

# 長期穩定引文分析方法與 臺灣引文資料庫之實徵應用

楊志堅\*

文獻引用在歷經長時間的分布情形是本文企圖探討及分析的議題，藉由實徵資料與數理統計學理基礎的交互驗證，本文提出 longitudinal citation scale (LCS) 的核心概念，希望 LCS 可以更精確而忠實的反映出文獻引用在包含長、短期的時間區段內實質狀況，尤其希冀 LCS 的概念可以鼓勵學術研究社群重視文獻引用的「長期」及「穩定」發展，以此特色有別於某些強調即時或短期的引用分析指標或方法。除了學理基礎的推演，本文也將提供 LCS 在臺灣引文資料庫之應用實例。

## 一、緒論

Henry Roediger (2013) 現職為美國心理科學學會出版委員會主席 (Association for Psychological Science Publications Committee)，他有句很值得所有從事科學研究及學術發表工作的同仁們仔細體會的話，他說 “I have always done my best to feign deep interest about IFs.”。IF 即 Impact Factor (影響指數) 為近年國內學術社群相當關心的議題，但是 Henry Roediger 卻說他必須「努力假裝」對這議題很感興趣？

從另一個角度來看，任何領域若對單一指標的過分依賴，其實很可能都是即將淪為濫用或誤用的開端，即便是高度量化的學科，依然需要從各個面向以檢驗其追求科學真理的歷程的合適程度。所以又何必對單一的指標，如 IF，給予如此這般的過度關愛？但是，IF 卻又是一般社會大眾或研究社群很

---

\* 國立臺中教育大學認知神經計量實驗室教授、國科會人文及社會科學發展處教育學門 (含體育學、圖書資訊學) 召集人。

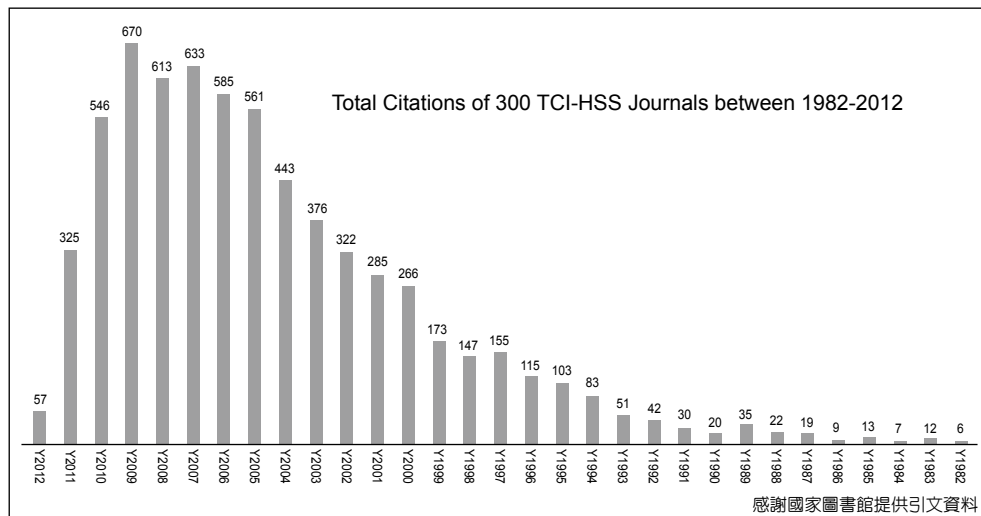


容易理解的量化數據，或許我們也應該「假裝」一下，假裝我們對這議題也很感興趣？尤其，人文社會科學研究更著重於對學術社群或社會文化的長期影響，目前常見的 IF 計算方式對於這樣的長遠影響特色能否正確的顯示出來？

