

線上消費者再購行為之探討： 機率分配和迴歸因子模型方法的比較

張順全¹ 唐璿璋² 吳全益³

摘 要

本論文為一實徵研究方法論的綜合論述，主要探討機率分配和迴歸因子模型在電子商務線上消費者再購行為實徵研究的適用情境。文章架構在 2003 年經濟學克拉克獎得主，對人類個體行為有深刻理解的實證經濟學者史蒂芬·李維特（Steven D. Levitt）一系列實證經濟學論述中，所揭櫫如何做好實徵研究的重要各面向。本篇論文除指陳若干使用「決定論」即傳統迴歸因子模型的線上消費者再購行為的實徵文章可能存在的問題，而再整理國外過去以「機率論」研究線上消費者再購行為的行銷科學相關文獻，同時以國內實際的電子商務資料驗證，並透過便於產學界應用的 MS-EXCEL 試算表實作計算。藉由實證分析線上消費者再購行為發現，應用機率分配模型確實可以作好預測未來銷售狀況，並可據以解釋再購行為的內隱機制。由於電子商務網站的客戶大多屬於非訂定合約（Non-contractual）型客戶，本論文亦討論以機率論為主體出發，加入迴歸因子為輔，對一般電子商務網站的客戶再購行為研究的可能性，這逆向思考將有別於一般以迴歸因子模型處理線上消費者再購行為議題的研究。

關鍵字：再購行為、機率分配模型、實徵研究

¹ 馬偕醫學院全人教育中心助理教授，通訊作者，E-mail: zhang@mmc.edu.tw

² 國立交通大學管理學院副院長暨經營管理研究所教授，E-mail: etang@mail.nctu.edu.tw

³ 國立交通大學管理科學研究所碩士，E-mail: stzachary10206089@gmail.com

收件日期：2014.09.19 ；修改日期：2014.09.22 接受日期：2014.10.08

Revisiting Online Repeat Purchase Behavior: Deterministic Versus Stochastic Behavioral Modeling

Shun-Chuan Chang¹ Ying-Chan Tang² Jung-Kuei Chu³

Abstract

The aim of this paper is to elucidate two intrinsically different methodologies, the stochastic versus deterministic behavioral modeling, to contribute to a better theoretical knowledge about consumer online repeat purchase behavior. It is based on Steven Levitt's series of pioneering work on empirical research in microeconomics, which engrave him to win 2003 John Bates Clark Medal, by applying natural variation incidences to answer important human behavior questions. Deterministic E-commerce studies about online purchase behavior are hard to perform and justify, and the generated panel data are often fortuitous in interpreting corresponding relationships. This revisiting research is characterized by investigating the drivers and the effects of consumer repeat purchase behavior from empirical evidences to test the behavioral postulates for predicting consumer's stochastic choices. We present the MS-EXCEL spreadsheet of a stochastic modeling process based on a real online transaction database. This calibrated data set comprises non-contractual customers who made purchases on particular CDs without certainties on consumer's repeat purchase. Results indicate the proposed stochastic model is robust in predicting online repurchase outcomes. Insights for convergence and validation of empirical relations in incorporating deterministic element into stochastic scenarios are also discussed.

Keywords: Repurchase Behavior, Stochastic Behavioral Modeling, Empirical Research

¹ Assistant Professor, holistic education center, Mackay Medical College, Corresponding Author, E-mail: zhang@mmc.edu.tw °

² Associate Dean, College of Management; Professor, institute of Business and Management, National Chiao Tung University, E-mail: etang@mail.nctu.edu.tw °

³ MBA, Institute of Management Science, National Chiao Tung University, E-mail: stzachary10206089@gmail.com °

「我們通常能預測的，是把昨天的趨勢延伸至明日。已經發生的是我們唯一能量化的，這些連續趨勢無論多重要，都只是將來的一面而已。」

-- 彼得·布拉克(1969)

壹、 續論

彼得·布拉克(1969) 在「不連續的時代」一書中作了預言：新科技將引發產業全面變革，創新將成為是否具有競爭力的關鍵。他預言現代企業所處的環境瞬息萬變，是以，把昨天的趨勢延伸至明日的外插法，這樣的預測方式僅是呈現未來趨向的一面而已。

電子商務及網路行銷正是彼得·布拉克(1969) 所言未來將出現新科技與創新的一種商業態樣。其中電子商務「線上再購」是指線上消費者藉由網際網路，買家向同一個賣家有第二次以上的購買行為。由於電子商務網站的客戶大多屬於非訂定合約(Non-contractual)型客戶，換言之，客戶首購消費後可能經常再購買、也可能隔一段相當的時間後再次消費、或永遠不再購買，職是之故，判斷客戶是否真正流失有一定的難度。事實上，可行的預測模型，我們不但要考慮再次購物的影響因素，更要考慮顧客回流的行為模式(何靖遠、賴宜楓，2012)。在過往論及線上再購的文獻中，常見資訊管理學者的傳統論述焦點乃著重於認知、信任的視角，透過消費者滿意度評價對於再購意圖或忠誠意圖的影響(喻奉天等人，2010；Yen and Lu, 2008)，至於國內分析線上消費者再購實際行為紀錄的學術論文(TSSCI)發表則首見於何靖遠與賴宜楓(2012)。該篇論文即使用國內 Yahoo! 奇摩拍賣中 T 恤類別商品在 2011 年 3 月份的真實交易資料，以及回溯買賣雙方的歷史交易資料所進行之研究。上述這篇國內的電子商務類論文吸引筆者詳視後著手撰寫本文，故本篇論文係呼應線上消費者再購行為的實徵研究之主題，而再蒐集行銷科學的相關文獻，嘗試作進一步闡述。

為了再論此一主題的實徵研究更具類推性或稱通則性(generalizability)，筆者參考了現任「美國芝加哥大學價格理論中心」(Initiative on Chicago Price Theory)執行長史蒂芬·李維特(Steven D. Levitt)所設計的課程：「如何做好實徵研究」(How to Do Empirical Research)¹中所臚列的重點。這位 2003 年經濟學克拉克獎(John Bates Clark Medal)得主，如同當年克拉克獎評審的頌詞略以：Steven Levitt is the most innovative empirical researcher in his cohort…。「如何做好實徵研究」該課程課綱所提到的第一個重點是成為首次解決特定主題的創作(Be the first to tackle a problem)，顯見相當肯定

¹網址: <http://pricetheory.uchicago.edu/levitt/home.html>

首次發表原著的實徵研究的重要性，一如啟發筆者研究緣起於何靖遠與賴宜楓(2012)在電子商務學報發表的論文，該文結合行銷領域的 RFM 模型以及拍賣網站所提供評價資訊，建立線上再購行為影響因素的研究模型，對國內電子商務學術研究實具貢獻。更值得一提的是，奠基過去的再論型創作也是重要，也是本篇所採取的討論面向如李維特的「如何做好實徵研究」課綱所揭櫫的，尚有其他重點例如：修補文獻的缺口 (Fix a literature that is broken)，引用新資料(Find new data)，更精進的驗證 (New and better sources of identification)，以及理論與實務的整合相應(Integrate theory and empirics)等，按此，李維特對精進實徵研究的見解，筆者擬引以再論線上消費者再購行為的實徵研究時，作為檢證論述的根據。

