

第三章、預測模型建構與實驗設計

在第二章的文獻探討後，本研究認為在 CPFR 流程下，由於銷售預測模型強調短期預測與考量事件資訊，因該採取三階段混合預測模型，結合時間序列方法模型於因果回歸模型中，如此可以同時納入上述兩種方法之優點，即銷售量時間變化趨勢與特殊事件影響效果。由於當 CPFR 流程中例外事件觸發，可促使協同夥伴規劃事件進行調整，藉由資訊即時更新，預測模型亦可快速呈現該例外事件觸發調整之銷售預測結果，使預測在考量最新資訊下進行，提升預測績效。另外，透過基因演算法來進行回歸模型參數之最佳化，期能藉由嘗試錯誤來自我學習調整，反應不同時間事件訊息與資料對於銷售量解釋程度不同的需要，改善模型與預測績效。

本研究預測模型採用三階段混合預測方法之原因可從企業預測技術採用趨勢與偏好、混合預測方法之優點、CPFR 流程短期銷售預測之內涵與特性等三部分來看，條例如下：

1. 企業預測技術採用趨勢與偏好：

由文獻探討中發現，企業在採取預測技術時，會考量成本效益、模型解釋性與可理解性、技術熟悉度、方便性等考量。另外，文獻中 Jain (2002) 的數據可以發現，決策者使用模型時，偏好可以產生易於理解、穩定預測值的模型，偏好操作簡單的時間預測模型與易於解釋因果關係的因果回歸模型，並認知到因果分析在複雜環境下，對於預測準確性的重要性。而因果分析模型的資訊取得成本隨者資訊科技與網際網路的進步、企業夥伴協同合作程度的提升而降低，直接資訊可得性與及時性提高，亦使該方法之可採用性增加。對於黑箱型難於理解之類神經預測方法雖有研究證明其預測績效較一般統計線性預測方法佳 (Chu and Zhang, 2003)，但較少被採用。

2. 混合預測方法之優點：

統計上每種方法有其優缺點與考量，Kolter (1991) 指出時間序列方法用於短期預測有不錯之績效，尤其是在擁有產品線過百個項目之公司，作有效且經濟之短期預測時，他認為可以使用指數平滑法。但是時間序列方法僅將過去銷售量作為未來銷售量之函數，不考慮其他實際需求與影響因素，而頻繁促銷事件與不穩定環境會影響預測績效。一般因果回歸模型在採用上，僅考量影響銷售事件之因果資訊，忽略時間因子。而由文獻探討中發現，計量經濟模式預測如因果回歸預測方法的結果並沒有比時間序列方法準確，但是可以提供變數間的因果關係。而許多學者指出，透過混合數種方法之預測模型可以互補長短，提升預測績效。透過時間預測模型與因果回歸模型之結合使用，可以同時考量銷售量的時間關係與具影響事件資訊。

3. CPFR 流程短期銷售預測之內涵與特性：

在 CPFR 流程下，強調進行短期、個體預測(Micro Forecasting)，根據少量且多元的過去、現在與未來資料，透過配合性較佳的函數作短期預測，並持續不斷的作模型修正，以求快速反應。而事件訊息與短期詳細資料易呈現複雜變異性、不同時間之資料與訊息所反映的解釋程度亦不同，需要藉由嘗試錯誤來自我學習調整，以求最佳化預測模型參數解，使預測績效提升。傳統統計預測方法求解時，利用普通最小平方法求模型參數解，在運算時需依賴大量的歷史資料(資料量大於 25 筆)，以求取一個長期平均趨勢線，再用外插法求預測值，其預測誤差率高，且無法滿足嘗試錯誤不斷自我學習調整之需求。透過短期模型參數最佳化運算，可以改善此問題。而模擬自然界演化程序，強調自然演進多點搜尋的基因演算法，相對於傳統最佳化搜尋法可以較快的獲得整體最佳解。由文獻探討 2.3.2 節中發現，其在使用上與邏輯上具有簡易性，並可以有效的解決函數最佳化和控制問題，在預測上已有許多學者使用基因演算法來求預測模型的參數，如 Jeong 等人(2002)研究發現利用基因演算法求多元回歸模型參數解中，確實可以改善其預測績效。因此，若透過基因演算法進行短期最佳化演算法求得函數參數之最佳值，將能有效且有效率的改進預測模型之績效。

本研究以 Jeong 等人(2002)所提之基因因果預測方法作為本研究預測模型之基礎，發展本研究之混合預測模型。

本章以下部份將詳細說明本研究所提出的混合預測模型、模型建構方法、與實驗設計。

3.1 混合預測模型整體架構

本研究中提出之混合銷售預測模型，包含三個部分：(1)時間序列—指數平滑模型；(2)經濟計量模型—多元回歸模型；(3)基因演算求最佳化之預測模型。有系統、順序的融合上述時間序列銷售趨勢特性與促銷等銷售量影響因子於預測模型中。混合預測模型架構圖如圖 11。

