

應用函數映射技術擴增產業創新資料庫之研究

江志民^{*1} 郭博文² 丁台怡³

^{*1} 經國管理暨健康學院助理教授

² 經國管理暨健康學院副教授

³ 創曜產業顧問有限公司副總經理

摘要

產業創新活動調查可幫助我們了解產業界中新知識的創造以及現有知識的新應用，尤其是我國中小企業在研究發展以外的創新活動是相當活躍的。Knell 與 Nas (2006) 及 Raymond (2006) 等人指出在研究創新過程以及「投入與產出」會有一些限制，即在收集創新調查資料時，調查資料時間點是一項根本的問題，必須以長期追蹤資料 (panel data) 才能完整呈現「投入與產出」的關聯性。

目前歐盟 CIS 調查已進行到第五次，我國創新調查也已進行兩次，由於調查樣本的不同，或調查問卷中問項題目會隨著不同梯次進行增修，造成了不同梯次的創新調查所得的資料並不相同，各梯次的資料庫無法直接串聯合併成長期追蹤調查資料庫。本研究嘗試以資料庫的函數映射 (functional mapping) 技術進行產業創新資料庫中變數與資料擴增，期能達成建構長期追蹤的創新資料庫之目標。研究結果顯示資料庫映射結果的相對誤差率 RER 維持在 1 上下，並沒有造成太多的誤差，結果令人滿意且穩定。表示利用資料庫函數映射技術確能將原有的資料庫不足之處加以擴增，解決資料短缺問題，形成長期追蹤資料，提升資料庫的價值。

關鍵詞：創新調查、函數映射、資料庫擴增

* 通訊作者

Research on Using Functional Mapping Technology to Expand Innovation Survey Database

Chiang, Chih-Ming^{*1} Kuo, Po-Wen² Ting, Tai-I³

^{*1}Assistant Professor, Ching Kuo Institute of Management and Health

²Associate Professor, Ching Kuo Institute of Management and Health

³Deputy General Manager, InnoValue Consult Incorporation

Abstract

Innovation is the key of survival for small and median enterprise (SME). It was important to understand the relation between innovation activities. However, Knell and Nas (2006) and Raymond et al(2006) indicated the limitation on studying the innovation process, such as the “timing” is the essential problem as collecting data. It was impossible to draw a whole picture about the relations of input and output without having panel data.

Nowadays Community Innovation Survey (CIS) had been conducted the fifth time in Europe and the second time in Taiwan. The questionnaire using in Taiwan Industry Innovation Survey had been modified 2 times. The contents and variables were not the same within these 2 databases. As the consequence, the 2 databases cannot be combined as a panel database. This study was trying to use functional mapping technology to expand the data and variables and to build a panel database for Taiwan innovation survey. According satisfied outcomes of this research, RER varied around 1. It meant using functional mapping technology to expand Taiwan innovation survey database was practicable.

Key Words: innovation survey, functional mapping, database expansion

* Corresponding Author

壹、前言

就知識經濟發展而言，須仰賴技術的生產、流通與應用，這些知識包括新知識的創造以及現有知識的新應用。在過去幾年間 OECD 認為早期我們重視的研發活動 (R&D activities)，在我國研發活動的調查與資料分析舉辦了二十多年，對於研發活動的調查與資料分析已相當清楚；但是技術及資訊的流動在個人與個人間、企業間、學術及研究機構間及彼此間的動態創新過程，對於那些幫助或阻礙技術創新的因素，我們瞭解得並不十分的清楚。因此歐盟自從 1991 年開始發展 Community Innovation Survey (CIS)，其目的在瞭解及掌握業界的創新活動，截至目前為止，歐聯已做過四次 CIS，而且根據 Oslo Manual (OECD 對創新調查所建議的規範)，對收集得來的 technological data 作分析與處理，以做國際比較。