在傳統行銷科學文獻中對一般顧客再購行為的研究已有相當久的時間，有利用顧客終生價值 (CLV, Customer Lifetime Value) 來追蹤客戶行為，體現顧客總終身價值之折現值(Rust, Lemon, and Zeithaml, 2004)；也有 RFM 分析模型(Hughes, 1994)應用於資料庫行銷中，利用客戶的同質性分群，企業透過篩選顧客群、與其保持聯繫、傳遞訊息等，據以提升視為公司策略性資產的顧客權益(Kumer and Petersen, 2005)。

申言之，常見如使用迴歸模式來預測顧客未來購買行為的因果推論，其中牽涉到使用許多可觀測的影響因子，諸如利用 RFM 分析模型中最近購買時間(recency)、購買頻率(frequency)及購買金額(monetary) 等來預測購買行為的思維屬於科學哲學的決定論(determinism) (Talbot, 2008)。但是，使用多元迴歸分析的經驗中吾人得知，如果自變項的個數很多，有時候就要用調整後的判定係數代替表示模型適配度，因為增加過多的自變項，往往誇大了其模型和資料的契合程度 (Neter et al., 1996)。

再者，迴歸模型設計允許有誤差項存在，掌握如何詮釋誤差項則歸於所謂的機率論(stochasticism)。行銷科學文獻使用機率論預測一般顧客未來的購買行為是相當豐富的，冀望應用不同的機率分配來掌握顧客購買行為中無法觀測到的動態變化，這些模型假設通常都是利用機率論中二項(Binomial)、卜松(Poisson)或指數(Exponential)分配來捕捉顧客購買的選擇、次數、以及時間。Ehrenberg (1959)也是利用 Poisson-Gamma 即負二項分配(Negative Binomial Distribution, NBD)模型，估計與預測顧客的購買次數，奠定此類消費行為探討的文獻基礎；Schmittlein、Morrison、與 Colombo (1987)在管理科學期刊(Management Science)提出 Pareto/NBD 模型，將購買模式的內涵分成交易過程(transaction process)與退出過程(dropout process)；至於線上消費者購買行為的研究，Fader 與 Hardie (2001)根據美國線上光碟銷售網站 CDNOW 的客戶交易資料，利用 Beta-Geometric 模型來預測該網站的線上消費者首次和重複購買行為，且無須使用任何自變項即可精確預估長期變化。尤有進者，Fader 與 Hardie 等

(2005a)仍使用 CDNOW 的客戶交易資料進行實徵研究，並提出 BG/NBD 模型 (Beta-Geometric/NBD)，有效解決原本 Pareto/NBD 模型計算不易的問題，並且根據每位顧客的歷史交易資料，預測未來的購買次數，並且評估顧客異質性，計算個別顧客之活躍機率(active probability)。是以，機率論在線上消費者再次購買議題，國際間已有許多重要文獻發表。但是，究竟機率分配和迴歸因子模型方法使用時機為何？模型預測力和解釋效果孰重？此又將如何從實證中回答？這些正是本文關注的研究問題。

本文總共分成五節，第一節在介紹研究背景、動機、目的與問題，第二節為文獻探討，著重在修補線上再購行為的研究文獻的缺口，第三節說明本文研究方法，著重在引用新資料進行實證，第四節是資料分析，著重在使用更精進的驗證方式論述線上消費者再購行為，最後是討論，則著重在理論與實務的相應整合。

貳、文獻探討

本節著重在修補線上再購行為的研究文獻的缺口。在前節介紹中，吾人可以得知國內以機率論視角論述線上消費者再購行為較為欠缺，如果因為對隨機現象的內隱機制漠視，而認為線上消費者再購事件的偶然性係完全不可預測，則是誤解。反之，如果全然假定決定論，即每一件再購事件都是可以掌握的，則或許也掩蓋了許多意料之外發生的實況。事實上，由決定論原理來說明線上再購行為的因果關係，和預測將來再購事件的發生，存在著邏輯的和實際上的困難。在邏輯上來講，因為推論時往往是處於一個複雜的因果關係鏈上，我們有可能在某關係階段探索的因，其實是其他階段影響的果，長此以往，我們可能陷於只看局部，或迷惘於一個沒有終止的因果鏈。邏輯上的「不必要」與「不可能」處理過度複雜的因果關係，均將迫使我們通過描述偶然性的結構來模型化事件，申言之，常用多元迴歸分析的殘差項通常就是在缺乏對其他因素的知識或測量上存在誤差(Neter et al., 1996)時，由殘差項的機率內涵來詮釋模型中的偶然性的結構。

再者，實務上，引起一個事件的原因可能會有諸多的因素，相關因素也可能在特定條件下影響力各有消長或存在共線性。由於使用的模型通常決定了如何詮釋行為的面貌，例如何靖遠與賴宜楓(2012)結合行銷領域的 RFM 模型以及拍賣網站所提供評價資訊，建立線上再購行為影響因素的實徵研究中，R、F 是顯著再購行為影響因素，過去累積交易金額即 M，並不會顯著地影響買家對同一賣家在六個月內的再購行為，但該文章事後的分層分析，則顯示單變量分析有影響，我們認為這表示該模型其中購買頻率 F 與購買金額 M 兩個變量間應存在共線性，可能因為當購買頻率增加，

累計金額也相對增加所致。另外，由此也顯示存有不同實徵研究認為 RFM 模型中的三個指標的重要性是不一樣的，和 Hughes(1994)原提出 RFM 模型中的三個指標的重要性是一樣的結論相左。

尚且，從科學哲學觀之，何靖遠與賴宜楓(2012)的研究推論係屬決定論的思考，該篇的研究限制部分也點出不同產品品類(例如該篇使用奇摩拍賣中 T 恤類別商品)對 RFM 屬性有不同的敏感性。換言之，某些品項對 F 屬性有較好的區隔顧客購買行為的能力，但其他產品則可能是對 R 較佳。另外，該篇使用邏輯迴歸模式以 R、F、M 為致因，建構再購行為影響因素，此模型乃適合於根據顧客過去的歷史交易資料來區隔顧客的再購於否，但該模型並不是可用以精確預測未來購買次數。

下圖 1 即機率分配和迴歸因子模型概念的比較，吾人可用以釐清兩者間如何的解釋與預測網際網路中買賣雙方一段時間內的序列交易所構成的再購行為。

因果決定論如圖上半部所示，關注的是利用統計方法，推論建立在歷史資料上，假設未來世界的發展趨勢和研究當下的情境相仿。但是，隨著當今電子商務崛起中起伏多變的氛圍，未來不同典範的世界可能造成所解釋與預測常常失靈，總體經濟現象如此，虛擬世界的商業活動亦然。機率論所關注的不同，如圖 1 下半部所示，與傳統決定論研究之不同處，在於利用顧客的歷史交易資料建模，求得顧客潛在特徵值/參數，甚至也考量顧客個體的異質性，再根據特徵值對未來線上交易行為作預測。這樣模型化「偶然(不確定)性」的過程中並無須使用任何可觀測的自變項加入模型。