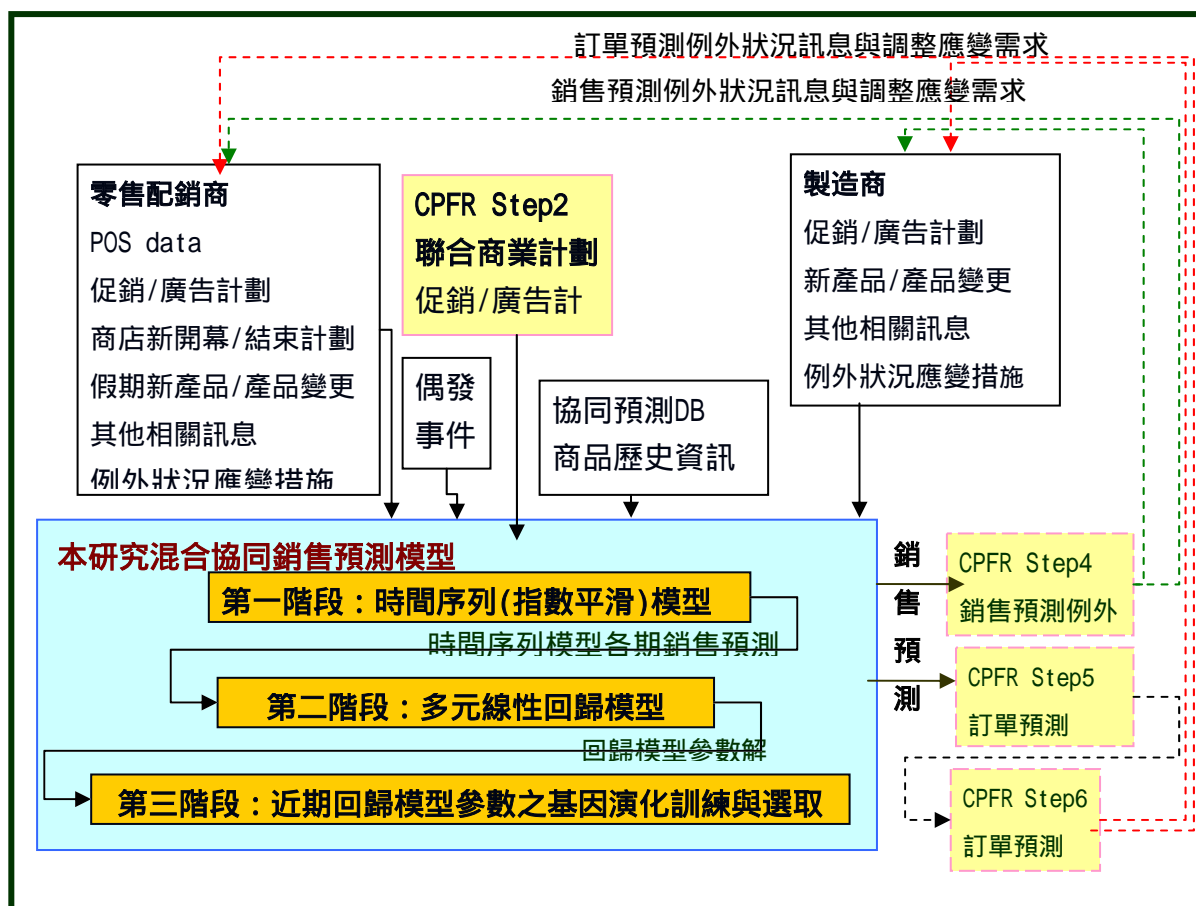


圖 11、混合預測模型架構圖

茲概述如下：

(1) 時間序列—指數平滑模型時間序列：

時間序列方法有許多種，但許多資料顯示指數平滑法與伯斯-堅肯斯(ARIMA)模型之績效並無太大差異 (Mills, 1990)，而後者在採用上複雜許多 因此，本研究中採用企業界常用且簡單之指數平滑模型 針對欲進行預測之產品的歷史銷售數據特性，選定合適的指數平滑預測方法，並利用敏感度分析，求取最佳之平滑參數值 透過指數平滑法，可得下第 $t+1$ 期之預測值，再將此預測值作為第二階段多元回歸模型之第 $t+1$ 期輸入變量之一

(2) 多元回歸模型；

由於時間序列模式中，無法有效納入事件因子，第二階段以多元回歸模型加強之 透過可得資料分析，求出影響銷售量之事件，並加以量化 透過統計軟體，如 SPSS 等，進行分析，求出多元回歸模型之參數解，作為第三階段混合預測模型，基因演化之初始參數解，以求短期最佳化參數解

(3) 基因演算求最佳化之混合預測模型

由於不同時間之資料與訊息所反映的解釋程度不同亦需要藉由嘗試錯誤來調整，以求最佳化預測解。若有最佳化演算法來求得函數參數之最佳值，將能改進其模型效率。多元回歸模型之參數解，透過基因演算法演化，可以進一步修正，求短期內最佳之預測模型。

