
讓台灣棒球的數據分析與美國齊頭並進

張家豪

長庚科技大學嘉義分部護理系
長庚科技大學慢性疾病暨健康促進研究中心

摘要

棒球打擊者能力的評估方法眾多，最常使用的傳統指標莫過於打擊率，一般而言，打擊率高的球員被認定打擊能力優秀，對職業棒球選手而言，更具有談薪的優勢，然而這一切的理所當然，建立於打擊率的可靠性，此文章想以數據分析提出更全面的評估方法，試想，兩個打擊率相同的球員，但其中一位更常出現雙殺打，是否應該有不同的評價？本文章所提出的新指標概念為：改善 D'Esopo 和 Lefkowitz (1977) 的跑者推進模型，逐打席檢查此新的推進模型與正式結果的吻合情形，接著，選取固定的 2017 年明星賽前中華職棒富邦悍將打席次數最多的 8 人名單，在所登錄的一軍 26 位打者名單當中，逐位選取 1 人加入以形成 9 人完整打線，執行 26 次新的跑者推進模型，並排序此 26 個值即可評估特定打者貢獻得分的能力。並提出特定的出局數、壘上狀況、對戰中信兄弟投手的不同情況，呈現富邦打者的能力排序，所得的結論為：在所有的情況下，楊承駿為能力排序第一的球員。得點圈有人的情況下是胡金龍。在主力投打評估方面，面對投手鄭凱文時，胡金龍表現佳。投手為鄭錡鴻時，林哲瑄表現佳。投手為羅曼時，林益全表現佳。

關鍵詞：跑者推進模型、貢獻得分的能力、打者的能力排序

壹、前言

遠在 30 年前，美國數據大師 Bill James (1986) 分析了美國職棒大聯盟 (MLB) 相關數據，並提出了解釋 MLB 球隊在一整個球季得失分與勝負場數的關係：某支 MLB 的球隊在某個球季的總得分數等於總失分數，他們的勝敗場數也大約一致，

也就是到達五成的勝率。很直觀的，球隊要獲勝必須提昇得分及控制失分，因此，在棒球的數據分析研究當中，有許多評估球隊得分的方法被提出及應用，例如，D'Esopo 和 Lefkowitz (1977) 的 runner advancement model (RAM; 跑者推進模型)，能估計球隊每場的得分數，爾後，Bukiet,

Harold, and Palacios (1997)提出在馬可夫鏈(Markov chain)方法之下，應用 RAM 以估計不同球員能力而產生的球隊期望得分，馬可夫過程(Markov process)是一種不具記憶的隨機過程，也就是 Markov process 的條件機率只和當前狀態有關，此一特質能完整詮釋棒球場上的得分變化。以上兩個有關棒球領域的理論和指標成了爾後棒球數據分析的重要依歸，諸如；Hirotsu (2011), Hirotsu & Bickel (2014)的重新安排棒次以得到最大期望場得分；Hirotsu & Wright (2004)的更換投手策略；Fritz & Bukiet (2010)的選取 MLB 賽揚獎得主的方法；在臺灣，張家豪(2017)的改善 RAM 以貼近了中華職棒大聯盟(CPBL)使用等，都是來自於 RAM 或是 Markov chain 的應用。

除了評估球隊得分的議題外，數據分析進一步的想知道哪一位打擊者能為球隊的得分提供較大的貢獻，Fangraphs Inc. 成立了美國知名棒球統計網站 www.fangraphs.com，創造了 wins above replacement (WAR；替代贏球效果)指標 (<http://www.fangraphs.com/library/war/war-position-players/>)，Brown, Kenneth, Paul 和 David (2012)、Jensen(2013)；Baumer, Jensen 和 Matthews(2015)都曾使用 WAR 來被評價球員，以下簡單的解釋 WAR 的應用方式；在 MLB，比如某打擊者 1.5 WAR，指的是球隊打線有他的加入，與一位 3A 程度球員在打線相較，一個球季能為球隊多贏 1.5 場球，因為此訊息的實用性，在目前的 MLB 數據分析網站 WAR 被

廣泛的應用 (http://stats.slw.sephlietz.com/research/run_value)。這樣的想法，類似籃球統計指標「+/-」，此指標將某球員上場時球隊的領先或落後分數加總以形成，「+/-」越大者，代表著某球員上場貢獻越顯著。

然而，台灣棒球界或是 CPBL，對於使用數據分析來改善球隊的得分卻十分罕見，背後存在著許多原因，例如，對數據分析的陌生、對數據分析缺乏信心、懷疑數據分析的準確性、擔心數據分析增加球隊的經營成本、以及台灣使用數據分析做決策的概念不普遍等。我們提出這文章最主要目的有二（流程圖見圖 1）：第一，藉由目前使 CPBL 的統計資料，例如：採用富邦悍將和中信兄弟象的前 20 場對戰的逐位打席數據來修正 RAM，更精準的符合 CPBL 球賽的真實得分，依此提出的新的跑者推進模型(new RAM；NRAM，內容詳見表 1)，並比較 NRAM 和真實得分值的一致程度與差異原因。第二，依照 NRAM 所考慮的打擊指標，建立 CPBL 球隊的得分期望值，進而排序出特定球員對球隊得分的貢獻程度(Runs above replacement；RAR；替代得分效果)，建立類似 WAR 效果的指標。

