，偏態則較常態大。

⁶GEP 代表 generalized exponential power。

蒙地卡羅實驗次數為 50000 次，顯著水準設為 $\alpha = 0.05$ ，樣本數與多變量維度分別為 $n = 20, 40, 60, 80, 100, 200, 400$ 及 $p = 2, 3, 4, 5, 7, 10$ 。圖 7 至圖 19 僅呈現較具分辨性或討論價值的情境結果，觀察如下：

1. MW 組合在面對 Pearson 家族時，檢定力最好，MBW 居次，而 MMBB 墊底。三者檢定力最大差別在 0.1 上下，且皆受惠於 MK 對於峰度的高敏感度。因此檢定力之高低取決於組合數量之多寡。參考圖 7 至圖 10。
2. 對於 Khintchine 家族的兩個成員，三個組合的表現完全不同。面對 Khintchine(normal marginals) 時，雖其單變量十分近似常態，但多變量的峰度值大於多變量常態，因此 MK 具相當高的敏感度，這解釋了三個組合的表現都很好的原因（圖 11）。但面對 Khintchine(GEP marginals) 時，其多變量的峰度與偏態皆十分接近多變量常態，導致三個組合在樣本數低時，檢定力都不好，直到樣本數大於 200 才呈現出檢定力。其中以 MW 組合最差，落後其他兩個組合甚多，原因可從圖 2(b) 看到，只有 $B(h_S)$ 的檢定力比較高。而 MBW 則因組合數量較 MMBB 少，檢定力略高（圖 12）。
3. Burr-Pareto-Logistic(normal marginals)：這是唯一讓 MMBB 表現最好的分配，而且領先其他兩組甚多，理由是這是唯一採用 MS 的組合，而 MS 對這項分配的敏感度最好（見圖 2）。這項分配也許是讓 MS 即使面對眾多其他對立假說的檢定力表現平平，也不得不被 Tenreiro (2011) 列入 MMBB 的組合名單。參考圖 13。
4. 多變量混合常態（位置混合）：三種以位置不同混合兩個多變量常態的分配中，前兩種混合方式略成薄尾，結果以 MBW 表現最好，略優於 MMBB，而 MW 居末且相差甚遠的原因在於缺乏對薄尾敏感的組員（見圖 14 與圖 15）。第三種混合方式形成厚尾，MW 與 MBW 幾乎一致且優於 MMBB，理由來自 WANG 對厚尾的高敏感度（見圖 16）。
5. 多變量混合常態（尺度混合）：以尺度不同的混合方式，同樣形成厚尾與薄尾的不同分配。三種組合的檢定力表現亦如前面所描述的，MBW 與 MMBB 對薄尾表現較好（見圖 17），而 MW 對厚尾分配全賴 WH 而表現最好，但相差無幾，因為 MBW 也有 WH，而 MMBB 則包含對厚尾分配表現好的 $B(h_L)$ （見圖 18）。

6. 多變量偏斜常態 ($\alpha = 4$) : MW 與 MBW 因成員 WH 對偏斜厚尾的高敏感度，表現較 MMBB 好，尤其在低維度甚至到 0.1 的檢定力落差（見圖 19）。