Fader 與 Hardie (2001)研究美國線上光碟銷售網站 CDNOW 的客戶交易資料即是把線上再購現象，利用一連串合理的機率論假設說明，由此，推導出具衡量可能性程度的結論。Fader 與 Hardie (2001)針對美國線上光碟片銷售網站 CDNOW 提出的線上交易預測模型，乃結合線上消費者初次購買模型及重複購買行為模型，共有下面五個基本假設：

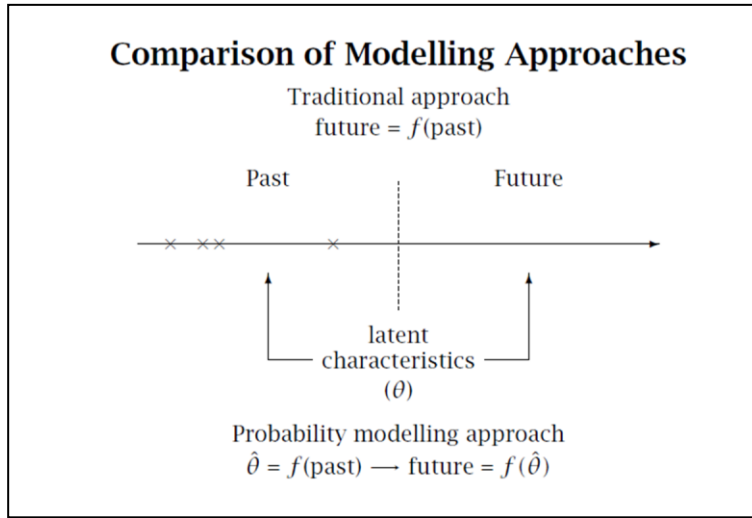


圖 1、機率分配和迴歸因子模型概念的比較

資料來源：http://www.brucehardie.com/talks/cba_workshop_paris_2007_handout.pdf, P219.

假設一：每位客戶第一次購買設為以 qT 為參數的 Geometric 分配。令變數 T_w 代表在 w 週首次購買的 CD 片數，此時 T_w 為大於或等於 0 之整數，直觀來說，當商品相對同質性高(如 CD 或 T 恤)，且控制在單一賣家時， qT 可表示客戶第一次購買決定 CD 數量的過程中，每加購一片的考量依據乃如同丟銅板， qT 為丟銅板正面的機率，遇到正面就一直買，直到出現反面就停止購買。越大的 qT 表現首次購買數量越多的傾向，由於市場整體表現是客戶個體行為的集合概念，越大的 qT 會決定市場上普遍多購的可能性較高，但是某些客戶結果不必然多買。基此，每位顧客購買片數機率 $P(T_w = x | q_T)$ 為以 qT 為參數的 Geometric 分配，其機率密度函數為：

$$P(T_w = x | q_T) = \begin{cases} q_T (1 - q_T)^{x-1}, & x = 1, 2, 3, \dots; 0 < q_T < 1 \\ 0, & x = 0. \end{cases} \quad (1)$$

假設二：加入考量顧客個體購買力的異質性，假設 qT 並非常數，而是每人有異。基此，不同顧客的 qT 設為以 αT 、 βT 為參數之 Beta 分配，其機率密度函數為：

$$g(q_T) = \frac{1}{B(\alpha_T, \beta_T)} q_T^{\alpha_T-1} (1 - q_T)^{\beta_T-1}, \quad 0 < q_T < 1; \alpha_T, \beta_T > 0. \quad (2)$$

結合兩式得到顧客首次購買片數機率 $P(T_w = x | q_T)$ 如下:

$$P(T_w = x) = \int_0^1 P(T_w = x | q_T) g(q_T) dq_T$$

$$= \begin{cases} \frac{B(\alpha_T + 1, \beta_T + x - 1)}{B(\alpha_T + 1, \beta_T)}, x = 1, 2, 3, \dots \\ 0, x = 0, \end{cases} \quad (3)$$

顧客首次購買片數 T_w 稱之為(Shifted) beta-geometric distribution。而其期望值為:

$$E(T_w) = \frac{\alpha_T + \beta_T - 1}{\alpha_T - 1}. \quad (4)$$

假設三: 每位客戶不買的機率為一個與時間相關的機率分配。

假設四: 客戶重複購買的片數機率為以 qR 為參數的 Geometric 分配。

假設五: qT 、 qR 互相獨立。

根據假設三, 顧客不買的機率是會隨時間改變。令 $\pi_{w|i}$ 代表在第 i 週第一次購買的顧客, 接下來第 w 週 ($w > i$) 不買的機率為:

$$\pi_{w|i} = 1 - \gamma(w - i)^\delta, w > i, \quad (5)$$

$\pi_{w|i}$ 代表隨時間的增加除了時差 ($w - i$) 會受到 γ 的校正外, 當 $\delta < 0$, 表示一般狀況隨著時差增加, 顧客會流失的機率更快, 這是冪級數的影響效果。當然 δ 也可能大於 0, 代表重複購買的機率與日俱增。這裡所指的“不買”(inactive)並非就表示顧客永遠流失, 只是表示顧客在未來第 w 週時當下無心入市購買, 將來還是可能再購。

根據假設四, 在第 i 週第一次購買的顧客, 在第 w 週 ($w > i$) 客戶重複購買的片數寫成 $R_{w|i}$, 重複購買的片數機率表示為參數 qR 的 Geometric 分配, 其機率密度函數為:

$$P(R_{w|i} = x | q_R) = q_R (1 - q_R)^x \quad x = 0, 1, \dots; 0 < q_R < 1, \quad (6)$$

而其中 q_R 依據假設二、五表示為以 α_R 、 β_R 為參數之 Beta 分配，其機率密度函數為：

$$g(q_R) = \frac{1}{B(\alpha_R, \beta_R)} q_R^{\alpha_R - 1} (1 - q_R)^{\beta_R - 1}, \quad 0 < q_R < 1; \alpha_R, \beta_R > 0. \quad (7)$$

上式 (6) 與首次購買片數的機率模型 (1) 有點不同：首次購買模型沒有買 0 片的情形。依據上面假設可以得到顧客重複購買片數 $R_{w|i}$ 的機率分配 $P(R_{w|i} = x)$ 為

$$\begin{aligned} P(R_{w|i} = x) &= \delta_{x=0} \pi_{w|i} \\ &+ (1 - \pi_{w|i}) \int_0^1 P(R_{w|i} = x | q_R) g(q_R) dq_R \\ &= \delta_{x=0} \pi_{w|i} + (1 - \pi_{w|i}) \times \frac{B(\alpha_R + 1, \beta_R + x)}{B(\alpha_R, \beta_R)}, \end{aligned} \quad (8)$$

其中 $x=0$ ， $\delta_{x=0}$ 為 1，否則為 0。而重複購買片數的期望值為：

$$E(R_{w|i}) = \gamma(w-i)^\delta \frac{\beta_R}{\alpha_R - 1}. \quad (9)$$

綜合五個假設，模型需估計的參數共有六個： α_T 、 β_T 、 γ 、 δ 、 α_R 、 β_R 。

以上乃是 Fader 與 Hardie (2001) 利用 Beta-Geometric 模型來預測該網站的線上消費者首次和重複購買行為，提供一種機率論思考的角度，其假設也具體刻劃了顧客將來會“不買”的行為模式。換言之，“不買” (inactive) 是一個表示線上消費者再購行為重要的概念，代表已首購後可能當下不想再購買，不一定是完全退出市場。