貳、新跑者推進模型

為了解釋以下表 1 NRAM 的推進內容，每一個項目 (DE1、E1、E2 除外) 都是單一打擊者發生該結果的次數除以打席 (PA) 所形成的機率值，而 DE1、E1、E2 則是採

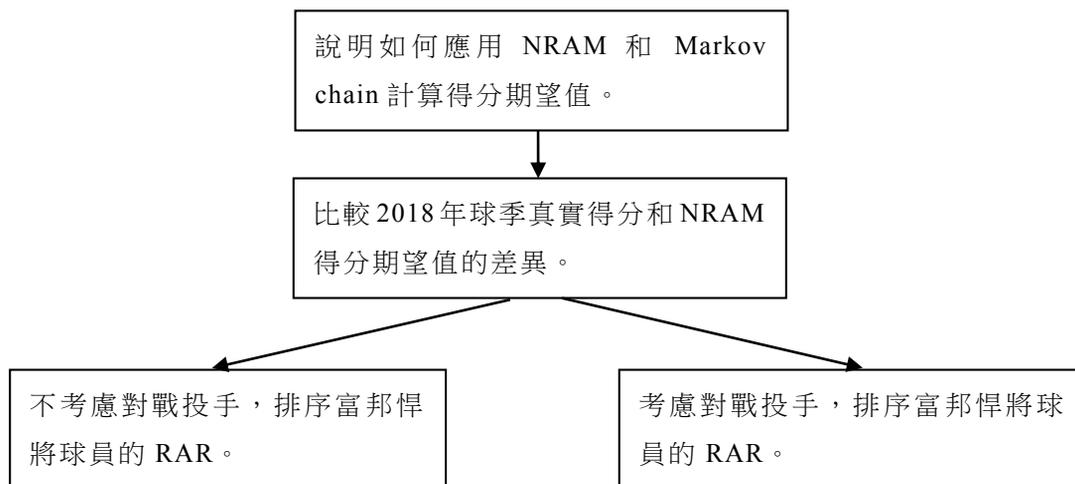


圖 1、研究的流程圖

用 CPBL 的平均機率值，我們將進行以下名詞定義，特別注意的是，以下的名詞定義與一般棒球英文縮寫定義有一些差異，原因是此計劃著重於壘上的推進效果。

BB：故意、無意的四壞球、觸身球或投手犯規。

1B：一壘安打。

2B：二壘安打。

3B：三壘安打。

HR：全壘打。

SH：犧牲觸擊。

PF：內野飛球出局。

FO：外野飛球出局。

GO：滾地球出局。

GDP：滾地球產生雙殺。

SO：三振出局。

DE1：對手守備瑕疵 1；因對手守備瑕疵推進一個壘。

DE2：對手守備瑕疵 2；SH、PF、FO、GO、SO 出現，因對手失誤而上壘。

E1：跑壘失誤 1；跑壘失誤而產生出局。

E2：跑壘失誤 2；跑壘失誤少推進一個壘。

$$BB+1B+2B+3B+HR+SH+PF+FO+GO+GDP+SO+DE2=1$$

$$BB=BB\cap DE1+BB\cap E1+\Delta BB$$

$$1B=1B\cap DE1+1B\cap E1+1B\cap E2+\Delta 1B$$

$$2B=2B\cap DE1+2B\cap E1+2B\cap E2+\Delta 2B$$

$$3B=3B\cap DE1+3B\cap E1+\Delta 3B$$

$$SH=SH\cap E1+SH\cap E2+\Delta SH$$

$$PF=PF\cap E1+\Delta PF$$

$$FO=FO\cap E1+FO\cap E2+\Delta FO$$

$$GO=GO\cap E1+GO\cap E2+\Delta GO$$

$$GDP=GDP\cap E1+\Delta GDP$$

$$SO=SO\cap E1+\Delta SO$$

$$DE2=DE2\cap E1+\Delta DE2$$

表 1、新跑者推進模型 NRAM

打擊情況	結果
BB	打者上一壘，若有擠壘才會推進
1B	打者上一壘，在無人或一人出局情況下，壘上跑者均推進一個壘；在二人出局情況下，壘上跑者均推進二個壘。
2B	打者上二壘，在無人或一人出局情況下，壘上跑者均推進二個壘；在二人出局情況下，壘上跑者得分。
3B	打者上三壘，壘上跑者得分。
HR	打者得分，壘上跑者得分。
SH	打者出局，一、二壘跑者各推進一個壘，三壘跑者強迫取分成功。
PF	打者出局，壘上若有跑者均無變動。
FO	打者出局，二人出局的情況之外，三壘跑者得分。
GO	在滿壘的情況下，三壘跑者出局，仍維持滿壘；三壘無人的情況下，打者出局，一、二壘跑者推進一個壘；三壘有人的情況下，三壘跑者不動。
GDP	打者出局，在無人出局的情況下，一壘跑者出局，二壘跑者推進一個壘、三壘跑者得分；在一人和二人出局情況下，該局結束。
SO	打者出局，壘上若有跑者均無變動。