IF 的計算基礎不管是 ISI (Institute of Science Index) 的 JCR (Journal Citation Reports) 報告或是 Jorge E. Hirsch 的 h-index，本質上，皆是分析或計算文獻被引用 (citation) 的狀況，姑且不論普受歡迎 (Popularity) 或常被引用的文獻是否即可代表在學理上產生實質的影響力？單就引用資料的統計內涵而言，這些引文的量化資料，其實是起源自文獻引用的次數與頻率，尤其最關鍵的問題是這些資料在固定被引視窗內 (例如：兩年內或五年內) 的比較與相對表現。

## 二、實徵範例

先從實際的例證來看，圖一是從臺灣國家圖書館選取的國內 300 本歷史悠久的學術期刊，所進行的引文資料統計圖。在 2012 全年，這 300 本期刊從 1982 年起至 2012 年所引用的歷年期刊的逐年分布情形。於 2012 這年，例如這 300 本期刊引用 2011 年 (Y2011) 的文獻的次數合計為 325 次，它們引用 2009 年 (Y2009) 的文章的總次數則為 670 次，其他年份請依此類推。



圖一

很明顯的，整體文獻的引用趨勢在最鄰近現今（2013 年）的二、三年內（2012, 2011, 2010, …），引用次數會逐年上升，但是經過了引用的顛峰期後，整體的文獻引用頻率會隨年代久遠而減少。文獻的引用趨勢就如同某些新聞事件，某一事件剛發生時，需要一小段時間傳播或流傳，當該事件的傳述逐漸到達沸騰或廣為周知之後，人們會隨著時間的增加淡忘這些事件，對該事件的關注會逐漸減低，只有極為重要的少部分事件能夠在長時間之後依然持續被人們所提起。所以能夠長期且穩定的被引述的文獻相對於僅能於短時間內喧騰是否更能有機會產生實質而深遠的影響力？

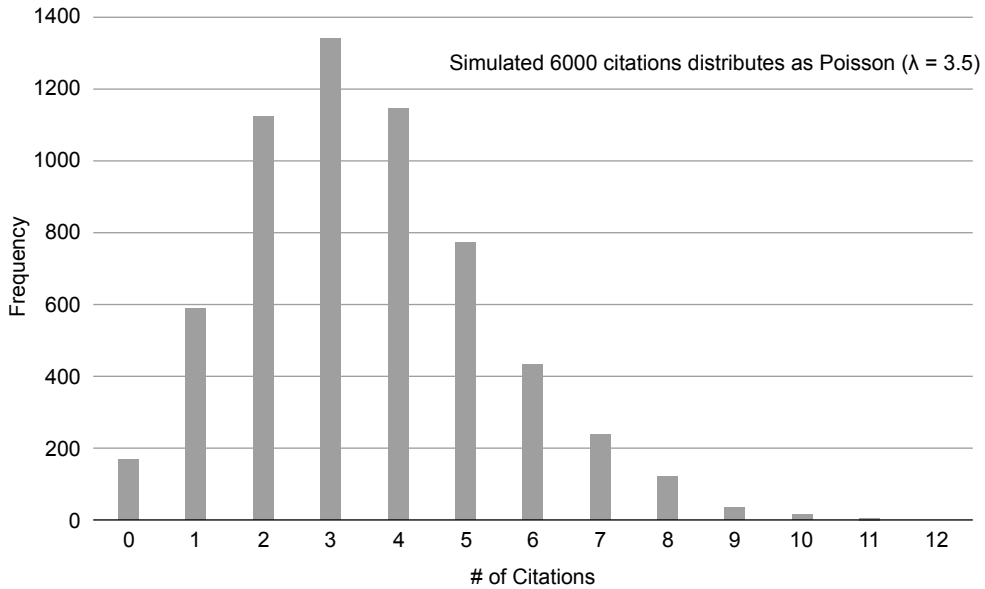
### 三、統計學基礎

如果再從數理統計的基礎上來詮釋這些資料的特徵，大部分的基礎統計教科書都會介紹到一個統計分配 Poisson（例如：Hogg, McKean, & Craig, 2012），它的基本數學表現如下：

$$\text{Prob}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots, \lambda > 0$$