3.2 資料蒐集整理與應用

本研究中，蒐集以「週」為單位之銷售資料、促銷等事件資料、產品資訊，並將資料劃分為兩區段，第一區段，為模型訓練建構期，第二區段為預測結果績效測試期。模型訓練建構期中，又區分為兩大部分，一為統計回歸訓練部分，透過統計多元回歸可得知其影響相關性，模型解釋度；二為基因演化部分，以週為單位進行最佳預測模型演化。

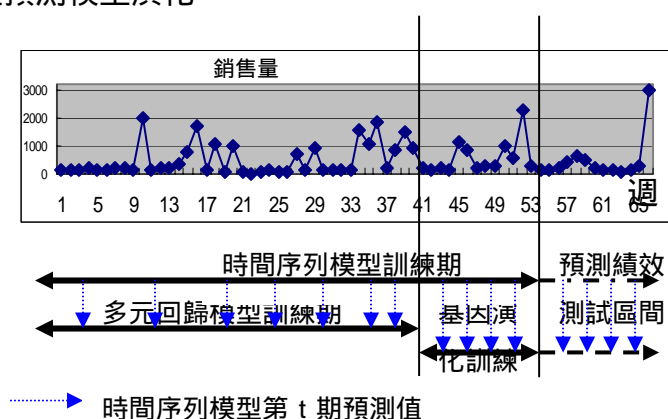


圖 12、本研究各階段劃分與資料應用示意圖

3.3 時間序列—指數平滑模型時間序列：

Kolter(1991)指出，在擁有產品線過百個項目之公司，作有效且經濟之短期預測可以使用指數平滑法。另外，由於短期預測中，銷售量售促銷等事件影響變動性大（舉例如圖 13），為了能夠反應該變動性，對不同時期資料採取不同資料加權，本研究捨去固定加權之平均法，採用簡單且成本低之平滑法。而考量「週」為單位之銷售量資料變化性大、季節特性不明顯，所以將採用賀特(Holt)兩個參數指數平滑法於本模型中。

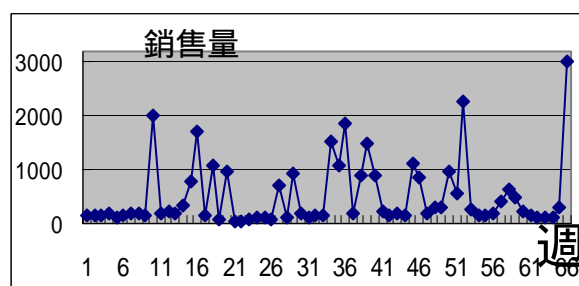


圖 13、週銷售量曲線舉例

賀特的雙重指數平滑法—忽略季節性因子，並對趨勢作二次平滑 該方法如下所示：

$$S1_{(t)} = a * X_{(t)} + (1-a) * [S1_{(t-1)} + S2_{(t-1)}]$$

$$S2_{(t)} = b * [S1_{(t)} - S1_{(t-1)}] + (1-b) * S2_{(t-1)}$$

$$Est_X = S1_{(t)} + S2_{(t)} * m$$

其中，

S1 表一次平滑因子

S2 表二次平滑因子

$X_{(t)}$ 表第 t 期銷售量

a, b 表平滑係數

m 表示往後預測之期數

由於平滑參數的選定，對指數平滑模型之績效影響甚大，故以敏感度分析求之，以該產品預測之 MAPE 值欲小者為佳

3.4 多元回歸模型

由於時間序列方法僅將過去銷售量作為未來銷售量之函數，不考慮其他實際需求與影響因素，本研究試圖透過多元回歸統計方法將這些影響因素考量進來

由 2.2.6 文獻探討中發現，影響銷售量之因素有很多，因為消費者購買行為一直是難以預測與理解的，長久以來許多學者試圖找出其主要影響因素 許多行銷相關研究(如 Shankar & Krishnamusthi 1996, Mulhern, Williams & Leone 1998, Voss & Seoders 2003 等)都指出產品本身價格因素(定價、售價、價格變動)、零售商與品牌商廣告、促銷活動(價格則扣幅度、促銷期長度)、賣場展示、特定假期與時間等因素會影響消費者購買行為，影響短期產品銷售量 而其他相關產品(如替代性、相依性產品等產品)之價格與促銷活動、品牌效益等亦會影響銷售量 而依據 VICS 所提之 CPFR 流程模型與 Armingier(2002)發表之文章中建議銷售預測應考量買賣方銷售促銷、行事曆規劃事件(如新產品上市、下市等)、產品歷史銷售資訊、季節與特殊節日等資訊

然而，雖然預測模型中考量的變數愈多其涵蓋的資訊愈多，預測可以較準確，但就實際預測作業而言，變數愈多資訊處理與蒐集成本愈高且其邊際績效貢獻度隨變數數增加遞減，且產業別與產品別解釋變項亦有不同 因此本研究中僅針對一般常見之重要變數，如價格、特徵廣告與展示、促銷活動、產品所處之生命週期階段、假期與特殊節日或季節性、競爭產品價格與促銷活動等作為模型解釋變數 並透過彙整方式，將可得之量化資料透過統計相關分析過濾