- DE1 因對手守備瑕疵推進一個壘，例如： $BB \cap DE1$ 可解釋為打者四壞球後盜壘成功。
- DE2 SH、PF、FO、GO、SO 出現，因對手失誤而上壘，推進效果同 1B。
- E1 跑壘失誤而產生出局，例如： $BB \cap E1$ 可解釋為打者四壞球後被牽制出局。
- E2 跑壘失誤少推進一個壘，例如： $SH \cap E2$ 可解釋為打者犧牲觸擊失敗，1B 往 2B 打者出局。

資料來源：本研究者整理。

下表2為2017年富邦悍將與中信兄弟象對戰的前20場比賽當中真實得分與NRAM得分值，完全吻合的有16場，差異±1分的有4場，以下將具有差異的場次提出說明：

- (一) 日期：2017/3/29，第3局2人出局一、二壘有人，林哲瑄1B帶有一分打點，球回傳到二壘觸殺回壘不及的跑者形成三出局，因為跑者在跑回本壘之前，二壘跑者已被觸殺，得分不算。然而NRAM無法判別先跑回本壘或二壘跑者先被觸殺，因此NRAM得分值多給1分。
- (二) 日期：2017/3/30，第2局1人出局一、二壘有人，陳凱倫1B帶有一分打點，一、二壘有人。然而NRAM在1人出局的情況之下，1B只能推進一壘，也就是二壘跑者無法回本壘得分，如果紀錄DE1將形成二壘跑者回本壘得分，一壘跑者同時會推進到三壘，與真實情況一、二壘有人不符，因此我們此

並無紀錄DE1，NRAM得分值少給1分。

- (三) 日期：2017/4/8，第8局1人出局，二、三壘有人，陽冠威打擊時投手暴投，三壘跑者回到本壘得分，二壘推進三壘，之後陽冠威形成BB，一三壘有人。NRAM先執行BB形成滿壘，再執DE1，因此形成得1分二、三壘有人，之後胡金龍的1B和投手暴投讓NRAM執行二壘跑者回本壘再得1分，然而真實情況一壘跑者無法回本壘得分。
- (四) 日期：2017/5/31，第7局1人出局，二壘有人，張建銘擊出1B，二壘跑者回到本壘得分，一壘有人。然而NRAM在1人出局的情況之下，1B只能推進一壘，也就是二壘跑者無法回本壘得分，如果紀錄DE1將形成二壘跑者回本壘得分，張建銘同時會推進到二壘，與真實情況一壘有人不符，因此我們此並無紀錄DE1，NRAM得分值少給1分。

表 2 、2017 前 20 場對戰真實得分與 NRAM 得分值差異

日期	真實得分	NRAM 得分值	差異
2017/3/25	8	8	0
2017/3/29	5	6	1
2017/3/30	8	7	-1
2017/4/7	4	4	0
2017/4/8	7	8	1
2017/4/9	6	6	0
2017/4/25	8	8	0
2017/5/5	0	0	0
2017/5/6	3	3	0
2017/5/7	0	0	0
2017/5/12	7	7	0
2017/5/13	1	1	0
2017/5/14	3	3	0
2017/5/23	2	2	0
2017/5/25	7	7	0
2017/5/31	12	11	-1
2017/6/9	6	6	0
2017/6/10	3	3	0
2017/6/11	6	6	0
2017/6/20	3	3	0
2017/7/12	6	6	0

資料來源：本研究者整理。

因此，由以上說明和表2可推測；NRAM能準確的描述CPBL的壘上轉移狀況。

參、蒙地卡羅馬可夫鏈 (Monte Carlo Markov Chains) 得分期望值

以下為使用 Markov process 推論棒球場上的期望得分簡略說明。考慮進攻球隊所面臨的出局和壘上狀態可歸納成以下 25 種狀態（見表 3），分別為當 0、1、2 出局數之下的 8 種不同的壘上情形，分別為：一壘有人，二壘有人，三壘有人，一二壘有人、二三壘有人、一三壘有人、一二三壘有人，最後加上 3 人出局。

表 3、出局和壘上狀態的 25 種編碼

編碼	(壘上狀態, 出局數)	編碼	(壘上狀態, 出局數)	編碼	(壘上狀態, 出局數)
1	(0, 0)	9	(0, 1)	17	(0, 2)
2	(1, 0)	10	(1, 1)	18	(1, 2)
3	(2, 0)	11	(2, 1)	19	(2, 2)
4	(3, 0)	12	(3, 1)	20	(3, 2)
5	(12, 0)	13	(12, 1)	21	(12, 2)
6	(13, 0)	14	(13, 1)	22	(13, 2)
7	(23, 0)	15	(23, 1)	23	(23, 2)
8	(123, 0)	16	(123, 1)	24	(123, 2)
				25	(X, 3)