它所描述的是隨機變數  $X$  等於  $k$  的機率將等於  $k$  與參數  $\lambda$  之間的關係，換成實際的例子來說，這個公式所要描述就是當  $X$  為某個事件發生的次數為  $k$  時發生的機率，例如：某個期刊  $X$  總共被引用  $k$  次的機率是可以利用  $k$  與  $\lambda$  之間的數學關係精準計算。也就是被引次數的機率是可以利用期刊平均被引用率（ $\lambda$ ）與被引次數的階層數（ $k!$ ）之間的指數關係計算出來。若利用電腦數值方法（例如：Yang, 2006）去模擬（simulate）平均引用率為 3.5 次（ $\lambda = 3.5$ ）的期刊群在某固定被引視窗中之總被引數 6000 次的 Poisson 分布圖，引文次數的分布情況將大致如圖二。

其中 Y 軸代表被引次數（X 軸）的發生頻率，例如：被引次數為 0 的發生頻率大約將近 200 次、而被引次數為 3 的發生頻率最高，可以超過 1300 次，之後隨著被引次數的增加，這類情況發生的機率會逐漸遞減，直到成為極稀少的事件，例如：極高被引數的高品質期刊是少見的。如果將這張圖與前面國內 300 篇期刊論文的被引範圍圖比較，兩者不是具有很高的相似度嗎？這個問題很容易就可以利用現有的實徵資料與統計分配的適配度檢驗方法進行驗證及解答。



圖二

如果期刊的被引資料與 Poisson 分配兩者可以類推的話，將有兩點問題需要特別注意：

1. 不同被引用次數的發生頻率是不同的，根據 Poisson 分配，當平均引用數為 3.5 ( $\lambda = 3.5$ ) 時，而且被引次數僅為 1 次 ( $x = 1$ ) 的機率為 0.105690842，但是在相同的 Poisson 分配中，若要被引次數高達 12 次 ( $x = 12$ )，此時的發生機率僅剩 0.000213034；換言之，不同被引用次數的發生機率無法等同視之。此 Poisson 分配各個次數及其發生機率的分布情形，請參考圖二。也因此可以推論出以下第二個問題：
2. 高被引的期刊、一般被引的期刊與低被引的期刊，因其發生機率的根本不同，所以，明顯的，它們將分屬在不同的量尺 (scale) 上，如果沒有經過校正，就直接以其原始被引次數進行計算的排序或指標，將會被其機率之差異所扭曲，雖然有一些標準化 (standardize) 或常態化 (normalize) 的過程可供使用，但是這些程序或許較為複雜。

#### 四、固定被引視窗內的逐年差異

如果將上述的情境拓展至多年期的期刊縱貫被引資料的情境中，從文獻的實徵資料及統計學理基礎，都可顯示出這類的資料將擁有兩個關鍵特色：一是各個時間區段內的事件發生頻率或出現機率是不同的，甚至會有極大的差異；二是這些資料的跨時段間會有相對明確的巔峰期出現。因此，很明顯的，若是直接將某幾個時段內的發生頻率加以平均計算，在巔峰期時間點內的相對大多數資料將可能主宰了其他相對少數時段的少數資料。如何兼顧每一年的不同的引用頻率（機率）不同，也就是引用難易程度各年不同的現實下，還能公平的反映出每筆引用文獻的正確相對位置？

臺灣有位很知名的經濟學家，針對類似問題舉了一個很有意思的例子：五百元的紙鈔一張在數字上或許可等於 500 個一元硬幣，但是一次拿得出來一張 500 元紙鈔跟需拼湊了很多次才能集滿 500 個一元硬幣，這兩者背後的實質意義當然大不相同。換言之，在一個固定被引視窗中的各個年度的引用數將如何被公平的對待？尤其，當統計學理或簡單的直觀都可指出久遠年度的期刊被引機率，根本上，就不可能等同於新近年度的被引率。也就是說，目前在固定被引視窗內將遠近不同的各年度的被引用率一視同仁加總後取平均值的作法，其實已經忽略了經典論著、期刊或文章的長遠影響力。