找出該產品銷售量之主要解釋變數

本研究回歸模型變數與其定義和方法如下說明，並對欲預測之商品簡稱產品 A：

解釋變數 1：產品售價(Selling Price)：銷售點之售價

解釋變數 2：價格差率：售價與定價差異比率， $(\text{售價}-\text{定價})/\text{定價}$

解釋變數 3：主要替代性產品與該產品之價差率： $(\text{替代性產品售價}-\text{產品 A 售價})/\text{產品 A 定價}$

另透過促銷活動與銷售量相關分析，找出影響銷售量之關鍵促銷活動，並視為解釋變數之一

解釋變數 4：商店關鍵促銷活動之有無

解釋變數 5：產品 A 關鍵促銷活動之有無

由於零售端促銷活動多，使用搭配上重複性高，為求在關鍵活動訊息外，仍包含所有活動績效訊息，以針對個別產品之週促銷活動數量、週促銷活動天數作為解釋變數：

解釋變數 6：個別產品之週促銷活動數量總計：各週中，有產品 A 促銷活動之活動數量總計

解釋變數 7：個別產品週促銷活動天數總計；

各週中，有產品 A 促銷活動之活動天數總計 由於零售商店之整體促銷活動(欲預測產品不一定有促銷活動)會影響產品銷售量(帶動買氣，或抑制買氣)，因此亦將此設為一可能之解釋變數

解釋變數 8：零售店週別整體促銷活動數(Storewide Promotion activity Quantity)：各週零售店舉行之促銷活動數量

解釋變數 9：該期所屬月之尖峰類別代表數

由於個別產品週別資料波動較大，不易得知產品銷售量之尖離峰與季節特性，本研究中透過人工方式，由產品類別之各月總計售量依大小排列，利用四分位點概念，劃分出最高峰期(大於第四分位點者)、次高峰期(第二分位點~第三分位點間者)、最低期(小於第一分位點者)、次低期(第一分位點~第二分位點間者)月份 高峰期指數設為 4、次高峰期指數設為 3、次低峰期指數設為 2、最低期指數設為 1

解釋變數 10：產品所處生命週期

產品隨所處之生命週期階段不同，其銷售策略、生產策略會有不同。創新採用理論亦指出產品所在之產品生命週期不同，消費者對產品接受喜好度會有不同，影響銷售量。但由於事先判定各別產品生命週期困難，以人工方式認定本研究中之簡單設定判別方式如下：以產品推出、產品推出後第 n 個月、製造商推出新替代性產品、製造商規劃產品下市(CloseOut)，事件作為分界點：

階段一(導入期)：產品推出~產品推出後第 n 個月；

階段二(成長期)：產品推出後第 n+1 個月~製造商推出新替代性產品；

階段三(成熟期)：製造商推出新替代性產品~製造商推出新替代性產品第 n 個月或尚未規劃下市；

階段四(衰退期)：製造商規劃產品下市(CloseOut)且製造商推出新替代性產品第 n 個月以後

以上 n 值之界定，可由產品類別銷售量之時間序列曲線歸納分析概取得之。依上述方法，則導入期指標值設為 2，3 為成長期，4 為成熟期，1 為產品末期，將進行結束生產與銷售。

產品製造商之活動有時會影響產品銷售量。如新世代產品推出等。本研究中，將有無新產品導入該零售店銷售，視為一個可能之解釋變數。最後，加上時間序列模型該期之預測值。

解釋變數 11：該期是否有同類別之新產品導入

解釋變數 12：時間序列模型之預測值

基於解釋變數資料特性(各變項數值差異大)，若單純用線性，則易使指標型變數(如值小於 10 者)其解釋效果被忽略，若採用半對數型線性函數，則又容易使預測值過度膨脹失真，本研究將採用雙對數線性回歸模型，並依據自變數特性，對自變數作適度轉換，模型預設如下：

$$\begin{aligned} \text{Log}X(t) = & b_0 + b_1 * \text{logSell}(t) + b_2 * \text{logSDisctRate}(t) + b_3 * \text{logCDiffRate}(t) + b_4 * (\text{StoreKeyAc}(t) \\ & + b_5 * (\text{ProdKeyAc}(t)) + b_6 * \text{AdQty}(t) \\ & + b_7 * (\text{AdDays}(t)) + b_8 * (\text{SotreAc}(t)) + b_9 * (\text{ProdAc}(t)) + b_{10} * (\text{Peak}(t)) + \\ & b_{11} * (\text{PDLC}(t)) + b_{12} * (\text{NewProd}(t)) + b_{13} * \text{logArEst}(t) + e \end{aligned}$$