Markov chain 為時間分立的 Markov process，假設時間序列 $T = \{t_0, t_1, t_2, \dots\}$ ，當中 $t_0 < t_1 < t_2 < \dots$ ，在此時間序列中所有可能的狀態集合為 $X = \{x_1, x_2, \dots, x_{25}\}$ ，因此 $x(T)$ 代表了在某個時間點的系統狀態，由時間點 t_{n-1} 的狀態下 x_i ，轉變成時間點 t_n 的狀態 x_j 的機率值，便可以用條件機率 $p\{x_i(t_n) | x_i(t_{n-1})\}$ 來定義，以下我們將此條件機率簡寫成 $p_{ij}(t_{n-1})$ 或 p_{ij} ，棒球場上的意思為：在本場比賽第 $n-1$ 人次上場打擊時，

壘上狀態由 x_i 轉變成 x_j 的機率， $i=1, 2, \dots, 25$ 。舉例而言：某球員上場打擊時的壘上狀態為無人在壘，出局數為 0 人出局，該打擊者完成打擊之後，壘上狀態為 1 壘有人，出局數仍為 0 人出局，因此對應表 3 的編碼可定義為編碼 1 轉換至 2，依定義用 $p_{1,2}$ 來表示轉換的機率。考慮 25 種狀態的場上情況的轉換機率，可以使用以下 25×25 的轉移矩陣來表示：

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,25} \\ p_{2,1} & p_{2,2} & \dots & p_{2,25} \\ \vdots & \vdots & \ddots & \vdots \\ p_{25,1} & p_{25,2} & \dots & p_{25,25} \end{pmatrix}, \text{ 當中 } Q = \begin{pmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,24} \\ p_{2,1} & p_{2,2} & \dots & p_{2,24} \\ \vdots & \vdots & \ddots & \vdots \\ p_{24,1} & p_{24,2} & \dots & p_{24,24} \end{pmatrix} \text{ 為所有非終止狀態所組成的次}$$

矩陣， $R = \begin{pmatrix} p_{1,25} \\ p_{2,25} \\ \vdots \\ p_{24,25} \end{pmatrix}$ 為所有非終止狀態單次轉移至終止狀態之機率所組成的次矩陣。

如果場上由編碼 1 轉換至編碼 2 的狀態是經由 2 次打擊完成的，則轉換機率可寫成

$$PP = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,25} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,25} \\ \vdots & \vdots & \ddots & \vdots \\ p_{25,1} & p_{25,2} & \cdots & p_{25,25} \end{pmatrix} \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,25} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,25} \\ \vdots & \vdots & \ddots & \vdots \\ p_{25,1} & p_{25,2} & \cdots & p_{25,25} \end{pmatrix}$$

$$\text{當中 } (P^n)_{1,2} = \sum_{i_{24}=1}^{25} \cdots \sum_{i_1=1}^{25} p_{1,i_1} p_{i_1,j} \cdots p_{i_{24},2}$$

至於 $p_{i,j}$ 可以使用下述 2 種方式來計算，在此使用 $p_{1,2}$ 來當說明：

第一，統計 CPBL (或是某打擊者) 總共出現了多少次編碼 1 的場上狀態，又在這些狀態之下，發生多少次下一個打席形成編碼 2 的場上狀態，此即為整體 CPBL (或是某打擊者) 的 $p_{1,2}$ 。

第二，依據 NRAM 所考慮的推進內容，因為編碼 1 轉移到編碼 2 需要打擊結果為，統計某特定球員以上 3 種狀態的機率總和，此即為某球員的 $p_{1,2}$ 。

以目前 CPBL 官方網站的統計資料，並無法直接獲得以上兩種資料，然而透過比賽記分板的資訊，http://www.cpbl.com.tw/games/play_by_play.html?&game_type=01&game_id=184&game_date=2017-08-24&pbyear=2017

可以將官方網站的文字敘述內容，統整成 NRAM 所考慮的 15 種推進內容，依此，可以計算以上所提的第二種 $p_{i,j}$ ，這也是本文章轉換機率值的來源。

藉由 $p_{i,j}$ ，我們想要算出得分期望值，因此我們必須應用蒙地卡羅馬可夫鏈 (Monte Carlo Markov Chains ; MCMC)，

以下為應用 MCMC 的簡單說明。首先定義目前場上半局得分、上壘、出局狀況的矩陣 Un ，當中， Un 的「行」表示目前的得分數，總共有 21 行 (0, 1, ..., 20)，最後一行訂為 20 分的原因是；在 MLB 的單局最高得分紀錄當中，1953 年波士頓紅襪和底特律老虎曾經在單局拿到了 19 分。 Un 的「列」表示目前的壘上狀況，所以總共會有 25 列，而 n 表示該局第 n 位上場打擊的球員。而轉移矩陣 P 可以拆解成 $P0$ 、 $P1$ 、 $P2$ 、 $P3$ 、 $P4$ 五種得分效果，舉例而言： $P_{(n)0}$ 則表示該局第 n 位上場打擊球員的打擊結果得到 0 分的機率，同理， $P_{(n)4}$ 則表示該局第 n 位上場打擊球員的打擊結果得到 4 分的機率。所以 Un 提供了該局第 n 位上場打擊的球員所產生的得分結果和機率的訊息。因此，在開賽前 $Un_{[1,1]}=1$ ，其餘的矩陣元素皆為 0，意思為此時發生得 0 分、0 人在壘、0 人出局的事件機率為 1；而第一棒上場打擊後得 4 分的機率情形如下：