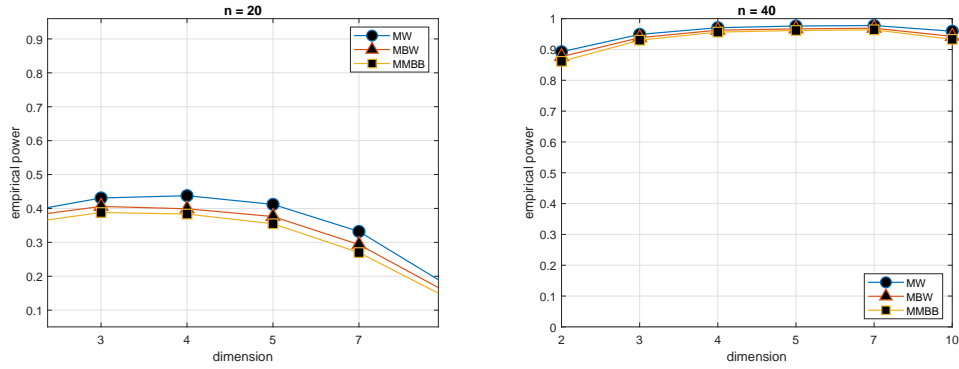


圖 7: 當對立假說為 Pearson Type II ($m=0$) 時，比較三種不同組合的檢定力。

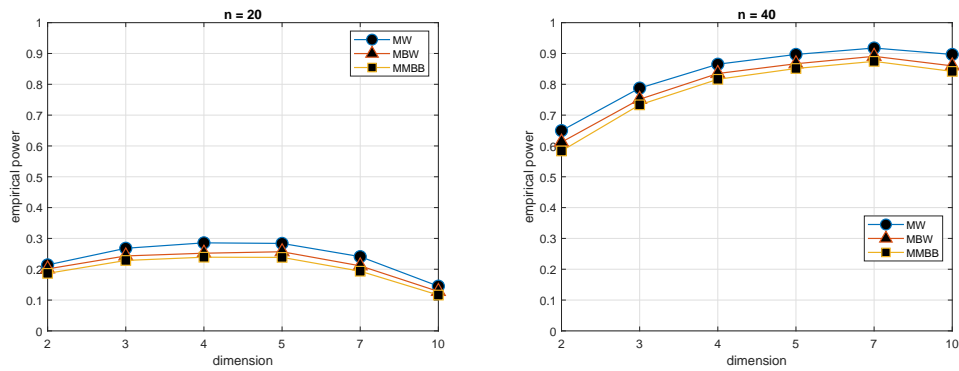


圖 8: 當對立假說為 Pearson Type II ($m=0.5$) 時，比較三種不同組合的檢定力。

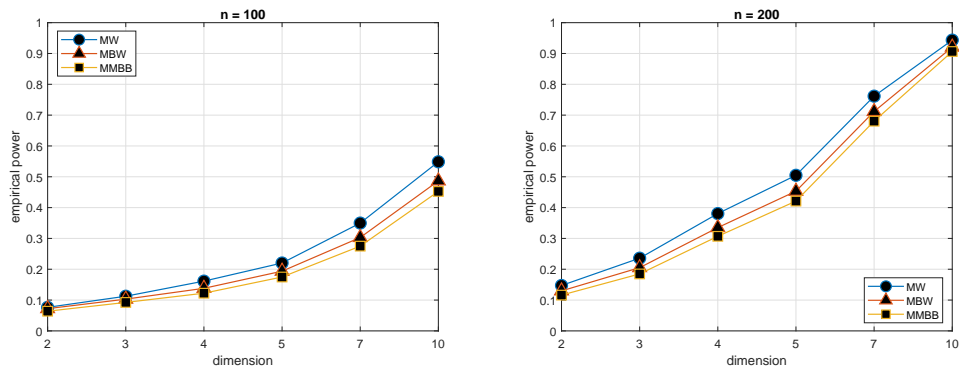


圖 9: 當對立假說為 Pearson Type II ($m=10$) 時，比較三種不同組合的檢定力。

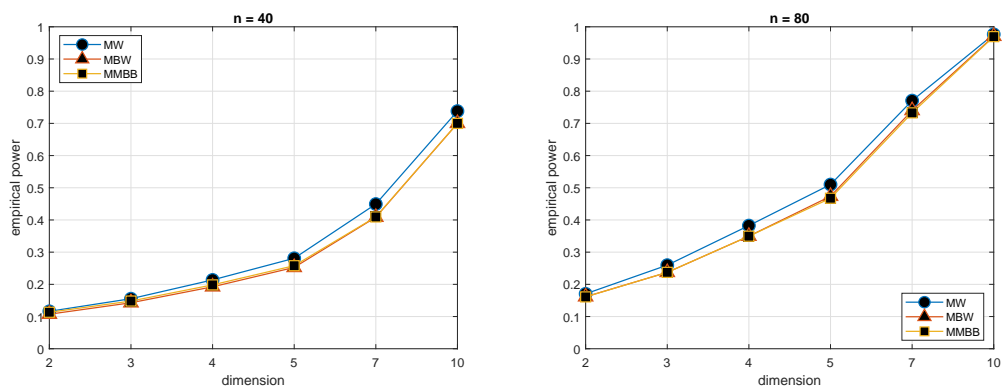


圖 10: 當對立假說為 Pearson Type VII ($m=10$) 時，比較三種不同組合的檢定力。

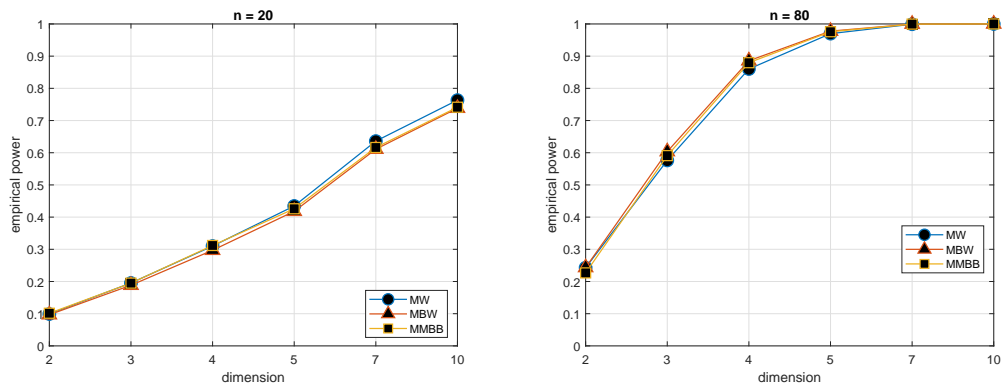


圖 11: 當對立假說為 Khintchine (normal marginals) 時，比較三種不同組合的檢定力。