我國產業結構以中小企業為主，員工人數在 200 人以下之企業約佔全體企業的 98% 以上。根據國科會執行科技動態調查之經驗，我國產業界研發活動主要集中於大型企業，以國內研發經費在前 300 名之大企業為例，其研發經費即佔產業整體研發經費 70% 以上。以我國中小企業之活躍情況推論，我國中小企業雖然沒有太多的研發活動，仍然能夠推出許多新產品，顯示我國中小企業在研究發展以外的其他創新活動應該是相當活躍的。因此，在 2001 年我國第六次全國科技會議中，國科會與經濟部不約而同地在第一、三議題各自提出了辦理全國「技術創新調查」之建議。隨後召開的經濟發展諮詢委員會也對「研擬與國際接軌之知識經濟指標，並定期評估。」提出建言。國科會與經濟部乃共同出資籌辦了台灣地區第一次技術創新調查 (Taiwan Technological Innovation Survey, 以下簡稱 TTIS)，歷經兩年的調查與分析，對於我國的技術創新已有相當程度的瞭解。

鑑於技術創新調查對於知識經濟的重要性，目前我國由國科會贊助下已完成台灣地區第一次的技術創新調查 TTIS1 與第二次的產業創新調查 TIS(2004~2006)。從創新調查的分析結果發現，台灣的產業發展形態有其特殊性，因此除需持續進行產業創新調查與觀察之外，並需要透過多次的產業創新調查，長期追蹤並蒐集產業創新相關資料，建立國內外產業之技術創新 panel data 統計資料庫與資訊平台，才能顯現技術創新調查之實質意義及發展趨勢。

欲了解創新活動的過程與演進，必須建立創新活動中「投入與產出」間的關聯性，但在研究創新過程以及「投入與產出」會有一些限制，Knell 與 Nas (2006) 及 Raymond 等人(2006)指出在收集創新調查資料時，調查資料時間點 (timing problem) 是一項根本的問題。目前創新調查 CIS 的設計，投入部分是收集過去三年公司內部對於創新活動的努力，而產出部分則是收集最後一年的創新成果。若要依此分析「投入與產出」的關聯性則必須有很強的假設：「過去三年的創新活動需為每年持續並相似的投入，才導致最後一年的成果。」但這項假設可能會因不同產業會有不同的狀況，故必須以長期追蹤資料 (panel data) 才能完整呈現「投入與產出」的關聯性。

目前國內所進行兩梯次創新調查樣本大部分係經由抽樣而來，另針對 500 大的少部分企業進行普查，使得調查樣本並非都是相同的企業或廠商，如第一次的技術創新調查

TTIS1 的有效樣本為 3,356，而第二次的創新調查 TIS(2004~2006)的有效樣本為 10,017。故所建立的創新調查資料庫中每一梯次受調查的廠商不盡相同，在建置關聯式資料庫時無法直接串聯，以致造成大部分廠商在某些梯次的調查之中是屬於無反應的狀況 (unit non-response)。若欲改善此項問題，則必須完整進行長期追蹤的研究不可；然實際上可能無法達成，因為長期追蹤的研究必須要花費龐大的人力與物力，且還需要廠商完整的配合才可行，使得現階段的 CIS 調查與我國的創新調查仍無法真正進行長期追蹤的研究。

另一方面，目前歐盟的 CIS 調查已進行到第四次，我國的創新調查也已進行兩次，調查問卷內的問項題目會隨著不同梯次進行增修。倘若 CIS 調查或我國創新調查可進行長期追蹤研究，但每一次的問卷內容及部分問項題目皆有可能會不同，如 CIS 到了第四次才增列了行銷創新及組織創新的相關問項，這也造成了不同梯次的 CIS 調查所得的資料並不相同，亦增加了分析的困難度。事實上，我國第二次的創新調查 TIS(2004~2006)增加了許多第一次的技術創新調查 TTIS1 所沒有的問項，比如「創新類型」是屬於 TIS(2004~2006)新增加的。