當吾人比較因果決定論的研究設計，觀察“不買” (inactive) 的意義，會發生一種因果推論的時序疑問是：回溯性研究和前瞻性研究思考是不同的嗎？例如如何靖遠與賴

宜楓(2012)一文中蒐集國內 Yahoo! 奇摩拍賣中 T 恤類別商品在 2011 年 3 月份的交易資料，回溯買賣雙方的歷史交易資料。以邏輯迴歸分析(logistic regression)的依變數為二元變數，其值為「1」時表示該交易為「再購」；否則其值為「0」表示該交易是「非再購」。邏輯迴歸分析是適用在依變數為二元類別資料，主要的目的是為了找出依變數與一組自變數之間的統計關係，也可以克服依變數必須服從常態分配的假設，且可以估計依變數發生的再購機率。自變數包括兩次交易的時間間隔，雙方累積的交易次數和交易金額，以及賣家的總評價等，該模型如下：

其中， X 為自變數， β 為待估計參數， $p(x)$ 表示自變數取值 x 時的再購機率。

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \quad (10)$$

被視為「再購」的交易表示該交易的買方和賣方在過去一段時間內有成交的紀錄。若是沒有歷史的成交記錄，可以斷定該買家和賣家是頭一次在線上交易。但是，這些頭一次在線上交易的人，事實上也可能會在非統計期間的未來重購。職是之故，根據買方和賣方的成交的紀錄，進行回溯性研究或前瞻性研究，所探討的線上消費者再購行為意義是不同的。如此，前者回溯性研究在邏輯上會變成也可用以探討「非再購」的緣由，此處「非再購」並非不再購買，而是代表資料期間只觀測到一次購買(包含回溯統計時間內未發現二次以上購買所歸類的首購、未來會永久退出買方角色的首購、以及非統計期間的未來可能發生重購的首購)；後者前瞻性研究才是探討首購後將來隨時間變化可能“不買”(inactive)的情形。是以，**傳統迴歸因子模型的確較不容易掌握電子商務“不買”(inactive)的意義。**

另一方面，較值得一提的是，何靖遠與賴宜楓(2012)在電子商務學報的論文，有另一貢獻應是該文的因果推論中，將再購行為結合行銷領域的 RFM 模型進行論述。機率論的重要文獻也有連結 RFM 模型建構線上消費者再購行為的論文，列舉其中一則是 Fader 與 Hardie 等(2005b)的 Journal of Marketing Research 的年度最佳論文所勾勒 RFM 和 CLV 之間清楚的數學函數關係架構；二則是 Fader 與 Hardie 等(2005a)改進 Pareto/NBD 模型，藉由基本假設上的小幅變動，而得到了一個新模型，稱為 BG/NBD 模型。

Pareto/NBD 模型是 Schmittlein、Morrison 和 Colombo(1987)建立，也稱為 SMC 模型，SMC 模型可求出個別客戶的活躍性機率及預測購買次數。該模型定義客戶仍有購買力時稱為活躍(active)，如果客戶跑掉不再購買稱為流失(dropout)，線上顧

客歷史交易資料包含從定義時間 0 到一段時間 T 的購買次數為 X ，而最後一次購買時間點為 t_x ，即 $0 < t_x < T$ ，因此 Pareto/NBD 模型僅需要利用個別客戶的交易資訊為 (X, t_x, T) 。

BG/NBD 模型與 Pareto/NBD 模型唯一的差別在於客戶何時與如何流失的假設不同。Pareto/NBD 假設客戶的流失可以發生在任何時間點，於實際交易情形無關。但 Fader 等人認為客戶的流失應該緊接在購買行為之後。BG/NBD 模型建立於五個基本假設如後，其中前兩個假設與 Pareto/NBD 模型的假設一樣，Fader 等學者於 2005 年提出 BG/NBD 模型建構，也同步公布使用工具簡便的套裝軟體(Excel)工作表²即可建立 BG/NBD 模型的概似函數、求解最適參數、以及計算個別顧客未來購買次數。模型有以下前提假設：(1)當顧客具備活躍性時，其顧客交易量會隨著交易率 λ 而呈現卜松(Poisson)分配，也就是在兩次的交易時間當中，交易率 λ 服從指數(Exponential)分配；(2)顧客之間交易率 λ 的異質性設為伽瑪(Gamma)分配；(3)在每次交易過後，顧客會流失的機率為 p ，因此顧客歷時多久會退出的時間點假設為以 p 為參數的幾何(Geometric)分配；(4) 每人的流失的機率 p 的異質性設為貝他(Beta)分配；(5)顧客交易率 λ 與退出率 p 在顧客之間獨立。BG/NBD 模型只是將 Pareto/NBD 客戶流失模型部分由指數分配轉換成幾何分配，Fader 與 Hardie 等(2005a)證明 BG/NBD 模型與 Pareto/NBD 模型在預測顧客未來的消費行為的適用性及預測能力，兩者的效力相同。但 BG/NBD 機率模型容易計算的特性更適合實際應用。

參、研究方法

在首節介紹中，李維特「如何做好實徵研究」該課程課綱所提到的第三個重點是引用新資料(Find new data)。要言之，在談到實徵研究方法中的基石，就是必須存在至少一組符合研究旨趣的好資料可供驗證。例如：何靖遠與賴宜楓(2012)選擇國內最具代表性的網站—Yahoo! 奇摩拍賣平台，以服飾商品中的「T 恤」當作關鍵字，進行搜尋相關商品的交易資料。透過網頁中所揭露的成功交易為研究對象，找出所有再購的交易行為。本章節將舉出筆者整理幾類可供學界先進同儕，後續研究線上消費者行為議題的實徵研究資料，除介紹其重要的個別變項意義，也對變項之間關聯性的結構問題提出說明。

²網址: <http://www.brucehardie.com/notes/004/>

一、美國 CDNOW Master Data Set³

Fader 等學者於 2013 年 10 月釋出了前節所提 Fader 與 Hardie 等(2005a; 2005b)以及 Fader 與 Hardie (2001)多篇使用機率論模型研究線上消費者行為的原始資料集。CDNOW 原是美國一家線上光碟片銷售網站，成立於 1994 年，目前已由 Amazon.com 併購。在它初營運的五年間，吸引至少全世界七十萬人上網購買(Fader & Hardie, 2001)，這組 CDNOW Master Data Set 包括所有 1997 年第一季在 CDNOW 首購後，直到 1998 年六月底所有線上消費者再購紀錄，合計 23,570 人，69,659 人次/筆數，共四大變數：顧客編號(the customer's ID)，交易日期(the date of the transaction)，購賣數量 (the number of CDs purchased)，以及所花金額(the dollar value of the transaction)。例如第一筆資料為編號 00001，19970101 首購，購買 1 片 CD，共 11.77 元美金。廠商只要將顧客歷史交易資料(individual-level transaction)中的交易次數、最後一次交易時間及交易時間長度輸入 Excel 試算表 (Fader et al., 2008)，應用最大概似法透過 Excel 「規劃求解」功能便可以輕鬆、快速地取得 BG/NBD 模型的四個重要參數值，求出相關交易期望值預測，進而作為客戶關係管理的重要參考依據。舉例來說以週為單位，假設觀察期間為 30 週所觀察到的交易情形，某顧客重複交易次數為 4 次，表示總購買次數為 5 次；重複交易的最後一次購買時間為 25.43 週；交易時間長度定義為 28.4286 週，表示第一次購買時間在 1.5714 週，即 $(30 - 28.4286)$ 。