$$U1_{[5,]} = U0_{[5,]}P_{(1)}0 + U0_{[4,]}P_{(1)}1 + U0_{[3,]}P_{(1)}2 + U0_{[2,]}P_{(1)}3 + U0_{[1,]}P_{(1)}4$$

$$\therefore U0_{[5,]} = U0_{[4,]} = U0_{[3,]} = U0_{[2,]} = P_{(1)}4 = \{0,0,\dots,0\}_{1 \times 25}$$

$$\therefore U1_{[5,]} = \{0,0,\dots,0\}_{1 \times 25}$$

為了讓讀者更輕易了解 Un 矩陣，以下列舉了某第一棒打擊者在第一局他完成打擊後的 $U1_{21 \times 25}$ 矩陣，第1行（得0分）和第2行（得1分）向量機率分別為：

$$U1_{[1,]} = U0_{[1,]}P_{(1)}0 \text{ 和}$$

$$U1_{[2,]} = U0_{[2,]}P_{(1)}0 + U0_{[1,]}P_{(1)}1$$

其餘向量皆為0向量，因此第一局第一棒的 $U1$ 矩陣元素組成為：

$$U1_{[1,2]} = \Delta BB + \Delta 1B + \Delta DE2$$

$$U1_{[1,3]} = (BB + 1B) \cap DE1 + \Delta 2B$$

$$U1_{[1,4]} = 2B \cap DE1 + \Delta 3B$$

$$U1_{[1,9]} = (BB + 1B + 2B + 3B) \cap E1 + \Delta SH + \Delta PF + \Delta FO + \Delta GO + \Delta SO$$

$$U1_{[2,1]} = HR$$

可得知，第一局第一棒的期望得分值為：

$$0 * \text{第1行機率總和} + 1 * \text{第2行機率總和} + \dots + 20 * \text{第21行機率總和} = 1 * HR$$

因此，在半局之下，打線串連起來的得分期望值為：

$$0 * U0P_{(1)}P_{(2)} \cdots P_{(9)}P_{(1)}P_{(2)} \cdots \text{的第1行機率總和} +$$

$$1 * U0P_{(1)}P_{(2)} \cdots P_{(9)}P_{(1)}P_{(2)} \cdots \text{的第2行機率總和} +$$

...

$20 * UO P_{(1)} P_{(2)} \cdots P_{(9)} P_{(1)} P_{(2)} \cdots$ 的第21行機率總和

依此直到 $Un_{[25]}$ 的總和機率大於 0.999，也就是給定21種得分的情況之下，三出局的邊際機率大於0.999半局隨即中止，依此方式持續進行9個半局即可獲得單場比賽的得分期望值。在Bukiet et al. (1997) 的文章中提到：D'Esopo和Lefkowitz使用RAM計算每場比賽的期望得分，所得到的期望得分值略低於真實得分值（少了6.7%），而Markov chain的方法所計算出的結果與RAM相當一致。基於以上文獻結論，在本文章我們可以直接採用NRAM來獲得類似MCMC的結果。

肆、資料處理與 RAR 結果

以下章節呈現NRAM的模擬得分值，在執行 R 3.0.3 (R Development Core Team 2014)程式的模擬過程和參數方面；我們的作法為：選取固定的富邦悍將打席次數最多的8人名單，在富邦悍將所登錄的一軍26位打者名單當中，逐位加入1人以形成9人完整打線，DE1、E1、E2取此20場比賽的出現頻率機率值，分別為4.89%、1.47%、2.57%，此特定打線模擬10,000次取平均NRAM得分值，此為所評估打者的NRAM平均得分值，因此最終會有26筆

NRAM平均得分值，比序此26個值的大小，即可獲得排序的RAR，也就是可以用來推測該隊26位球員貢獻得分值的能力。若讀者對於本篇研究的程式感興趣，可以與作者連絡。

2017年富邦悍將與中信兄弟象對戰的前20場比賽當中，有上場打擊紀錄的球員如下：林哲瑄、張建銘、方克偉、李宗賢、陳凱倫、申皓瑋、林旺衛、蔡友達、陳品捷、張冠廷、黃智培、石翔宇、張清虎、楊承駿、胡金龍、吳宗峻、張詠漢、高國麟、林益全、高孝儀、楊冠威、林琨笙、林威廷、林瑋恩、高國輝、于孟雄，因本研究刊物為公眾閱覽性質，之後的內容如果涉及部分負面打擊能力的打者姓名，本文已經以代碼（FB1-FB26）代替。若球團對於本篇結果有興趣，可以與作者連絡。