上述的問題與一般常見的缺失值不盡相同，缺失值指的是少部分或某些人資料點無法取得，而本研究的問題則是指在分析的過程中少了某些變數，即缺少整個變數的資料點，如在 TTIS1 資料庫缺少了創新類型的變數。為解決此類型的問題，故本研究嘗試以資料庫函數映射 (functional mapping) 的方法，針對創新調查資料庫中因問卷題項增修後所造成的資料及變數短缺的問題，進行資料庫的變數與資料擴增，以建立完整的創新調查長期追蹤研究資料庫。

貳、文獻查證

一、創新指標的發展應用

國際間對於創新調查的推展是以歐盟為首，歐盟自從 1991 年開始發展 Community Innovation Survey (CIS)，其目的在瞭解及掌握產業界的創新活動，截至目前為止，歐盟已做過四次 CIS，而且根據 Oslo Manual 2005 (OECD 對創新調查所建議的規範)，對收集得來的 technological data 作分析與處理，以做國際比較。內容主要涵蓋：(1) 非技術創新及其影響、(2) 技術創新及其影響、(3) 創新的資訊來源、(4) 阻礙創新活動的因素、(5) 創新對企業績效的影響、(6) 創新的擴大應用、(7) 特殊問題、(8) 企業基本資料。

對於創新調查資料之應用的部份，歐盟國家創新調查由歐盟統計局 (EUROSTAT) 統籌各國的調查資料，集結成「創新計分板」(Innovation Scoreboard) 中，也將美、日兩國各項創新相關表現與歐盟各國作比較。近年來，OECD 在會員國部長級會議之要求下，也著手進行知識經濟相關指標之研究，出版了「科學、技術與工業計分板」(OECD Science, Technology and Industry Scoreboard)，研擬了知識創造與流通 (The creation and diffusion of knowledge)、資訊經濟(The information Economy)、全球化整合(The global Integration of Economic Activity)、生產力表現(Economic Structure and productivity)等四類指標以瞭解其會員國間之差異。

周彥宏(2003)研究結果發現，技術創新績效指標模式的系統構面之重要性依序為：「產品創新」、「純粹技術創新」及「制程創新」；而三個系統構面下的主要構面皆以「整體績效構面」最為重要；在指標方面則以「產品創新構面」中的「顧客對此項產品的滿意程度」最為重要，其次依序為：「與公司過去類似的技術創新活動比較具有相對的優勢」、「與競爭者發展的新產品比較，新產品的品質水準」、「新技術對公司獲利的貢獻程度」、「技術創新成果與原定技術目標的符合程度」、「產品能符合市場潮流或滿足顧客需求」。

劉孟俊，（2004,12）一、產業科技創新相關調查本研究利用因素分析 (factor analysis)，針對廠商資料進行因素萃取，以選取影響技術創新指標。其次，運用技術創新指標進行群聚分析 (cluster analysis)，對具有相似創新指標的廠商進行分類，歸納各廠商的創新模式。製造業可萃取五個創新指標，分別為「產品技術創新」、「技術創新資訊仰賴市場提供」、「制程技術創新」、「電子商務應用」、「價購外部知識與人員訓練投入」。製造業廠商與服務業廠商的創新模式均可歸納 5 類。

臺灣地區經濟部中小企業處（2007），參考 OECD Oslo manual、韓國 Inno-Biz 篩選準則及中國「中小企業技術創新測度與評估研究」相關評估準則，雖然各有不同的衡量構面及準則，但大致可歸納為七大衡量構面：創新投入、創新管理、創新策略管理、創新能力、創新實現、創新成效及知識管理。

二、產業創新型態及關聯性探討

吳思華等（2009）研究發現（1）臺灣廠商的規模越大，越容易進行各種類型的創新，但是創新的傾向會有邊際遞減（呈倒 U 形曲線）的現象；（2）市場範圍越大，進行各類型創新的傾向越高；（3）新設立公司的創新傾向普遍較低；（4）各類型創新皆具有互補關係，可見公司創新是為了完全解決問題，故通常會同時進行各類型創新，惟各類型創新間互補程度不同。