二、美國 CDNOW Sample Data Set⁴

本組資料乃是將 CDNOW Master Data Set 中 23570 人以系統抽樣抽出 1/10 的樣本而構成 6919 筆資料，增加了一個變數即顧客新編號(從 1 到 2357)，Fader 與 Hardie (2008)利用本組抽樣的銷售資料為基礎，建構 Excel 試算表，輸入原始資料後即可計算出每人 R、F、M 的統計，此資料較不龐雜，也適合後續研究線上消費者行為議題的實徵研究據以參酌使用。

三、台版類 CDNOW 資料集

筆者自行建構的一資料庫為國內某線上銷售日本動畫及影音光碟網站之 2006 年 1 月到 2007 年 2 月的完整歷史交易資料，我們可以把它視為驗證 CDNOW 模型的參

³網址: <http://www.brucehardie.com/datasets/>

⁴網址: <http://www.brucehardie.com/datasets/>

考資料，其中總共有 1341 位消費者，交易筆數共 2853 筆。原始資料包含客戶 ID、消費日期、消費金額、郵遞區號、單筆訂購片數等資料，亦表示一般電子商務網站的顧客消費資料大多和 Fader 等學者釋出的 CDNOW Data Set 資料型態相似。

表 1 係單一銷售網站常見的新舊客戶混雜的資料庫型態，和顧客歷史交易的資料(individual-level transaction)不同，已經沒有顧客編號(the customer's ID)。換言之，第二週購買光碟片的客戶可能是新客戶也可能是第一週的舊客戶。此一總體層次資料(aggregate-level data)，為一般電子商務業主所容易紀錄的資料型態，資料已經失去個別顧客歷史交易(individual-level transaction)的訊息，卻可利用 Beta-Geometric 模型拆解總體層次資料，預估該網站的線上消費者首次和重複購買行為，這提示了我們面對總體層次資料，利用機率論解構資料內涵的可能性，故表 1 的特殊資料結構殊值注意。

本論文下一節資料分析將只用到客戶 ID、消費日期時間及單筆訂購片數的資料。使用者 ID 用來作為辨認新舊客戶及重複消費之依據，消費時間取其中的月份，將資料分成 2006 年 1 月到 2006 年 12 月共 2396 筆交易及 2007 年 1、2 月份共 457 筆交易兩部分，利用前者建立模型及求出相關參數，並利用後者加以驗證。

由於本研究的資料沒有 CDNOW 交易資料的龐大，因此不用週為單位而改用月份為單位計算，將 2006 年全年資料依據月份分成 12 個月，並統計每位消費者每月購買片數，由於購買片數超過 10 片以上的交易次數大約佔全部交易次數的 9%，為增加分析之準確性，因此將購買片數分為 20 組，依據每月數量分成 1-19 片及購買 20 片以上的 20+ 共 20 組，使得第 20 組之交易數僅佔全部交易的 1%。

表 1、台版類 CDNOW 資料集購買片數及月份分佈表

數量	月份												總計
	1	2	3	4	5	6	7	8	9	10	11	12	
0	0	78	215	290	409	511	598	694	767	842	905	914	
1	19	44	44	52	45	48	56	63	42	35	47	62	557
2	13	36	42	49	32	37	34	34	38	42	40	48	445
3	15	34	21	41	27	39	26	32	24	27	32	32	350
4	16	23	20	20	28	32	20	25	23	25	17	36	285
5	11	16	8	14	15	17	20	21	18	15	24	27	206
6	4	12	6	16	20	11	11	11	19	10	9	20	149
7	4	5	8	5	3	3	13	9	13	10	9	11	92
8	4	4	4	4	7	4	9	7	6	6	2	10	64

	月份												
數量	1	2	3	4	5	6	7	8	9	10	11	12	總計
9	2	3	1	4	3	5	5	2	4	2	7	5	43
10	4	3	6	3	1	3	2	2	2	6	5	6	43
11	3	3	4	2	1	2	5	2	4	1	1	4	32
12	1	3	2	2	3	0	2	3	1	4	3	2	30
13	1	4	3	0	2	1	1	0	3	2	1	4	22
14	1	1	0	3	0	2	1	1	3	4	0	1	17
15	0	0	3	1	2	1	4	0	0	1	0	2	14
16	0	1	1	0	1	1	0	0	0	5	0	0	9
17	0	0	0	1	0	1	0	0	2	1	0	0	5
18	0	0	1	0	0	0	1	0	1	0	1	0	4
19	0	0	0	0	0	0	2	0	1	0	0	0	3
20+	0	1	0	1	1	2	3	3	3	5	1	6	26
總銷售量	423	757	690	792	750	797	958	787	949	984	755	1199	
單月購買人數	98	193	174	218	191	209	215	215	207	201	199	276	
新增購買人數	98	173	118	119	92	120	93	96	65	69	61	86	
重複購買人數	0	20	56	99	99	89	122	119	142	132	138	190	
累計購買人數	98	271	389	508	600	720	813	909	974	1043	1104	1190	

資料來源：本研究整理

在下一節，筆者將要逐步利用此一國內地區線上光碟銷售網站的實際交易資料，透過分析及整理後，應用 Microsoft 的試算表軟體 Excel，建立完整銷售預測模型，並透過 Excel 本身的規劃求解功能，得到銷售模型需要的參數。並且，再一次檢視本土類似 CDNOW Data Set 資料型態，是否也可以得到準確的再購預測。同時藉此一實徵研究結果，提供當今電子商務企業一個不同於問卷調查、點閱率、網站流量的評估方法的佐證。礙於資料取得的限制，本研究只有利用 2007 年 1、2 月，兩個月的資料作為驗證資料，只能使用模型驗證短期的預測效果。如取得更多月份的交易資料，可以更進一步驗證該模型中長期的預測能力。

肆、資料分析

在本節筆者將利用 Fader 與 Hardie (2001) 根據美國線上光碟銷售網站 CDNOW 的客戶交易資料所建構的 Excel 試算 5，利用 Beta-Geometric 模型來解釋和預測台版 CDNOW 資料集的線上消費者首次和重複購買行為。