其中，

$\text{log}X(t)$ 表第 t 期銷售量對數值，

logSell = 實際售價之對數值

SDisctRate：價格差:售價與定價差異比率， $(\text{售價}-\text{定價})/\text{定價}$

$\log\text{SDisctRate} = \text{SDisctRate}$ 之對數值

CDiffRate=主要替代性(競爭)產品與該產品之價差率： $(\text{替代性產品售價}-\text{產品 A 售價})/\text{產品 A 定價}$

$\log\text{CDiffRate} = \text{CDiffRate}$ 之對數值

StoreKeyAc=該週是否有商店層之關鍵整體促銷活動，若有，值為 1，若無以 0 表示

ProdKeyAc=該週是否有該產品之關鍵促銷活動，若有，值為 1，若無以 0 表示

AcQty=該月產品 A 之廣告活動週次數， $0 \leq \text{AcQty} \leq 5$

AdDays=產品 A 之廣告活動總計天數， $0 \leq \text{AdDays} \leq 7$

SotreAc=該週零售商店整體之廣告與活動數量

ProdAc=該週產品別之廣告與活動數量

Peak=該期所屬月之尖峰類別代表數，將 12 個月劃分四個類別 1 為表最低尖峰期，4 表最高尖峰期 2 與 3 分表低與高

PDLC=該期該產品所處之生命週期階段，分為 1-4 期 2 為導入期，成長期為 3，成熟期為 4，1 為產品末期，將進行結束生產與銷售

NewProd=該週是否同類別新產品導入期，若是為 1，無為 0

ArEst=時間序列模型預測值

$\log\text{ArEst} = \text{ArEst}$ 之對數值

針對可得資料與統計工具分析變數之合適性，求出合適之回歸預測模型，以 SPSS 統計軟體，進行模型解釋度與相關性、解釋變數間共線性分析 由於變數間若存在共線性會降低模型之績效，以變異數膨脹因子(Variance inflation factor, VIF)來作多元回歸共線性分析 通常 VIF 值應小於 10 為佳

3.5 基因演算求最佳化之混合預測模型

3.5.1 基因演算法之染色體編碼與適應函數

3.5.1.1 編碼

本研究中將多元回歸模型中的 n 個解釋變數視為基因，共有 n 個基因，而染色體係指有 n 個基因依照順序排列組合 一組基因組合(回歸模型參數解組合)稱為一條染色體

基因演算法中染色體基因編碼部分，由於實數編碼一般而言比二進位編碼在速度上與結果上來的快且佳，因為二進位編碼須對染色體基因(問題解)進行二位元形式編碼，如此可能會造成結果失真現象，且額外的編、解碼動作增加

執行時間，因此本研究中編碼採用實數編碼

3.5.1.2 適應函數

由於適應性函數建構的好壞，會影響基因演算之結果，而 Jeong 等人(2002)之研究發現，除了當期誤差、對過去期間誤差外，另對於不合基因值規範域之染色體合適值部分，乘以前 t-1 期之誤差均方根可以提升演化結果預測模型之績效，故採用之 因此本研究之適應函數與 Jeong 等人(2002)大致相同

本研究適應性函數共分為三個部份 結合(1)當期的預測誤差；(2)當期以前之所有誤差的平均值；(3)基因值不合規範域之誤差 本研究中績效判斷指標為 MAPE，不同於 Jeong 等人(2002)等人研究中採用平均絕對誤差值(MAD)，故適應函數之前二個部分皆採為誤差率 為避免適應值過小影響判斷，乘以 100 三部分公式如下：

$$\begin{aligned}
 (1) \text{ 當期的預測誤差率：} & \quad \frac{|Y(t)-y(t)|}{Y(t)} * 100 \\
 (2) \text{ 當期以前所有誤差率之平均值：} & \quad \frac{\sum_{i=1}^{t-1} \frac{|Y(i)-y(i)|}{Y(i)}}{(t-1)} * 100 \\
 (3) \text{ 基因值不合規範域之誤差值：} & \quad \frac{\sum_{i=1}^n (1+P1_i + P2_i + P3_i) * 100 * \sqrt{\sum_{j=1}^{t-1} \sum_{i=1}^n (A_i \cdot B_{j,i})^2}}{t-1}
 \end{aligned}$$

Jeong 等人(2002)等人對於基因值不合規範域部份，僅就基因值變化上下限作規範，本研究中針對銷售預測特性，另加入兩指標 故本研究中基因值規範域部份，主要有三個指標：

指標一(P1)：基因值變化之上下限(L,U) 當該基因超出範圍，指標設為 1

指標一(P2)：為避免個別變項在預測時因為該參數對應之基因值為 0 被忽略 基因值不可為 0，當該基因值=0，指標設為 1

指標一(P2)：由於各期初始參數值中，其正負號代表該變項與銷售量間之正負相關性，當基因值與該初始值正負號相異時，指標設為 1

透過權重值加總判斷合適性，愈小者佳 其目的是希望能促使該最佳解能適當的反應期間特性，亦即愈接近預測期的誤差值之權重愈大，懲罰性愈大，並考慮其他限制性因素，以加速收斂 合適性函數完整公式如下：

$$F_{Penalty} = a * \frac{|Y(t)-y(t)|}{Y(t)} * 100 + b * \sum_{i=1}^{t-1} \frac{|Y(i)-y(i)|}{Y(i)} / (t-1) * 100 + r * \sum_{i=1}^n (1+P1_{i,t} + P2_{i,t} + P3_{i,t}) * 100 * \frac{\sum_{j=1}^{t-1} \sqrt{\sum_{i=1}^n (A_{i,j,t} - B_{j,i,t})^2}}{t-1}$$