吳思華等的研究也發現會創新的廠商也會進行各種類型的技術合作創新，這種產官學研之間的連結是國家創新系統最關鍵的架構。本研究發現臺灣的國家創新系統是雙核心系統。主要核心的產業內上下游之間的合作研發，這顯示臺灣產業網路的創新活力。第二個核心則是產學與產研的合作研發。政府的補助會引導廠商與大學、研究機構以及競爭者一起合作創新，是連結這兩個核心的關鍵力量。

林志維（2002）研究主要在於探討「競爭策略」、「技術環境」、「技術創新」與「創新績效」四者間的關係，研究以傳統化工業與高科技業為調查物件，結果證實公司的管理者能針對自己產業的技術環境與競爭策略，將公司的創新重心著重在不同的技術創新上，而這些不同的創新也影響不同的創新績效。

李宜穗（2010）研究臺灣地區產業創新調查後發現，1.近年企業創新活動有大幅改變，從傳統的技術創新（含產品創新、制程創新），逐漸演變成非技術創新（行銷創新及組織創新）。2. 創新活動阻礙因素，已從原先的成本及知識因素演變為市場因素。3. 創新保護措施方面，仍以「比競爭者搶先推出產品快速進入市場」為第一考慮。

本研究整理出目前國內外對於創新調查應用之相關研究，這些研究通常都屬於橫斷面的調查研究，較少有縱貫性的研究，故本研究即是欲克服創新調查資料庫無法串連之不足之處。

參、研究方法

根據上述的研究目的，即「如何針對研究中的資料庫進行實際資料與變數的擴增」，本研究嘗試以資料庫的函數映射（functional mapping）方法進行資料庫中變數與資料擴增，期能達成建構長期追蹤的創新資料庫之目標。

一、材料來源

本研究主要研究為建立不同梯次之創新調查長期追蹤研究資料庫，所使用的為兩次創新調查資料庫，分別為第一次的技術創新調查 TTIS1 的有效樣本為 3,356，調查期間為民國 90 年 8 月 1 日至 91 年 7 月 31 日；及第二次的創新調查 TIS(2004~2006)的有效樣本為 10,017(以下簡稱 TTIS2)，調查期間為民國 96 年 8 月 1 日至 96 年 10 月 31 日。

二、資料庫函數映射

所謂「函數映射」主要是指當兩個資料庫之間無法直接串聯時，如我國創新調查 TTIS1 與 TTIS2 兩個資料庫，所調查的企業廠商不盡相同時，此時首先必須建立兩資料庫之間的虛擬聯結關係。接著，於兩資料庫中抽取出等量且具有代表性的預測樣本（predictive samples）。將此兩樣本資料配對轉換，建立兩資料庫間的虛擬聯結以及預測樣本資料庫，以此預測資料庫建立適當的加值預測函數。透過此預測函數即能將輔助資料庫的加值訊息映射至主資料庫之中，此過程稱為「資料庫函數映射」。（江志民，2006）

「資料庫函數映射」主題之假設前提為：存在主資料庫（main database 或 target database）與輔助資料庫（auxiliary database 或 assistant database），主資料庫為有興趣研究且欲被加值的資料庫，輔助資料庫則是提供加值訊息的資料庫，且兩者皆有相同或相似定義的基本變數，但不具有可對應之關聯關係。因此在「資料庫函數映射」過程中首要解決的問題，即是「如何建立主資料庫與輔助資料庫兩者之間的關聯性」。也就是說，如何將兩個有關聯但卻無法直接合併的資料庫間，建立虛擬聯結關係，本研究稱為預測資料庫（predictive database）。本研究的想法係以統計之抽樣方法，在兩資料庫之中抽出等量且具有代表性的樣本進行一對一的配對。

當建立了預測資料庫之後，接著必須尋找適當的預測函數，透過此預測函數的函數映射（functional mapping）方式，在主資料庫中進行加值預測，使得主資料庫能夠獲得最佳的加值效果。圖 1 說明了整個資料庫函數映射之加值概念流程，假設圖中的左邊為主資料庫的關聯表內容，右邊則是輔助資料庫的關聯表內容。首先假設此兩個資料庫為同性質但不相同的資料庫。兩者分別都是以關聯表形式所建立的資料庫，但兩資料庫之間卻又不具有關聯表可供對應與合併。