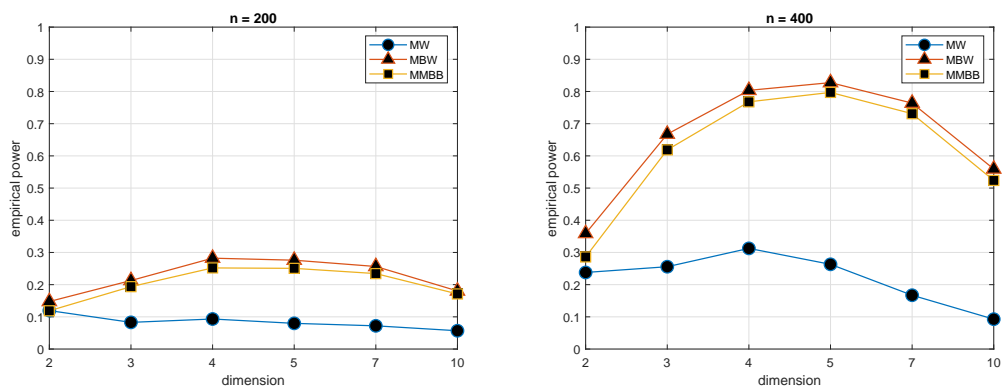


圖 12: 當對立假說為 Khintchine (GEP marginals) 時，比較三種不同組合的檢定力。

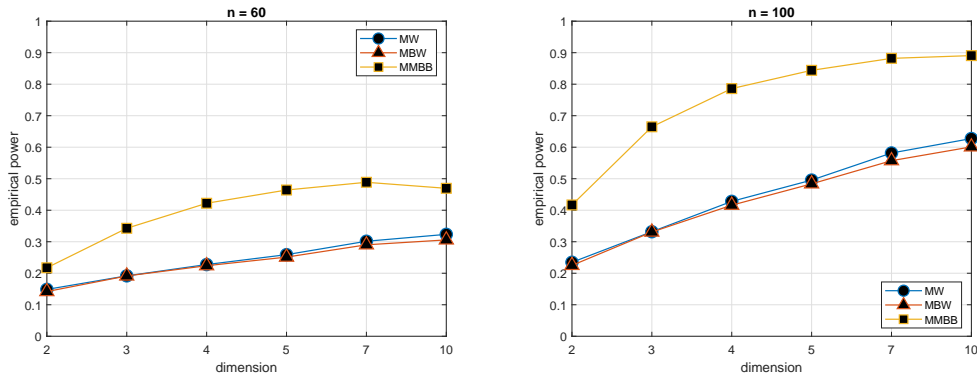


圖 13: 當對立假說為 Burr-Pareto-Logistic(norml marginals) 時，比較三種不同組合的檢定力。

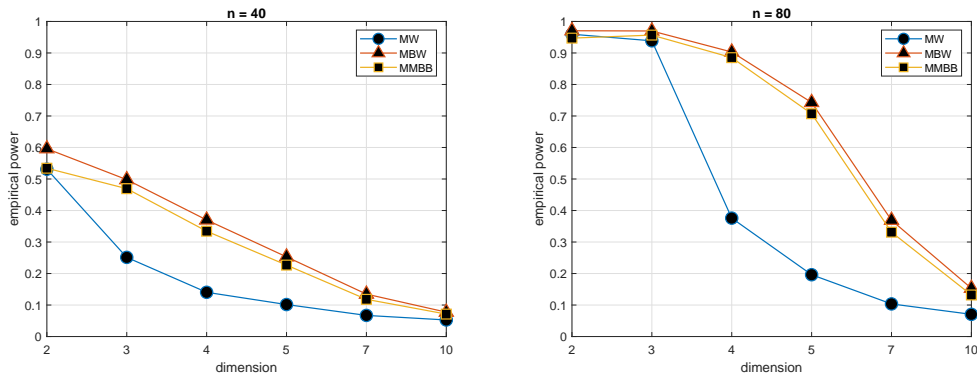


圖 14: 當對立假說為多變量 (位置) 混合常態分配 ($\pi = 0.5$) 時，比較三種不同組合的檢定力。

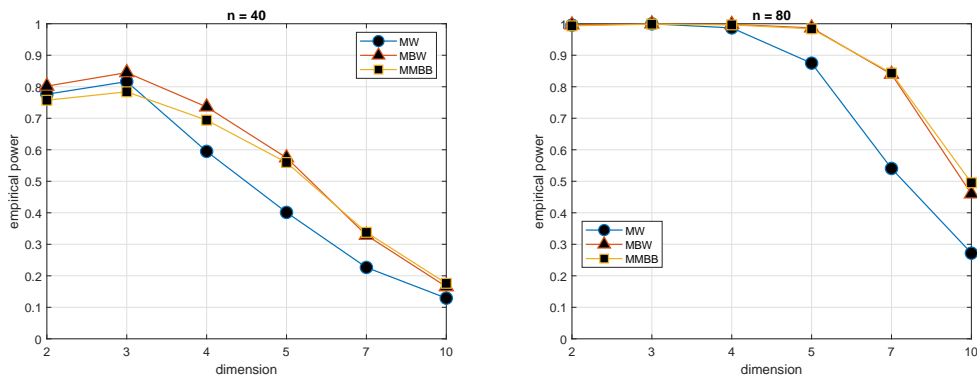


圖 15: 當對立假說為多變量 (位置) 混合常態分配 ($\pi = 0.79$) 時，比較三種不同組合的檢定力。

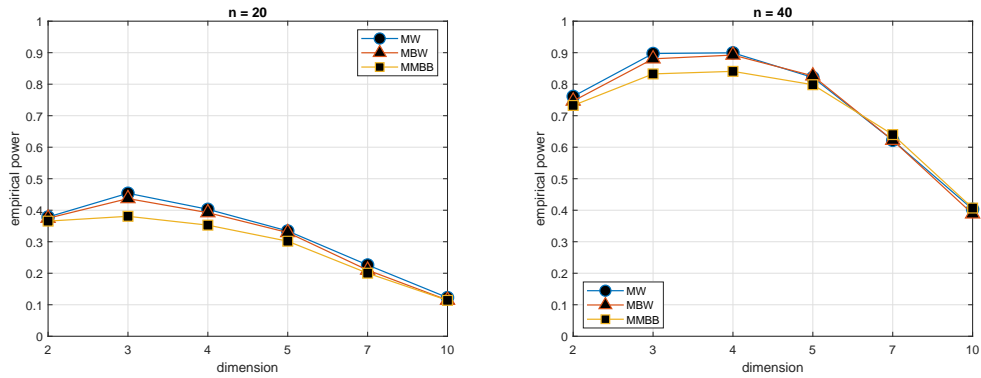