其中， i, t 表期間， n 表基因個數(參數量)， a, b, r 為權重值， $a+b+r=1$

A_i =替代染色體之第 i 個基因，

$P1_i$ ：若($A_i < L_i$)或($A_i > U_i$)， $P1_i=1$;

其他狀況： $P1_i=0$

$P2_i$ ：若($A_i=0$)， $P2_i=1$;

其他狀況： $P2_i=0$

$P3_i$ ：若($A_i * B_{t,i} < 0$)， $P3_i=1$;

其他狀況： $P3_i=0$

$B_{j,i}$ 表第 j 期最佳染色體第 i 個基因值

(L_i, U_i)：表基因 i 的變化範圍

3.5.2 基因演算法之染色體產生與複製

每期第一代染色體群的產生，係由每期之初始染色體(基因演化第一期，以多元回歸模型之參數解為之；第二期以後則以前一期最佳解為之)為中心，訂定正負 $a\%$ 之上下界進行隨機複製，其複製公式如下：

$$B_{t,i} = B_{tsi}^{MIN} + (B_{tsi}^{MAX} - B_{tsi}^{MIN}) * Rand$$

$$B_{t,i}^{MAX} = (1 + a) * INI_{t,i}, INI_{t,i} > 0 \\ (1 - a) * INI_{t,i}, INI_{t,i} < 0$$

$$B_{t,i}^{MIN} = (1 - a) * INI_{t,i}, INI_{t,i} > 0 \\ (1 + a) * INI_{t,i}, INI_{t,i} < 0$$

其中， $Rand$ 表隨機變數值

($B_{t,i}^{MIN}, B_{t,i}^{MAX}$) 表第 t 期，第 i 個基因變化之範圍

$INI_{t,i}$ 表第 t 期初始染色體之第 i 個基因值

a 表變化範圍定義因子

在代染色體群的產生部份，除了第一代以上述隨機方法複製產生，第二代以後，不作複製，以上一代產生演化出之所有染色體以隨機方法進行交配

為避免基因演算法中過早收斂(區域最佳解)問題，Jeong 等人(2002)所提

之預測模型中加入了族群差異指標(Population Diversity Index, PDI)—藉由比較最近一期最佳解與最佳解集合內所有解的相似度，計算族群差異指數，以做為判斷族群相似度的指標，以跳脫區域最佳解之陷阱 使用之方式，乃在第三期以後，判斷 PDI 指數是否大於 PDI 隨機變數判定值 R，若是，表示最佳解集合內所有解的相似度很高，以隨機方式重新定義最佳染色體基因組合 本研究中在前測實驗中發現，將 PDI 用於期最佳解集合之相似度判斷，無法快速且有效的跳脫區域最佳解，但若將 PDI 指數之判斷用於每期各代最佳解集合相似度判斷中，可幫助跳脫區域解，使每期最佳解在少數代演化下產生，並且有不錯的預測績效 故將其用於每期第十代以後之各代最佳解集合相似度比較

本研究中 PDI 之公式如下：

$$PDI = \frac{\sum_{j=1}^{t-2} (\log_{t-1}(1+j)) * (\sum_{i=1}^n |F_i(t-1) - F_i(j)| / n)}{\sum_{j=1}^{t-2} \log_{t-1}(1+j)}$$

其中，t 表代數；t>=10；

n: 基因個數(參數量)；

i: 基因位置

j: 期間

$\log_{t-1}^{(1+i)}$ 表權重值，期間差距愈大者，權重欲小

3.5.3 基因演算法之交配與突變、子代選擇方式

基因演算法之交配與突變方式有許多種，為比較不同交配與突變方式配改善基因演算法區域最佳解與太早收斂、難以收斂之成效，本研究中採用兩種方式作為比較：一為 Jeong 等人(2002)所使用的交配突變方法；二為 Herrera 等人(1997)所提之模糊連結基礎之交配(Fuzzy connectives based crossover)，利用模糊運算探索概念，增加基因之變化性，與 Jeong 等人(2002)模型中之簡單單點交配加內嵌式突變不同 子代選擇部分，採用親子互競方式，與 Jeong 等人(2002)之方法同

3.5.4 基因演算法之演算流程

基於上述交配與突變方式，本研究模型有兩種基因演化流程 以下將採用 Herrera 等人(1997)之模糊連結基礎之交配模型稱為基因演化流程模型一；將採用 Jeong 等人(2002)之交配突變方法稱為基因演化流程二 敘述如下：