一、台版類 CDNOW 模擬資料中第一個月份首次購買的機率模型分析

首先根據本文第二節列出的首次購買機率公式 (3) 僅試第一個月份，可以得知一月首次購買 0 片到 19 的機率，而購買 20 片以上的機率 (儲存格 C24) 為 1-(購買 0 ~19 片的機率總和)。至此，我們已經利用 EXCEL 將一月份購買 x 片數之機率算出。接著再將購買 x 片之人數 (儲存格 B5 到 B24) 乘上該購買 x 片機率 (儲存格 C5 到 C24)，取對數可得到 D5 到 D24 格之個別對數概似值，對數概似值總和放在儲存格 D2。再利用 Excel 中的規劃求解功能，使得對數概似函數最大化，經過計算後解得 α_r, β_r 值，分別放在 B1 到 B2，故此時的 α_r, β_r 值稱為最大概似法估計值。

	A	B	C	D	E	F	G	H	I
1	$\alpha_r =$	2134661511	LL=	-228.97			期望值	4.33	
2	$\beta_r =$	7108427796					$E(T_1) = \frac{\alpha_r + \beta_r - 1}{\alpha_r - 1}$		
3							$\alpha_r - 1$		
4	數量	人數	$P(T_1 = x)$	個別LL	$E(n_{1x})$	Chi-Square			
5	1	19	0.2309	-27.85	22.63	0.5831			
6	2	13	0.1776	-22.47	17.41	1.1152	98*C5		
7	3	15	0.1366	-29.86	13.39	0.19461			
8	4	16	0.1050	-36.05	10.29	3.16208			
9	5	11	0.0808	-27.68	7.92	1.20052			
10	6	4	0.0621	-11.11	6.09	0.71648			
11	7	4	0.0478	-12.16	4.68	0.09947	$=(B10-E10)^2/E10$		
12	8	4	0.0367	-13.21	3.60	0.04419			
13	9	2	0.0283	-7.13	2.77	0.21377	$\frac{[n_{1x} - E(n_{1x})]^2}{E(n_{1x})}$		
14	10	4	0.0217	-15.32	2.13	1.64216			
15	11	3	0.0167	-12.27	1.64	1.13261			
16	12	1	0.0129	-4.35	1.26	0.05353			
17	13	1	0.0099	-4.62	0.97	0.00101			
18	14	1	0.0076	-4.88	0.75	0.08726			
19	15	0	0.0058	0.00	0.57	0.57296			
20	16	0	0.0045	0.00	0.44	0.44064	$\chi^2 = \sum_{x=1}^{20} \frac{[n_{1x} - E(n_{1x})]^2}{E(n_{1x})}$		
21	17	0	0.0035	0.00	0.34	0.33888			
22	18	0	0.0027	0.00	0.26	0.26061			
23	19	0	0.0020	0.00	0.20	0.20043			
24	20+	0	0.0068	0.00	0.67	0.66742			
25	總消費人數	98			χ^2	12.7269	$=CHIINV(0.05,F27)$		
26					Critical value	27.59			
27					自由度	17			
28									

圖 2、一月份首次購買機率計算結果

資料來源：本研究整理

⁵ <http://www.brucehardie.com/notes/002/>

為了檢驗這個模型的適用性，我們利用卡方檢定（Chi-Square test）檢驗。首先算出期望的交易人數，即一月份總購買人數 98 乘上各購買片數機率，存於儲存格 E5 到 E24 之中，其個別 Chi-square 值為儲存格 F5 到 F24，由於總和（儲存格 F25）=12.7 小於 5% 的顯著水準下的臨界值 27.59，證明這份資料符合 Beta-Geometric 分配。

二、台版類 CDNOW 資料中完整模型的機率分析

根據本文第二節所列出 Fader 與 Hardie (2001) 所提的公式，完整模型需估計的參數共有六個：分別是代表首購的 α_T 、 β_T 和代表重購的 γ 、 δ 、 α_R 、 β_R 。利用其 Excel 試算整理執行規劃求解，六個一起估計得到參數值為：

表 2、整模型規劃求解結果

變數	α_T	β_T	α_R	β_R	γ	δ
值	62166.96	172145.1	9.716096	29.98394	0.302316	-0.28408

資料來源：本研究整理

其中 $\delta < 0$ 表示隨時間增加，重複購買人數減少，根據（5）式計算結果，顯示新顧客在消費後，隔月不會馬上回來重複購買的比率約 70%。而利用（4）式得到首次購買片數期望值為 3.77 片。由（8）式得到的重複購買片數機率計算可知，即使在可能重複購買的客戶中，還是估計有 24.5% 的人最後完全沒有購買。

表 3、重複購買片數機率表

重複購買數量	0	1	2	3	4	5	6	7	8	9	10
概率	24.5%	18.0%	13.4%	10.0%	7.6%	5.8%	4.4%	3.4%	2.6%	2.1%	1.6%
重複購買數量		11	12	13	14	15	16	17	18	19	20+
概率		1.3%	1.0%	0.8%	0.6%	0.5%	0.4%	0.3%	0.3%	0.2%	1.1%

資料來源：本研究整理

根據圖 3 單次購買數量分佈情形，乃利用第二節所列出 Fader 與 Hardie (2001) 所提的公式（4）和（9）合成，可以發現模型預測購買人數隨單次購買片數增加而逐步遞減。而對應實際銷售情形可以發現單次購買數量大致符合模型分配，較值得注意的部分在於單次購買一片及七到九片的次數低於模型預測，而購買三到六片的次數又高於模型預測，推測可能因為交易成本造成之差距，當消費者購買時除了要支付該光碟的購買成本外，還要支付運費成本，商品單價平均為 300 元，單次交易未滿 900 元時須加上運費 80 元，如果只買一片光碟則運費相當於單片光碟購買成本的四分之一，因此購買一片的購買次數少於模型預測。而由於購買金額滿 900 元即免加運費之

優惠，可以發現單次購買數量超過 3 片（約 900 元）時，客戶的購買行為受免加運費鼓勵，使得購買次數高於模型預測。至於到了七片以上時，金額大約超過 2000 元，購買行為又趨向保守，可能 2000 元為該公司顧客在網路購物時的一個心理門檻。

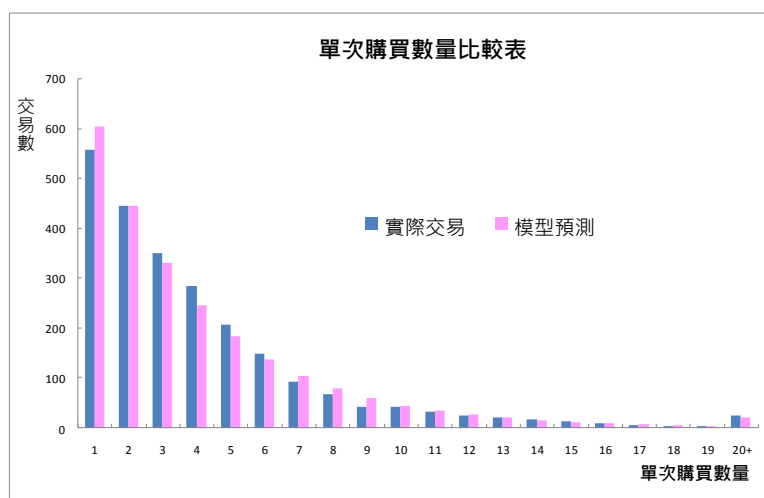


圖 3、每筆交易購買數量分佈

資料來源：本研究整理

由圖 4 虛線為模型預估值、實線為實際銷售量可以看出來，新顧客的交易量有下滑的趨勢，表示新客戶的加入有減緩之趨勢。而舊客戶的銷售量呈現穩定成長，比較模型預測及實際銷售數字可以發現，模型能夠掌握實際交易量的變化，提供穩定的趨勢預測。尤其是針對新客戶購買量的預測績效優良，虛線和實線契合。另外發現實際銷售情形在二、七、十二月的銷售量有較大幅度成長，推測原因應該與該網站之主要客戶群為學生及上班族，銷售額增加與寒、暑假及聖誕節之假期有關。