事實上，社會科學研究常見的調查方法中，也常需要面對取樣（sampling）時，母體（population）內的各族群的所占比例不平均的問題，族群不均等的問題可能會造成各族群間扭曲失真，透過對少數族群的加權（weighting），例如：分層隨機取樣（stratified random sampling）法與其對應的權重比例，就可以比較正確的趨近母體的原樣貌（Yang, & Tsai, 2008）。本文依此方向，將逐年的不同被引用率視為該年所占於所有引用母體的取樣機率，再計算這些取樣機率的倒數（inverse），就可以成為因在固定被引視窗內所產生的逐年引用率差異的權重（weights）。未來計算長期引用指標時，就應該乘上這些權重，如此，特定期刊的很多年前舊文的「少量」但「歷久彌堅」的引用情形才能公平的反映於新近議題文章的短暫高被引熱潮。因此，必須考慮以下特徵：

特徵一：能反映各年的引用機率是不均等的，就如同分層取樣時，  
必須考慮各個分層的大小可能不相同。

特徵二：每一年的引用數都是重要的，也就是強調長期且穩定的概



念，最近一、兩年內的引用情形與九年、十年前的引用狀況都必須列入考慮，並隨當年不同的引用率給予不同的權重。

其中，特徵二與被引半衰期的設計概念是截然不同的，半衰期的概念是從最鄰近點的引用情況為主要依據，然後再往前檢閱引用數的衰退或遞減情形；換言之，離最鄰近點越遠或太遠的引用數，或是斷斷續續、不均勻的引用數，最後可能都無法在這個指標的計算上發揮實質有效的影響力。

依據以上的設計概念及資料特徵，TCI Longitudinal Citation Scale (LCS) 加權法實際計算時，可以歸納成以下的執行步驟（感謝國家圖書館工作同仁的程式設計及協助整理）：

TCI Longitudinal Citation Scale (LCS) 的加權步驟（被引視窗設定以十年為例）：

- 步驟一：列出特定計算年度中，各期刊前十年歷年出版文章於計算年度的被引用次數。此後計算皆以該計算年度之數值為基準。
- 步驟二：求取各種期刊歷年的「出版文章平均被引次數」。
- 步驟三：求取歷年所有期刊「出版文章平均被引次數」的總平均數。
- 步驟四：以上述歷年總平均數中的最大值作為計算基數，求取各年度的加權指數。假設上述數值中，2009年為最大值，則以2009年的總平均數作為分子，除以歷年的總平均數，作為各年的加權指數。
- 步驟五：將各期刊歷年的出版文章平均被引次數，乘以該年度的加權指數。
- 步驟六：計算各期刊十年間，「加權後的出版文章平均被引次數」的平均數，此即為 TCI-LCS 長期影響指數。

## 五、可能限制與未來研究方向

TCI-LCS 既然是從統計分配推導而來，當然也就因襲了統計推論的不確

定性 (uncertainty) 的特質，TCI-LCS 指標面值上的大小差異，不適合以數學絕對值的算術差 (arithmetical difference) 去解讀這些差異，若僅以 TCI-LCS 表面值的數字大小進行的排序或排名，很可能排序中彼此鄰近的名次群，其實並沒有任何科學上的不同，TCI-LCS 是如此，其他常用的引用指標若也來自於對引用數的加工與綜合計算，當然亦復如此。

不同領域期刊的使用社群本身的人數大小不一，研究人員為數眾多的社群之引用數若是直接對比於全部的引文母體，其被引用機率自然容易高於小社群之引用數相對於全部母體的比例。如何適當的將研究社群分類，當然不是一件容易的工作，因此若是將引用指標數值直接進行跨社群比較，使用者必須十分的謹慎小心。