(1)基因演算流程一，大致如下：

步驟 0：取得多元回歸模型參數解作為初始解

步驟 1：初始化母體族群

隨機產生母體族群

步驟 2：PDI 判斷

若代數大於等於 10，計算 PDI 值

隨機產生介於 0~1 間之 PDI 判定值 R，若 (random number \geq PDI) 則 重新定義族群初始解

隨機產生代母體族群

若代數小於 3，跳至步驟 3

步驟 3 基因演化運作(Genetic operation)

步驟 3.1 由前一代演化產生新族群中，隨機選擇兩染色體，以進行交配

步驟 3.2 交配

隨機產生介於 0~1 間之交配率值 R，若設定之交配率 $<$ R，則進行模糊連結基礎運算子交配

步驟 3.3 突變

隨機產生介於 0~1 間之突變率值 R，若突變率 $<$ R，則進行單點置換突變(the swapping mutation)突變

步驟 3.5 子代染色體產生

計算新染色體之合適度，並依親子互競法，選出最佳與次佳染色體

步驟 3.5 若子代染色體 $<$ 設定之族群大小，重新回步驟 4

若子代染色體 = 設定之族群大小，重新回步驟 3.1，執行下一代演化

步驟 4 第 t 期，第 k 代，「代」最佳染色體評選

選擇最佳合適度(函數值最小者)，為該「代」最佳染色體

步驟 5 判斷第 t 期代數是否演化完畢

若演化完畢則重新回步驟 6，否則回到步驟 2

步驟 6 評選第 t 期各代最佳染色體，選出第 t 期最佳之染色體

步驟 7 判斷所有期數是否演化完畢，

若演化完畢則重新回步驟 8，否則回到步驟 1，進行下一期演化

步驟 8 產生最適預測模型

(2)基因演算流程二，大致如下：

步驟 0：取得多元回歸模型參數解作為初始解

步驟 1：初始化母體族群

隨機產生母體族群

步驟 2：PDI 判斷

若代數大於等於大於 10，計算 PDI 值

隨機產生介於 0~1 間之 PDI 判定值 R，若 (random number \geq PDI)

則 重新定義族群初始解

隨機產生代母體族群

若代數小於 3，跳至步驟 3

步驟 3 基因演化運作(Genetic operation)