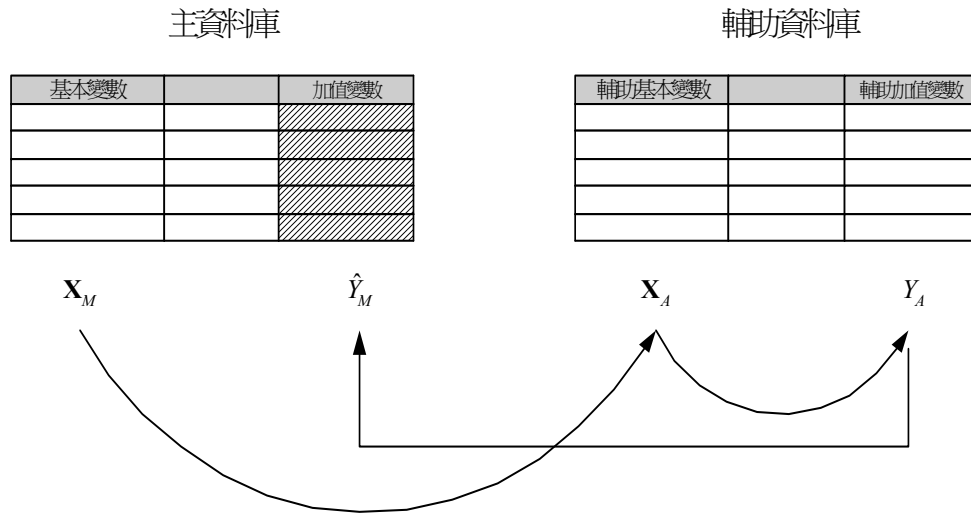


圖 1 資料庫函數映射示意圖

Figure 1 Functional mapping concept

根據上面同性質的假設，這兩個資料庫應該具有相同或相似定義的基本變數 (basic variables)，如 TTIS1 與 TTIS2 同時蒐集了各企業的各類行業別、員工人數、企業類型或所在地等等的資料。總而言之，只要兩個資料庫的來源不同，但卻又具有相似的資料內容，而兩者之間並不具有關聯表可供串聯與合併的特性。可將此兩個資料庫區分出欲進行加值的資料庫為「主資料庫」(main database)，以 Ω_M 表示之，即我們有興趣研究的資料庫；另一為提供加值輔助訊息的資料庫為「輔助資料庫」(auxiliary database)，以 Ω_A 表示之，僅提供訊息，並不會改變或增減原有資料庫之內容。

令主資料庫的基本變數為 \mathbf{X}_M ，輔助資料庫中的輔助基本變數 (auxiliary basic variables) 為 \mathbf{X}_A ；另外，輔助資料庫中的輔助加值變數 (auxiliary value_added variable) 為 Y_A ，而主資料庫中的被加值變數 (value_added variable) 為 \hat{Y}_M ，也就是加值過程中最終產出的資料變數。

函數映射的加值過程中，首先必須先找到兩個資料庫的虛擬聯結關係，此部分是以挑選兩個資料庫中適當的基本變數 \mathbf{X}_M 與 \mathbf{X}_A 作為對應關係。接著在兩個資料庫中，抽出具有代表性的樣本，進行排序與配對。配對完成後，若 \mathbf{X}_M 與 \mathbf{X}_A 的分布不一致時，可能需要透過轉換函數 $T(\cdot)$ ，能夠將 \mathbf{X}_M 映射至輔助資料庫 Ω_A 中，如式 (1) 所示。

$$T\mathbf{X}_M = T(\mathbf{X}_M) \in \Omega_A \tag{1}$$

接著，以轉換後的 $T\mathbf{X}_M$ 視為預測元 (predictor) 或自變數 (independent variables)， Ω_A 中的輔助加值變數位 Y_A 為應變數 (dependent variable)，建立適當的統計加值預測模式，如下式。