圖 16: 當對立假說為多變量（位置）混合常態分配： $(\pi = 0.9)$ 時，比較三種不同組合的檢定力。

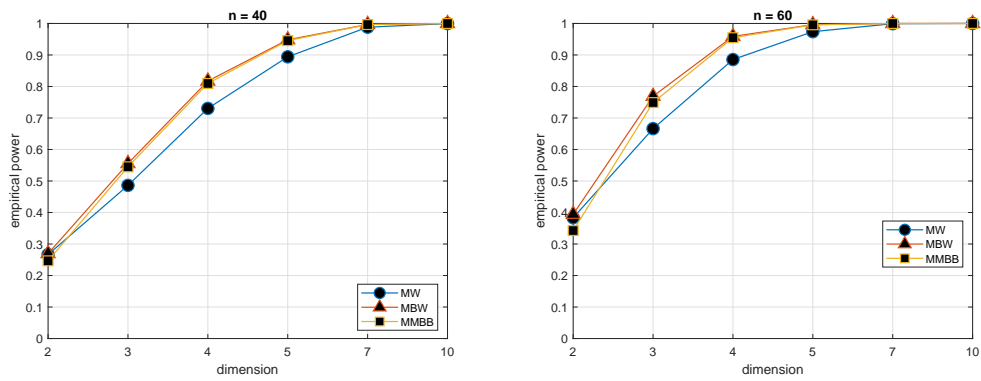


圖 17: 當對立假說為多變量（尺度）混合常態分配（ $\pi = 0.5$ ）時，比較三種不同組合的檢定力。

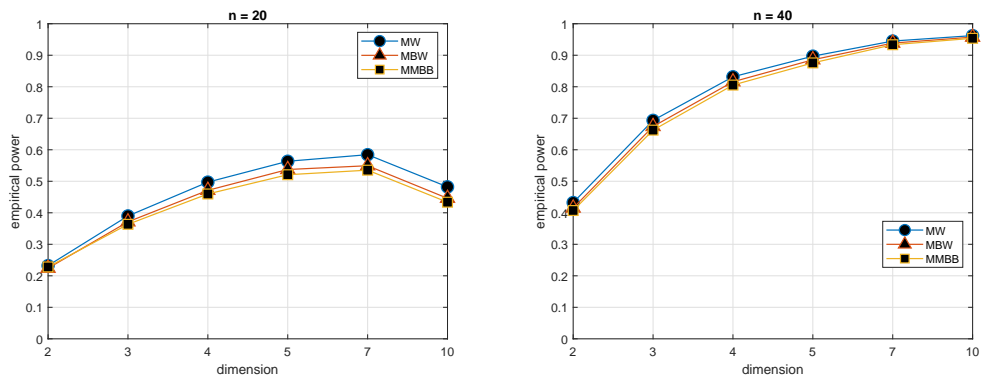


圖 18: 當對立假說為多變量（尺度）混合常態分配： $(\pi = 0.9)$ 時，比較三種不同組合的檢定力。

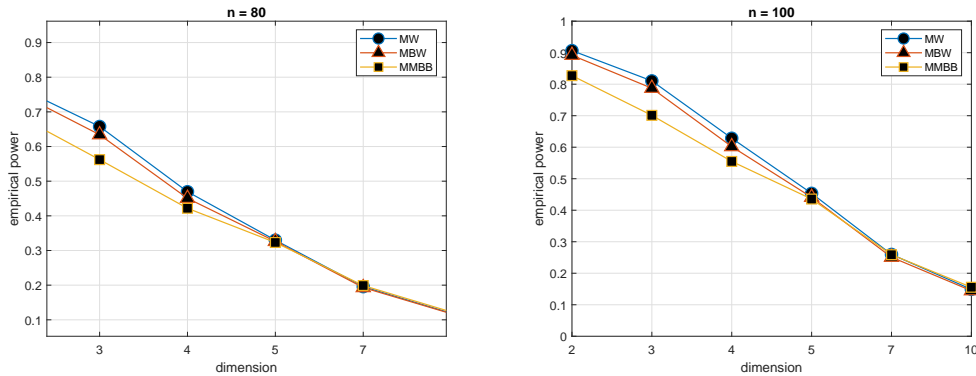


圖 19: 當對立假說為多變量偏斜常態分配 ($\alpha = 4$) 時, 比較三種不同組合的檢定力。

6. 結論

本文根據 Tenreiro (2011) 提出的多變量常態的多重檢定程序, 提出兩個多重檢定的組合, 分別含有兩個與三個具檢定力的統計量組合 MW 與 MBW, 並與 Tenreiro (2011) 建議的包含四個統計量的組合 (MMBB) 進行蒙地卡羅實證比較。