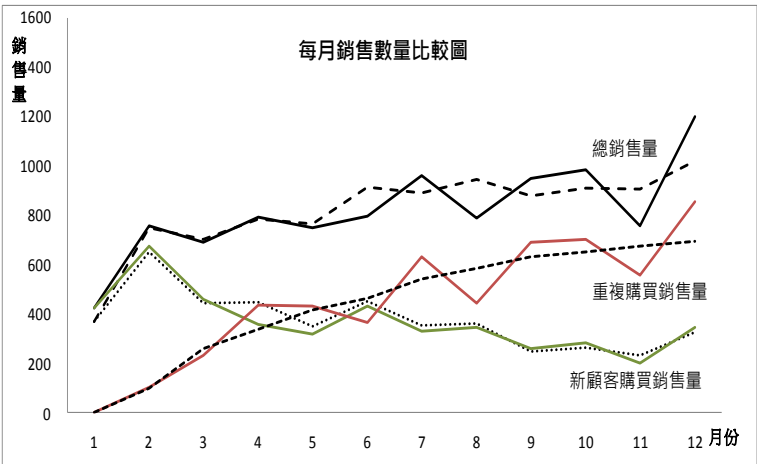


圖 4、每月銷售數量比較圖

資料來源：本研究整理

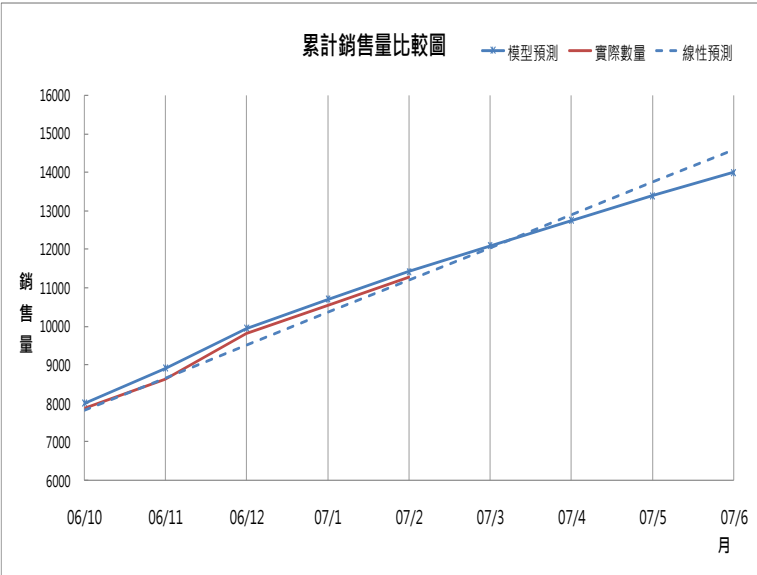


圖 5、累計銷售量模型預測與以時間為自變項的線性預測比較圖

資料來源：本研究整理

圖 5 係本研究機率分配和迴歸因子模型方法的實證比較結果釋例，由於每月銷售數量波動較大，將每月銷售量加以累計，來比較累計銷售量的預測情形。由圖 5 可以看出來在累計銷售量方面，研究模型與實際交易資料十分吻合，且多數時候都是模型預測值略高於實際銷售量，這對於預測未來銷售情形時，模型能夠充分預測，並且模型的預測值可以作為未來銷售量的上限參考值。現將前十二個月累積銷售量做迴

歸因子模型線性預測與本研究模型及實際資料加以比較，可以發現在 2007 年 1 月及 2 月，本研究模型比一般以時間為自變項的單元迴歸線性預測更能掌握實際銷售變化情形。以時間為自變項的單元迴歸線性預測模型如下：

其中， T 代表時間自變數， β 為待估計參數， ε 表示一般單元迴歸的誤差項。

$$\text{累計銷售量} = \beta_0 + \beta_1 T + \varepsilon \quad (11)$$

三、電子商務的資料型態的衍生思考:國內高鐵 T Express App 資料集

本文分別以機率論補充文獻；以國內的資料加以驗證機率分配模型在本研究主題上的適用性和重要性；並且批判何靖遠與賴宜楓(2012)一文中蒐集國內奇摩拍賣中 T 恤類別商品，以「回溯」買賣雙方的歷史交易資料計算未來再購行為可能的謬誤。

筆者手上另有一國內高鐵 T Express App 資料集說明另一種屬於電子商務交易的資料型態，該資料為國內某家商業銀行信用卡客戶從 2011 年 11 月至 2012 年 10 月曾利用國內高鐵 T Express App 應用程式購買高鐵票者。交易原始資料包含消費月份、客戶 ID 前六碼、生日、性別、居住地區、購買次數、購買金額、購買方式。國內高鐵於 2011.10.28 推出全新購票系統「國內高鐵 T Express」，免費開放下載手機 APP 應用程式，旅客透過該程式，訂票、付款、取票、搭乘一次完成。不但可以隨時隨地訂票或退票，或查詢最新列車時刻表及服務、優惠等活動訊息，還能減少紙張的浪費。旅客可透過下載該應用程式訂位搭乘，完成付款手續並取票後，會出現一組手機票證二維條碼，只要使用該條碼感應車站入口閘門，就可以進出站搭車。



圖6、國內高鐵 T Express App應用程式訂位畫面

資料來源：本研究整理

T Express App 資料集和上述 CDNOW 資料集不同之處有三，(1) T Express App 資料集是除交易資料之外，尚包括有顧客基本資料：性別、年齡以及居住地區；(2) T Express App 資料集也包含購買方式，意指曾利用國內高鐵 T Express App 應用程式購買高鐵票者，仍有臨櫃台購買票券的實際狀況發生；(3) T Express App 資料集和線上光碟銷售網站 CDNOW 的客戶交易情境是不同的，使用智慧手機是透過網際網路交易的新興電子載體媒介(electronic devices)，實徵研究的焦點應是探討智慧手機消費者對加值應用服務的購買行為。礙於上述原因，本文不以此資料作更深入的討論，在此僅作為一示例，表示完整、正確的資料來源和研究設計對實徵研究的重要，並且勾勒出後續研究線上消費者行為其他議題的實徵研究方向。

伍、結論與建議

檢視上述各種在線上消費者再購行為實徵研究的內涵可以發現，線上消費者購買行為的確具有個人異質性，亦即，若僅是依賴可觀測的異質性，例如測量會影響線上消費者再購行為的顯性因素(如 RFM 分析)等，卻往往因為言有未盡之處，無法完全掌握消費者再購行為。申言之，一般實徵研究常見焦點在於解釋自變項的效果的顯