$$Y_A = f(T\mathbf{X}_M) + \varepsilon \tag{2}$$

最終的加值預測模式即為式 (2) 中，不含誤差項的模式

$$Y_A = f(TX_M) \quad (3)$$

最後步驟即是在主資料庫中，各個觀測值的基本變數 X_M ，轉換成屬於 Ω_A 中的 TX_M ，再套入預測模式，式 (3)，得到

$$\hat{Y}_M = f[T(X_M)] \quad (4)$$

至此階段即完成加值的工作，其中 \hat{Y}_M 即是在主資料庫中的加值變數，式 (4) 也就是加值映射函數。

三、評估

經過前面資料庫函數後，得到主資料庫中映射（加值）完成的資料 \hat{Y}_M ，需對結果進行評估。基本上，於輔助資料庫之中建立一個與映射函數相同的模式，稱為「基準模式」（benchmark model）， $\hat{Y}_A = f(X_A)$ ，做為評估的基礎。無論是加值預測模式或基準模式，皆是屬於預測模式。一般而言，加值變數為離散型的預測模式之分類結果是否可靠，判斷的方法為錯分率 MCR（missclassification rate）（沈明來，1998），

$$MCR = \frac{\sum_{i=1}^n I(y_i \neq \hat{y}_i)}{n} \quad (5)$$

其中 n 為預測資料庫中的樣本數； $I(y_i \neq \hat{y}_i)$ 為指標函數，若預測值 \hat{y}_i 與實際觀測值 y_i 不相等，其值為 1，反之為 0。

對於連續型之加值變數的預測模式精確性與好壞的評估方法，常以均方差 MSE（mean square error）作為依據，主要想法即是探討預測值 \hat{y}_i 與實際觀測值 y_i 之間的平均差異，計算方式為

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k} \quad (6)$$

其中 n 為預測資料庫中的樣本數， k 為預測模式中所使用的參數數目。

無論是 MCR 或 MSE，所代表的意義皆是預測模式所造成預測錯誤的平均情況，當然錯誤情況越小越好。也就是這兩個統計量越小越好，因此可用來判斷預測模式其預測能力的精確度。

故可計算出評估輔助資料庫中基準模式的 MCR_A 或 MSE_A ，及兩資料庫透過串聯與轉換後之函數映射模式之 MCR_M 或 MSE_M 。接著，為評估本研究提出建立函數映射模式優劣之依據，本研究以「相對有效性」（relative efficiency）的想法（Ott, 1993），建立所謂的「相對誤差率」（relative error rate, RER），即加值預測模式之 MCR_M 與 MCR_A （或 MSE_M 與 MSE_A ）之比率，可視為是一種評估的指標，作為探討本研究提出之加值預測模式的評估依據。

$$RER = \frac{MCR_M}{MCR_A} \quad \text{或} \quad RER = \frac{MSE_M}{MSE_A} \quad (7)$$

RER 的範圍應大於 0，倘若 RER 大於 1，表示函數映射模式的加值預測效果較基準預測模式差；RER 越接近 1，代表映射模式的加值的效果越好，越接近基準預測模式的效果；若 RER 等於 1，則代表映射模式的加值與基準預測模式效果一樣好。反之，RER

若小於 1，表示映射模式的加值預測效果較基準預測模式為優良，而越接近 0，則越優良。

肆、結果

本研究利用資料庫函數映射方法與步驟，將第一次的技術創新調查 TTIS1 視為主資料庫，第二次的創新調查 TTIS2 視為輔助資料庫。兩個資料庫之基本變數皆選擇產業別、區域別、員工人數別，其中產業別分成製造業及服務業，區域別分成北中南三區（因東部區域調查資料過少而不列入分析），員工人數別分成：(1)19 人以下、(2)20~49 人、(3)50~249 人、(4)250 人以上；此三個變數亦作為分層隨機抽樣的分層依據，共分成 24 層。接著，分別於兩個資料庫的 24 層每層中抽取 20 家合併後成為預測資料庫，此時主資料庫與輔助資料庫分別有 480 家的資料。