在面對 13 種刻意挑選易與多變量常態混淆的對立假說資料時, MBW 與 MMBB 都展現全方位的檢定力, 且除了面對 Burr-Pareto-Logistic(normal marginals) 分配外, MBW 皆略優於 MMBB, 代表兩者的組成成員對絕大部分對立假說具敏感度, 而 MBW 因組合數量較少的先天優勢, 檢定力略高於 MMBB。另, 組合數量更少的 MW, 雖對某些對立假說的資料有最高的檢定力, 但因缺少對薄尾分配的敏感度, 尚不具全方位的檢定力。

本文的實證研究顯示, 若要提高多重檢定的檢定力, 的確可以從降低組合數量著手, 前提是組合內的統計量能涵蓋對廣泛對立假說的檢定力, 譬如圖 7 至圖 10、圖 16、圖 18 與圖 19 都證實了這個想法, 其檢定力的排序都是 MW > MBW > MMBB。

最後要說明三個統計量的組合也可以從 MMBB 拿掉較不具競爭力的 MS 變成 MBB, 但限於篇幅並未將 MBB 列入評比。有興趣的讀者可以試著做做看, 甚至嘗試其他具競爭力的統計量, 譬如 Tenreiro (2011) 建議將 $B(h_S)$ 與 $B(h_L)$ 合成 $B(\hat{h})$, 其中 $\hat{h} = \frac{1}{2}h_S + \frac{1}{2}h_L$, 或是 Mardia and Foster (1983) 結合 MS 與 MK 組成 S_W^2 統計量。另外, 構成 MW 與 MBW 組合的 WH 統計量乃基於參數 $q = 5$ 的選擇, 而 Wang and Hwang (2011) 指出最適配的 q 值會因資料分配之不同而有所差異, 因此也許可以從不同 q 值的 WH 統計量組合成對廣泛對立假說具檢定力的新組合。類似的探索式研

究，將隨著電腦的運算速提升而變得可行，甚至更輕易做到。

致謝

作者感謝兩位審查委員寶貴的意見，提示文中錯誤與含糊不清之處，使得本文的論述更為清楚與完整。另，本文的 MATLAB 模擬程式有一部分來自其他論文作者公開的程式，在此一併致謝並公開本文所使用的 MATLAB 程式於 <https://ntpucw.blog/software/>，以共襄盛舉，期能協力推動研究工作之進展。