表4呈現經常守備位置、PA和該球員加入產生的RAR前9名。在所有的情況當中，楊承駿、胡金龍、吳宗峻為前三名打者；當無人出局的情況下，前三名為楊承駿、吳宗峻、于孟瑄；1人出局時，前三名為張詠漢、楊冠威、高國麟；2人出局時，前三名為楊承駿、吳宗峻、林益全；得點圈無人時，前三名為吳宗峻、楊承駿、方克偉；得點圈有人時，前三名為胡金龍、張詠漢、林哲瑄。

表 4、不同情況下的 NRAM 平均得分值排序

所有情況			無人出局			1 人出局		
代碼	PA	NRAM	代碼	PA	NRAM	代碼	PA	NRAM
3B 楊承駿	14	9.53	3B 楊承駿	6	10.04	LF 張詠漢	8	9.838
SS 胡金龍	59	9.09	LF 吳宗峻	4	10.01	1B 楊冠威	9	9.4
LF 吳宗峻	18	8.23	2B 于孟馮	7	9.65	SS 高國麟	14	8.975
FB4	18	8.18	FB14	36	9.3	FB3	6	8.775
FB5	49	8.05	FB16	13	9.26	FB16	6	8.763
FB6	86	8.03	FB10	20	8.99	FB2	11	8.525
FB7	44	7.86	FB19	8	8.91	FB10	16	8.45
FB8	25	7.61	FB15	19	8.85	FB6	27	8.175
FB9	33	7.6	FB2	25	8.81	FB17	28	7.8
2 人出局			2 人出局			得點圈有人		
3B 楊承駿	5	7.6	LF 吳宗峻	12	10.238	SS 胡金龍	11	7.363
LF 吳宗峻	8	6.14	3B 楊承駿	9	9.986	LF 張詠漢	6	6.588
1B 林益全	28	6.113	C 方克偉	28	9.512	CF 林哲瑄	24	6.388
FB4	4	5.8	FB10	33	9.438	FB6	29	5.888
FB16	20	5.5	FB13	17	9.188	FB8	11	5.838
FB13	4	5.45	FB11	7	8.98	FB18	16	5.563
FB2	23	5.363	FB6	57	8.875	FB22	2	5.375
FB14	23	5.288	FB5	36	8.838	FB10	18	5.3
FB15	13	5.088	FB2	48	8.775	FB16	11	5.3

資料來源：本研究者整理。

伍、面對不同投手的 NRAM 得分值

在富邦和中信兄弟前 20 場對戰當中，以下表 5 投手名單（包含面對 PA）為中信兄弟使用頻率最高的投手群，大致而言，

鄭凱文、鄭錡鴻、伍鐸、羅曼是中信兄弟象固定的四位先發投手，江忠城、林煜清、洪宸宇定位為長中繼投手，而陳鴻文為終結者。表 6 為胡金龍、林益全、林哲瑄這

三位富邦的主要而且薪資最高打者，面對不同投手的 RAR 排序；林益全擅長打武鐸、羅曼、洪宸宇和陳鴻文，胡金龍擅長打鄭凱文、江忠城和陳鴻文，而林哲瑄擅長打鄭錡鴻和陳鴻文。表 7 為面對不同投手的 RAR 排序前三名；當中，林威廷是全隊最擅長打武鐸的打者，胡金龍是最擅長打鄭凱文的打者，高孝儀是最擅長打鄭錡鴻的打者，高國輝是最擅長打羅曼的打者，林威廷是最擅長打江忠城的打者，張詠漢是最擅長打林煜清的打者，林益全是最擅長打洪宸宇的打者，胡金龍是最擅長打陳鴻文的打者。

陸、結論

世界的棒球認知當中，Bennett and Flueck(198)提出評估一位好打者大致有以下面向：一、傳統打擊指標優異，例如打擊率高。二、打擊者在關鍵時刻能提供球隊及時的幫助，例如得點圈有人時擊出安打(打點多)。三、打擊全面，不設限任何投手投球；對決不同投手都有好的打擊表現。然而以上的評估方式，都不如：該球員的加入能為球隊貢獻得分產出。而本研究的成果，能為讀者提供以上不同面向的建議，在統計的世界當中，很難有一個「完美」的理論或黃金模型，這也是為什麼打擊指標眾多而且新的研究層出不窮的原因，目前有關 MLB 球隊得分的研究，通常有兩個方向：Markov models 和 simulation

表 5、中信兄弟對戰富邦使用頻率最高的

投手群	
投手	面對富邦 PA
鄭凱文	131
鄭錡鴻	113
伍鐸	112
羅曼	108
江忠城	50
林煜清	41
洪宸宇	39
陳鴻文	28

表 6、主要打者面對投手 RAR 排名(N=26)

投打對戰	林益全	胡金龍	林哲瑄
先發			
武鐸	3	5	6
鄭凱文	16	1	11
鄭錡鴻	10	8	4
羅曼	2	12	11
中繼			
江忠城	7	2	10
林煜清	7	12	6
洪宸宇	1	9	8
後援			
陳鴻文	3	1	2