另外，在主資料庫中挑選 2000 年員工人數作為映射函數的自變數；在輔助資料庫中挑選 2004 年的員工人數為相對應的自變數，以及 2004 年的營業額作為應變數，即為欲映射至主資料庫的變數。

一、自變數轉換

由於兩個資料庫的相對應的自變數不盡相同，主資料庫中的「2000 年員工人數」，輔助資料庫為「2004 年員工人數」，故需先進行基本資料的自變數轉換，即以迴歸分析技術建立轉換函數。此時，本研究以輔助資料庫為「2004 年員工人數」作為應變數，主資料庫中的「2000 年員工人數」為自變數，經測試以建立二次迴歸模式較佳。轉換模式為

$$TX_M = 66.102 + 0.666X_M - 0.00005X_M^2$$

此迴歸轉換模式之 $R^2 = 0.531$ 。

二、加值函數映射

下一步即是將主資料庫中「2000 年員工人數」進行轉換得 TX_M 並做為自變數，輔助資料庫中的「2004 年營業額資料」則做為應變數 Y_A ，建立函數映射之迴歸加值模式，即 $Y_A = f(TX_M)$ 。另外並於輔助資料庫中以「2004 年員工人數」做為自變數 X_A ，而「2004 年營業額資料」做為應變數 Y_A ，建立函數映射之迴歸基準模式，即 $Y_A = f(X_A)$ 。

表 1 及表 2 中可看出不論加值模式與基準模式之迴歸分析參數結果都顯著，表示以此迴歸模式進行函數映射是有意義的。表 2 中亦可看到加值模式評估標準 RER 為 1.0610，接近 1，表示以此方式加值所得到的結果是相當不錯的。然而，亦看到兩個模式的 R^2 都不太高，且 MSE 都滿大的，應該是還有一些自變數還未納入模式中的關係。

表 1 資料映射函數模式參數估計結果

Table 1 Coefficient estimates of functional mapping model

模式	參數	估值	標準誤	t 值	p-value
加值模式	β_0	893830.919	409982.965	2.180	.030
	β_1	8605.289	1077.383	7.987	<.001
基準模式	β_0	1141610.201	365271.553	3.125	.002
	β_1	7463.497	761.921	9.796	<.001

三、穩定性討論

前面所介紹的是一套預測資料庫樣本資料下所進行的加值研究，此一小節中將以相同的過程與方法，但隨機另外抽取出不同的十套預測資料庫樣本進行比較，探討資料庫映射過程的穩定性，結果如表 3 所列。研究發現，此十套之樣本加值分析所得 RER 的平均值為 1.0589，相當接近於 1；且其中有兩套樣本資料的 RER 值甚至小於 1，顯見加值效果應該不致於太差的。此外，此十套之樣本的 RER 值皆在 1 上下，並沒有產生過大或過小的現象，變化程度不大，表示此兩資料庫的結構應大致相似；當然也顯示，以此模式進行加值分析的穩定性也是相當高的，所獲得的加值資料的可信度應該也不差。

表 2 資料映射函數模式檢定結果

Table 2 Goodness fitting of functional mapping model

模式	R^2	ANOVA		MSE	RER
		F 值	p-value		
加值模式	.118	63.796	<.001	5.381E13	1.0610
基準模式	.167	95.954	<.001	5.072E13	

表 3 10 組預測資料庫樣本估計結果

Table 3 RER of 10 predictive samples

樣本	模式	MSE	RER
1	加值模式	5.489E+13	1.0715
	基準模式	5.122E+13	
2	加值模式	5.478E+13	1.1082
	基準模式	4.943E+13	
3	加值模式	5.236E+13	1.0431
	基準模式	5.020E+13	
4	加值模式	5.393E+13	0.9948
	基準模式	5.421E+13	
5	加值模式	5.620E+13	1.1158
	基準模式	5.037E+13	