目前常見的引用指標中 (包含 TCI-LCS) 的每一本被引期刊或論述專著的品質，為了計算方便常常是被視為等值的；但是，引用網絡中的品質不均等卻是極為重要的議題，目前的 TCI-LCS 暫時未將引用網絡內的期刊品質列入考量，希望未來很快也能將此議題納入權重計算的核心之中。此外非固定引用視窗下的跨時間的發展，尤其是以期刊或專書為單位的縱貫表現之成長、衰退或維持情形，例如：初始的引用狀態與引用成長速率應該是很值得深入探究的議題。至於基於期刊引用的時間發展趨勢所做的分類或許也可以利用 Yang 與 Yang (2007) 的統計分類方法。

學術研究不應炒短線，而不重視研究內涵的深度，但也並非鼓勵貴古賤今。現有的即期引用指標 (例如：立即引用率 Immediacy Index) 當然有其貢獻，但是若想要了解長期且穩定的引文狀況，TCI-LCS 將鄰近年度或距今久遠的年度均重新給予公平的權重，是符合此一目的之合適工具。因為 TCI-LCS 極為重視**長期且穩定**的引用情形，所以若有人為或個別期刊因年度的權重不同而試圖操控某個年度的引文，相對於目前常見以兩年或五年為單位的其他引文指標的計算方式，若 TCI-LCS 的引文視窗加長至 20 年或甚至更長時，人為操控的難度將大幅升高，相對之下，TCI-LCS 應該是更穩定的指標。

另一個支持 TCI-LCS 強調長期且穩定的設計邏輯的理由是「一本好的期刊不會在短時間內有劇烈變化」，它必須具備 robust 的特質。或許現在我們再回頭讀 Henry Roediger 所說的「to feign deep interest about IFs」，就可以比較容易體會他的心境，如果 IF 真能反映期刊或專著的實質品質之良劣，又何必



費心無時無刻去關注 IF 的起伏變化？

不過，別忘了，數字不會說話，說話的是「人」、是「使用數字的人」，如果不能深入檢視期刊的本質及文章的內涵，再多的指標也不過就是一堆數字而已。本文提供了一個可能是相當有效且穩定的評量工具，但是，若沒有正確的使用態度，很可能就是另一個工具濫用或誤用的濫觴。希望研究者除了關注即時的影響能力之外，也能有相同的心力在努力發展具長遠影響力的重要研究、成就開創先河的領導型的論述或學研成果。

## 六、致謝

本文所提的期刊長期影響力研究方向，若沒有臺灣大學陳東升教授的持續極力堅持，就不會有 TCI-LCS 的原始想法，而國科會人文處鄧育仁處長、國科會人文處吳淑真副研究員的鼓勵與協調是促成實徵驗證的重要助力；當然，沒有國家圖書館工作同仁的辛勤整理，更不會有如此豐富的長期引用資料庫可供使用。本文作者衷心感激以上學者及相關工作同仁之協助，但本文所載之主張或若有錯誤，其反映的是作者的個人認識與獨自行為，概由作者個人承擔。

## 參考文獻

- Henry L. Roediger, III (2013). Journal Impact Factors: How Much Should We Care? Association for Psychological Science (APS), <http://www.psychologicalscience.org>
- Hogg, R., McKean, J., & Craig, A. (2012). Introduction to Mathematical Statistics, 7th edition, Pearson Prentice Hall.
- Yang, C.C. (2006). Evaluating Latent Class Analysis Models in Qualitative Phenotype Identification, Computational Statistics and Data Analysis, 50 (4), 1090-1104.
- Yang, C.C., & Yang, C.C. (2007). Separating latent classes by information criteria. Journal of Classification, 24, 183-203.
- Yang, C.C., & Tsai, L.T. (2008). Inferring Measurement Equivalence between Likert-type Questionnaires under Effects of Sampling Weights (in Chinese), Chinese Journal of Psychology, 50 (3), 257-269.
- 「臺灣人文及社會科學引文索引資料庫」簡介 (Taiwan Citation Index - Humanities and Social Sciences, 簡稱 TCI-HSS) <http://www.ncl.edu.tw/ct.asp?xItem=18669&ctNode=2187&mp=7>
- 「臺灣人文及社會科學引文索引資料庫」  
<http://tci.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/ccd=WqnIH8/aboutdb>