參考文獻

- [1] Fang, K. T., and Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman & Hall.
- [2] Fromont, M., and Laurent, B. (2006). Adaptive Goodness-of-Fit Tests in a Density Model. *The Annals of Statistics*, 34(2), pages 680-720.
- [3] Henze, N., and Zirkler, B. (1990). A Class of Invariant Consistent Tests for Multivariate Normality. *Communications in Statistics: Theory and Methods*, 19(10), pages 3595-3617.
- [4] Horswell, R. L., and Looney, S. W. (1992). A Comparison of Tests for Multivariate Normality That Are Based on Measures of Multivariate Skewness and Kurtosis. *Journal of Statistical Computation and Simulation*, 42(1-2), pages 21-38.
- [5] Johnson, M. E. (1987). *Multivariate Statistical Simulation*. John Wiley & Sons.
- [6] Mardia, K. V. (1970). Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, 57(3), pages 519-530.
- [7] Mardia, K. V., and Foster, K. (1983). Omnibus Tests of Multinormality Based on Skewness and Kurtosis. *Communications in Statistics, Part A-Theory and Methods*, 12(2), pages 207-221.
- [8] Mecklin, C. and Mundfrom, D. (2005). A Monte Carlo Comparison of the Type I and Type II Error Rates of Tests of Multivariate Normality. *Journal of Statistical Computation and Simulation*, 75(2), pages 93-107.

-
- [9] Roy, S. N. (1953). *Some Aspects of Multivariate Analysis*. John Wiley & Sons.
- [10] Royston, J. P. (1992). Approximating the Shapiro-Wilk W Test for Non-Normality. *Statistics and Computing*, 2, pages 117-119.
- [11] Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit. *Computational Statistics and Data Analysis*, 53(4), pages 1038-1053.
- [12] Tenreiro, C. (2011). An Affine invariant multiple test procedure for assessing multivariate normality. *Computational Statistics and Data Analysis*, 55(5), pages 1980-1992.
- [13] Wang, C. C. (2015). A MATLAB Package for Multivariate Normality Test. *Journal of Statistical Computation and Simulation*, 85(1), pages 166-188.
- [13] Wang, C. C., and Hwang, Y. T. (2011). A New Functional Statistic for Multivariate Normality. *Statistics and Computing*, 21(4), pages 501-509.

[Received March 2022; accepted August 2022.]

The Empirical Study of Multiple Test Procedures for Assessing Multivariate Normality

Chun-Chao Wang[†]

Department of Statistics, National Taipei University

ABSTRACT

This paper proposes two multiple test procedures for assessing multivariate normality, and compares the test power with the one proposed by the original author Tenreiro (2011) by means of Monte Carlo simulation. The idea of using a multiple test procedure comes from the fact that there is no uniformly powerful statistic, so when various testing methods make different decisions, which one to believe? Tenreiro (2011) established a multiple test procedure based on the modified Bonferroni correction method proposed by Fromont and Laurent (2006). The procedure contains four statistics each showing high power in facing alternative distributions of various shapes. And its Monte Carlo experiment also verified that this procedure showed overall good performance. Using this concept of multiple tests, and based on the principle of the Bonferroni correction method, this paper argues that a smaller number of multiple tests should provide higher test power, as long as the statistics that make up the group also have test power against a wide range of alternative hypotheses. Since the author of this paper have proposed a highly competitive statistic $W_{min,m}(5)$ (Wang and Hwang, 2011) that can replace two or even three of the four combinatorial members of Tenreiro (2011), two new combinations of the multiple test procedure are then checked. Through the same empirical research as Tenreiro (2011), it is found that among the new combinations, the multiple test procedure with three statistics outperforms the combination of the four statistics of Tenreiro (2011).

Key words and phrases: multiple test procedure, Bonferroni correction, multivariate normality, Monte Carlo simulation.

JEL classification: C15, C63, C65.

[†]Corresponding to: Chun-Chao Wang
E-mail: ccw@gm.ntpu.edu.tw