著性，甚至誇大了(現象=必然)的決定論可解釋一切的說法，這樣容易忽略了因果關係可能侷限於某些特定的條件才會發生，而觀察現象中，除了一般關注的因子效果解釋必然性部分外，其實還有偶然的成分存在。也就是說在實徵迴歸模型設計 (現象=必然+偶然) 中，往往誤差項的殘差意義容易被研究者忽視。

另一方面，當我們選擇刻意忽視「必然」的部分，機率論(現象=偶然)則代表關注的焦點移到不可直接觀察的異質性上，模型假設可能與線上消費一段時間實際交易資料內含的行為模式有所關聯。因之，前文補充的機率論過往文獻中，由於機率分配的機制可以有效的描繪「不可看見卻永遠存在的線上消費者行為異質性」，尤其這些模型通常假設都是從機率論中最基本的二項(Binomial)、卜松(Poisson)或指數(Exponential)分配出發，加入顧客異質性的分配設定，利用估計參數來捕捉顧客購買的再購於否、再購次數、以及何時再購。據此，本研究舉出相關機率分配模型經過實證說明，的確適合用以描述線上的消費行為和作出良好的預測。

綜言之，當前主要文獻中使用機率分配模型進行線上再購預測的常見作法有二：(1) 類似 Fader 與 Hardie (2001) 根據美國線上光碟銷售網站 CDNOW 的客戶交易資料所建構的總體層次資料(aggregate-level data)如本文表 1，並利用 Beta-Geometric 模型來加以解釋和預測線上消費者首購和再購行為。本研究實證資料分析部分也利用 Fader 與 Hardie (2001) 的作法來解釋和預測台版類 CDNOW 資料集的線上消費行為。由於僅就現有總體層次資料的交易時間及數量作探討，以解構整體的觀念，來剖析現有消費群重複購買次數和機率，並沒有完全利用原始資料中的個人交易資訊；(2) Fader 與 Hardie 等(2005a; 2005b) 改進過去研究，得到的 BG/NBD 模型，則是將顧客歷史交易資料(individual-level transaction)中的交易次數、最後一次交易時間及交易時間長度輸入 Excel 試算表，應用最大概似法透過 Excel「規劃求解」功能便可以取得 BG/NBD 的重要參數值，求得單一消費者可能重複購買的次數和機率。而且再加入金額的考量後，可以和 CLV 產生連結。

值得一提的是，本篇研究結論並非為了機率論可完全的取代決定論，而是由實證回答從事線上消費者再購行為的實徵研究，有另一個重要選擇，是可以應用機率論內涵來詮釋再購行為的隨機現象的內隱機制。由於實徵研究的經驗是可以累積的，學術研究像是一種傳承，除了期許有所創新跟突破，也希望後續研究者能站在學術先輩的肩膀上往前走。基於這樣的想法，本文盡可能地介紹機率論模型的相關線上消費者再購行為的研究內容，希望透過這些介紹，讓後續國內的研究者可以很快地進入整合相關的研究議題。

當然，線上消費者再購行為的研究，過去決定論也扮演了重要的角色，以「環境決定論」(environmental determinism)觀點而言，可用以強調人類活動的基礎乃建立於線上特殊的交易情境之上，如透過網際網路交易則有行動裝置和固定電腦不同之處。但是環境決定論卻不能代表身為交易環境中具有主動能力的顧客的價值、態度及異質性，甚至賣方的角色或評價等。然而「環境決定論」或是「顧客主導決定論」等思維都是容易流於偏於一隅。本文筆者整理「機率論」的論述欲擺脫片面以及因果關係的限制，而承認人類與環境相互關係的複雜性，並且提出具體可行方案來解決此一複雜關係所引發的問題，即本論文所主張的逆向思考方向，那就是應重視機率論為主體，甚至實證表示其足以做出良好的預測，再考慮加入迴歸因子的決定論為輔，對一般電子商務網站的客戶再購行為研究的可能性。

例如：前述國內高鐵 T Express App 資料集是除交易資料之外，尚包括有顧客基本資料：性別、年齡以及居住地區，這些變數可能引起的差異，吾人就可以做完機率分配模型後，事後加入決定因子的分析，據以區別再購行為的人口變項差異。另外，T Express App 資料集也包含不同購票方式，意指曾利用國內高鐵 T Express App 應用程式購買高鐵票者，仍有臨櫃台購買票券的實際狀況發生，吾人也可以嘗試利用所謂有限混合機率模型(finite mixture model)而達成拆解臨櫃台購買和使用 App 購買的不同交易行為，會在同樣的人身上發生。

本文提出線上消費者再購行為的實徵研究應進行機率分配模式為先的概念，經實證顯示其預測效果可能已遠勝於決定論的想法去挑選了過多或不足的決定因子變項。事實上，機率分配模型應用在行銷上，國際間已見不少學術論文討論，但是在國內，將機率分配模型概念應用在電子商務領域仍為少見，相信未來尚有很大的發展空間。

參考文獻

中文部份

1. 何靖遠、賴宜楓 (2012)。線上消費者再購行為的實徵研究。電子商務學報，第 14 期，第 2 卷，頁 307-328。
2. 喻奉天、蘇國璋、許孟祥、王昭文、林盈君 (2010)。不確定性與信任對線上消費者購買意願之影響，電子商務學報，第 12 期，第 3 卷，頁 431-456。

英文部份

1. Drucker, Peter. (1969). *The Age of Discontinuity*. New York: Harper & Row.
2. Hughes, A. M. (1994). *Strategic Database Marketing*. Chicago: Probus Publishing.
3. Talbott, W. (2008). Determinism. In E. N. Zalta (Ed.). *The standford encyclopedia of philosophy*. 5th ed. Stanford. CA: Stanford University.
4. Neter, Kutner, Nachtsheim, Wasserman. (1996). *Applied Linear Regression Models*. 3rd edition.
5. Ehrenberg, A. S. C. (1959). *The Pattern of Consumer Purchases*. Applied Statistics. 8, 26-41.
6. Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005a). Counting Your Customers: the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science*. 24(2), 275-284.
7. Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005b). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research*. 42(4), 415-430.
8. Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2008). Computing P(alive) Using the BG/NBD Model. From http://www.brucehardie.com/notes/021/palive_for_BGNBD.pdf
9. Fader, P. S., Hardie, B. G. S. (2001). Forecasting Repeat Sales at CDNOW: A Case Study. *Interfaces*. 31(3), 94-107.
10. Kumar, V., and Petersen, J. A. (2005). Using a Customer-Level Marketing Strategy to Enhance Firm Performance: A Review of Theoretical and Empirical Evidence. *Journal of the Academy of Marketing Science*. 33(4), 504-519.
11. Rust, R. T., Lemon, K. N., and Zeithaml, V. A. (2004). "Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*. 68, 109-127.
12. Schmittlein, D. C., Morrison, D. G., and Colombo R. (1987). Counting Your Customers: Who Are They and What Will They Do Next?. *Management Science*. 33, 1-24.

13. Yen, C.H., and Lu, H. P. (2008). Effects of e-service quality on loyalty intention: An empirical study in online auction. *Managing Service Quality*. 18(2), 127-146.