表 7、富邦打者面對投手 RAR 前三名

姓名	PA	NRAM	姓名	PA	NRAM
武鐸			江忠城		
2B 林威廷	9	7.91	2B 林威廷	2	10.08
RF 張建銘	9	7.8	SS 胡金龍	4	8.41
1B 林益全	12	6.86	LF 高孝儀	1	7.04
鄭凱文			林煜清		
SS 胡金龍	6	6.95	LF 張詠漢	2	4.05
LF 吳宗峻	5	6.91	2B 陳凱倫	4	3.66
LF 張詠漢	4	6.38	C 林琨笙	3	3.34
鄭錡鴻			洪宸宇		
LF 高孝儀	6	8.24	1B 林益全	3	4.59
3B 楊承駿	5	8.05	1B 楊冠威	2	4.48
C 方克偉	4	7.7	LF 張詠漢	1	4.36
羅曼			陳鴻文		
RF 高國輝	4	8.44	SS 胡金龍	5	6.93
1B 林益全	12	7.63	CF 林哲瑄	2	6.04
2B 于孟馥	5	7.55	1B 林益全	3	5.86

資料來源：本研究者整理。

models, Sugano(2008)和 Hunter(2014)提到 simulation models 大多發展成可以販售給 MLB 的專利型商業套裝軟體，而 Bukiet et al. (1997)建立了 Markov model 則是基於 MLB 的數據，無論在球隊的預算方面或是模型的適用性，以上兩種模型都無法直接使用在 CPBL。筆者認為，無論任何統計模

型，都脫離不了估計，有估計就有誤差，如果本篇文章的方法用來預測 CPBL 的球隊得分誤差較低，此研究就可以提供讀者參考。本文章結果顯示：只要提供球員個別的打擊結果，即可透過 NRAM 準確的估計該場的比數，更往前進一步思考，在安排打線之前，只要蒐集好 NRAM 所需的球

員數據，即可做得分最佳化的打線安排。更特定一些，我們將場上的情況分別討論；例如，文章中提供了打點圈有人時的 NRAM 平均得分值排序前三名，或是對戰個別投手的優秀球員等，都可以提供球隊最即時的訊息。

如果不探討個別情況，截至本研究所蒐集的資料區間，楊承駿的 NRAM 平均得分值排序第一，代表該球員上場時，提供球隊實質的得分貢獻最多，如果取 PA 大於 40 的打者來評估，胡金龍會是提供球隊實質的得分貢獻最多的打者。在得點圈無人的情況下，吳宗峻、楊承駿這些非主力球員的表現很優異，相反的富邦的三位倚重的打者 FB6、FB2、FB14 的貢獻只能列全隊的第 7、9 和 17 名，但在得點圈有人的情況下，他們的貢獻可名列全隊的第 4、3 和 1 名，或許球團可以依此現象安排上場球員或是給定來年薪資。

在面對不同的投手方面，我們發現了 FB6、FB2、FB14 這三位主力球員有趣的現象；當對方投手為鄭凱文時，FB2 的打擊名列第 1，然而，FB14 和 FB6 名列 9 名之後。當對方投手為鄭錡鴻時，FB14 的打擊名列第 4，然而，FB2 全隊第 8，FB6 第 10。當對方投手為羅曼時，FB6 的打擊名列第 2，然而，FB14 和 FB2 名列 9 名之後。簡單而言，此三位主要打者面對對方三位主力先發投手時，表現都不全面：FB2 只擅長打鄭凱文，FB14 只擅長打鄭錡鴻，FB6 只擅長打羅曼。在面對中信兄弟主要救援投手陳鴻文時，三個打者表現十分優

異，分居球隊 1、2、3 名。同時，我們不禁感到好奇，哪一位打者面對這三名主戰投手表現的一致且優異，答案是：沒有。勉強來說，高國麟可以打羅曼和鄭錡鴻（第 5、6 名）的球，面對鄭凱文只列 12 名；吳宗峻可以打鄭凱文和鄭錡鴻（第 2、5 名）的球，面對羅曼只列 15 名；申皓瑋可以打羅曼和鄭凱文（第 4、5 名）的球，面對鄭錡鴻只列 17 名。

本文章的樣本數略顯不足，理想而言，為了說服球團採用 RAR，需要增加對戰場次比數數據（本文章只有 2017 年富邦和中信的前 20 場，但整個球季也只會 40 場而已），以證明此指標具有準確性高的推估能力，再者，為了獲得 RAR，必須先蒐集球員在 NRAM 所需要的 15 項指標，然而，此 15 項指標的搜集，需要閱讀 play by play 過程轉換成 Excel 資料庫甚為繁瑣，不管如何，本文章提供一個初步且筆者認為可行的方法供讀者參考。

筆者認為，科學方面的研究或許可以參考世界各國研究人員的研究成果加以延伸，但涉及到特定人種、制度、市場規模的研究，本國人應該著手進行本地的資料蒐集和研究，就好比本文章提出的職業棒球研究議題；CPBL 的規模、數據和淺在商業利益不大，很難吸引國際學者的目光從事研究，MLB 的球員特質、打擊的策略跟 CPBL 絕對有很大的差異，如果要引用他們的研究成果之前，或許要進一步評估適用性，這就是本文著手完成的主要動機。