樣本	模式	MSE	RER
6	加值模式	5.760E+13	1.0724
	基準模式	5.371E+13	
7	加值模式	5.395E+13	1.0930
	基準模式	4.936E+13	
8	加值模式	5.887E+13	1.0798
	基準模式	5.452E+13	
9	加值模式	5.401E+13	0.9896
	基準模式	5.457E+13	
10	加值模式	5.563E+13	1.0204
	基準模式	5.452E+13	
RER 平均			1.0589

伍、討論

本研究以資料庫函數映射的加值概念為出發，企圖將 TTIS2 特有的資料內容映射至 TTIS1，而能夠完成資料加值的功能，以及生成產業創新調查的模擬長期追蹤資料，使得研究創新流程時更能完整描繪出創新的投入與產出。本研究所得到的結果，RER 為 1.0610 可說是讓人滿意的，也表示以此方法對於 TTIS1 資料庫的擴增是可行的且穩定的。

研究中也看到加值模式與基準模式的 R^2 都不太高，且 MSE 都滿大的，應該是還有一些自變數還未納入模式中的關係，這就需要花更多時間投入，以找到最完美的加值模式。然而，如同 Box、Berry 與 Linoff 等人所指出的：所有預測模式無法保證能夠得到百分之百的精確，但都能提供寶貴的訊息。然而，我們能夠努力的方向即是：透過資料加值的過程，讓原本主資料庫中缺乏的資料，從無到有，並盡量將預測誤差減少，達到能夠針對資料庫之中每一個樣本都有豐富且適當的資料，讓分析研究人員不再為資料缺乏的問題困擾。可見若時間與成本充足，本研究的使用方法應可繼續完成產業創新調查資料庫之擴增，以及串聯兩次調查資料庫成長期追蹤調查資料庫。

此外，在研究的過程中有些問題仍須於未來可供討論的，其中最主要的兩項為：

1. 資料中若存在離群值可能會造成資料函數映射結果不盡理想，應尋找出一套資料純化的方法。
2. 函數映射結果的評估之理論探討。

上述兩項主題亦為本人欲繼續進行深入研究之課題，若能完成此二項課題的研究，將使得資料庫函數映射流程更加完備，增加結果的可信度與可使用度。

致謝

本研究經費獲得經國管理暨健康學院 100 年度校內研究計畫補助。

參考文獻

- 江志民（2006）資料庫中資料加值程序之研究：以農業資料庫為例，台北，數據分析，1卷2期。
- 吳思華、謝邦昌、黃文璋、鄭宇庭、徐怡、蘇志雄、許牧彥、江志民（2009）臺灣地區第二次產業創新活動調查研究，臺北，政治大學。
- 李宜穗（2010）探討台灣地區產業創新調查 87-89 年 Panel 資料創新情形之研究，台北，輔仁大學應用統計研究所在職專班碩士論文。
- 周彥宏（2003）技術創新績效指標模式之建立，碩士論文，台南，南台科技大學企管所。
- 林志維（2002）競爭策略,技術環境,技術創新與創新績效之關係研究-以電子高科技業與化工業為研究對象，碩士論文，台南，國立成功大學企業管理研究所。
- 劉孟俊（2004）產業科技創新相關調查之產業研發議題綜合分析，臺北，中華經濟研究院。
- 經濟部中小企業處（2007）96 年度運用科技再造中小企業競爭，臺北，優勢政策研究計畫成果報告。
- Knell, Mark, Nas, S. O.,(2006). What is missing in the analysis of input-output relationships of innovation processes?. Ottawa. Blue Sky II Forum on “What Indicators for Science, Technology and Innovation Policies in the 21st Century?”
- OECD/EUROSTAT (2005) OSLO MANUAL 3rd Edition.
- Raymond, W. P. Mohnen, F, Palm, S. Schim van der Loeff (2006). Persistence of innovation in Dutch manufacturing: Is it spruious?, Cirano Scientific Series 2006s-04.