步驟 3.1 由前一代演化產生新族群中，隨機選擇兩染色體，以進行交配

步驟 3.2 交配與內嵌式突變

步驟 3.2.1 隨機產生介於 0~1 間之交配率值 R，若設定之交配率 $<$ R，則進行簡單單點交配

步驟 3.2.2 隨機產生介於 0~1 間之交配率值 R，若設定之突變率 $<$ R，則進行單點置換突變(the swapping mutation)突變

步驟 3.3 突變

隨機產生介於 0~1 間之突變率值 R，若(1-交配率) $<$ R，則進行單點置換突變(the swapping mutation)突變

步驟 3.5 子代染色體產生

計算新染色體之合適度，並依親子互競法，選出最佳與次佳染色體

步驟 3.5 若子代染色體 $<$ 設定之族群大小，重新回步驟 4

若子代染色體=設定之族群大小，重新回步驟 3.1，執行下一代演化

步驟 4 第 t 期，第 k 代，「代」最佳染色體評選

選擇最佳合適度(函數值最小者)，為該「代」最佳染色體

步驟 5 判斷第 t 期代數是否演化完畢

若演化完畢則重新回步驟 6，否則回到步驟 2

步驟 6 評選第 t 期各代最佳染色體，選出第 t 期最佳之染色體

步驟 7 判斷所有期數是否演化完畢，

若演化完畢則重新回步驟 8，否則回到步驟 1，進行下一期演化

步驟 8 產生最適預測模型

本研究基因預測模型流程如圖 14：

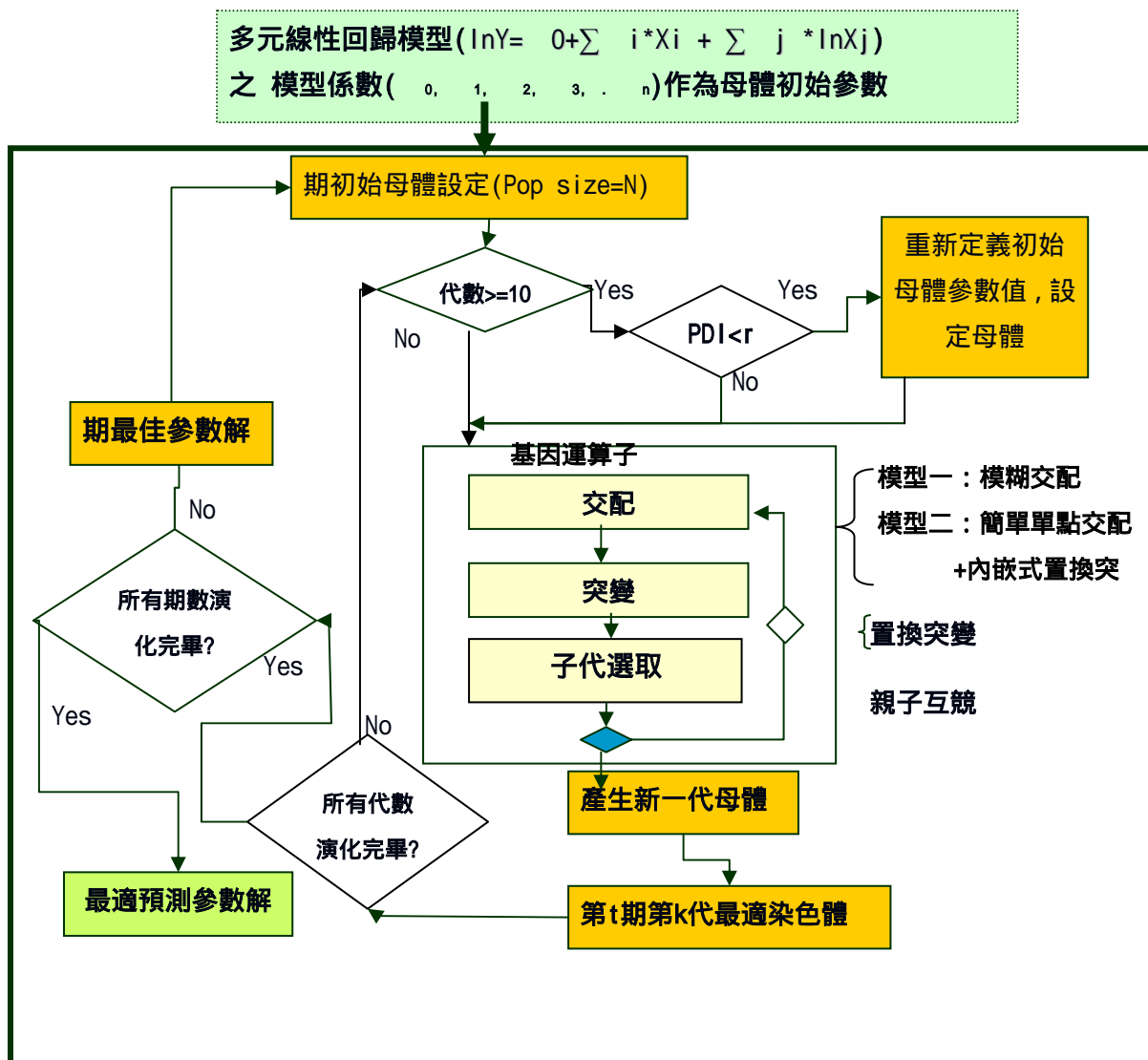


圖 14、本研究基因預測模型流程圖

資料來源：本研究

3.6 驗證方法、工具與績效衡量指標

在基因演化之混合預測模型驗證系統之建立工具部分，本研究將以 Access 內建之環境與工具開發系統模組，以進行系統雛型之展示與預測處理程序之進行

本研究之預測模型由於採用基因演化流程之不同，如 3.5.4 節所述，有兩種 第一種為採用基因演化流程一之混合預測模型，以下簡稱為本研究預測模型一；第二種為採用基因演化流程二之混合預測模型，以下簡稱為本研究預測模型二

在預測績效驗證部分，由於一般企業之銷售預測每月進行一次，一次進行三個月之預測，而一個月通常以四週為單位，故本研究將以 10 次實驗結果之四

週、八週、十二週平均績效值來檢視，並與以下三個模型之預測結果比較：

- (1) Jeong 等人(2002)之預測方法(二階段預測模型，不含時間序列之多元回歸模型，以基因演算法作最佳化)，以下簡稱為「Jeong 等人預測模型」
- (2) 多元回歸模型(不含時間序列模型)：利用統計工具 SPSS 作分析，求出該回歸模型求出該回歸模型參數，並對預測模型測試期間銷售量作預測，以下簡稱為「一般回歸模型」
- (3) 時間序列(指數平滑)模型：賀特兩個參數之指數平滑法，即本研究中第一階段時間序列模型

表 10、本研究模型與欲比較之模型特性對照表

	時間序列模型	多元回歸模型	基因演化流程	
			簡單單點交配+內嵌式突變	模糊連結基礎之交配
本研究模型一	@	@		@
本研究模型二	@	@	@	
Jeong 等人預測模型		@	@	
一般回歸模型		@		
時間序列(指數平滑)模型	@			

在衡量指標份，基於希望對極端值的影響衝擊不要過於敏感且可使預測人員迅速瞭解預測精確度，故選取平均絕對誤差百分比(MAPE)做為預測效度之衡量與比較基準 其公式如下

$$MAPE = \sum_{i=1}^m \frac{|Y(t)-y(t)|}{Y(t)} / m$$

MAPE 值愈小者越佳。另以預測與實際值比¹平均、誤差平方平均(MSE)、絕對誤差平均(MAD)，作為參考 而預測與實際值比平均，越接近 1 者為佳；誤差平方平均(MSE)、絕對誤差平均(MAD)皆越小越佳

¹預測與實際值比¹平均=(預測值/實際銷售值)