參考文獻

- 中華職棒大聯盟(2017)：球員個人紀錄。2017 年 10 月 4 日，取自 http://www.cpbl.com.tw/players/apart.html?player_id=H044&teamno=A02。
- 中華職棒大聯盟(2017)。比賽記分板。2017 年 10 月 4 日，取自 http://www.cpbl.com.tw/games/play_by_play.html?&game_type=01&game_id=184&game_date=2017-08-24&pbyear=2017。
- 張家豪 (2017)：跑者推進模型在中華職棒的打擊資料之應用。*智慧科技與應用統計學報*，**15(1)**，31-46。
- Baumer, B. S., Jensen, S. T., & Matthews, G. J. (2013). openWAR: An open source system for evaluating overall player performance in major league baseball. *Journal of Quantitative Analysis in Sports*, 11(2), 69-84.
- Bennett, J. M. & Flueck, J. A. (1983). An evaluation of major league offensive performance models. *The American Statistician*, 37, 76-82.
- Brown, K. H., Paul, E. G., & David, G. S. (2012). An Inquiry into the Pay Structure of the New York Yankees: 1919-1941. *Eastern Economic Journal*, 38(4), 449-59.
- Bukiet, B., Harold, E. R. & Palacios, J. L. (1997). A markov chain approach to baseball. *Operations Research*, 45, 14-23.
- D'Esopo, D. A. & Lefkowitz, B. (1977). The distribution of runs in the game of baseball. In S. P. Ladany and R. E. Machol (Eds.), *Optimal strategies in sports* (pp. 55-62). Amsterdam: North-Holland Publishing Co..
- Fritz, K. & Bukiet, B. (2010). Objective Method for Determining the Most Valuable Player in Major League Baseball. *International Journal of Performance Analysis in Sport*, 10(2), 152-169.
- Hirotsu, N. (2011). Reconsideration of the Best Batting Order in Baseball: Is the Order to Maximize the Expected Number of Runs Really the Best? *Journal of Quantitative Analysis in Sports*, 7(2)
- Hirotsu, N. & Bickel, J. E. (2014). Optimal Batting Orders in Run-Limit-Rule Baseball: A Markov Chain Approach. *IMA Journal of Management Mathematics*, 27(2), 297-313.
- Hirotsu, N. & Wright, M. (2004). Modeling a baseball game to optimize pitcher substitution strategies using dynamic programming. *Economics, Management and Optimization in Sports*, 131-161.
- Hunter, M. (2014). *10 Lessons I Learned from Creating a Baseball Simulator*. Retrieved October 4, 2017, from www.hardballtimes.com
- James, B. (1986). *The Bill James Historical Baseball Abstract*. New York: Villard Books
- Jensen, S. (2013). A statistician reads the sports pages: Salaries and wins in baseball. *CHANCE*, 26(1), 47-52.
- Research: Run Value of Major Batting Events. (n.d.). Retrieved October 4, 2017, from http://stats.slv.sephlitz.com/research/run_value
- Smith, Z. J. (2016). *A Markov Chain Model for Predicting Major League Baseball*. The University of Texas at Austin.
- Sugano, A. P. (2008). *A Player Based Approach to Baseball Simulation*. UCLA Statistics.
- WAR for Position Players. (n.d.). Retrieved October 4, 2017, from <http://www.fangraphs.com/library/war/war-position-players/>

投稿日期：107 年 02 月 05 日

接受日期：107 年 04 月 10 日

Improving Taiwanese Baseball Data Analysis Goes Hand in Hand with U.S.A

Chia-Hao Chang

Department of Nursing, Chang Gung University of Science and Technology, Chiayi Campus
Chronic Diseases and Health Promotion Research Center, Chang Gung University of Science and
Technology

Abstract

There are many ways to evaluate the ability of baseball batters, the most commonly used traditional indexes is AVG. In general, high AVG batters were identified as excellent players. The professional baseball players also have the advantages of salaries. However, the reliability of AVG is questionable. This article does not negate the effect of traditional indexes, but propose a more comprehensive assessment index. The concept of the new index proposed in this article is to improve the runner advancement model by D'Esopo and Lefkowitz (1977). We also check the coincidence of the new model and the formal results. In the list of 26 roster players, select one of them to a fixed 8 batters one by one. Finally, we obtain 26 values from the new model and sort them to evaluate the ability of batters. Moreover, we rank players by dividing game conditions into the location of runners on base, outs, and pitchers. The conclusions are: Yang is the best batter all the times. Hu is the best batter in scoring position. When batters face pitcher Kevin Cheng, Hu is the best batter. When face pitcher Cheng, Che-Hsuan Lin is the best batter. When face pitcher Roman, Yi-Quan Lin is the best batter.

Keywords: runner advancement model, evaluate the ability of batters, players' rank