

國立宜蘭大學應用經濟與管理學系

碩士論文

Department of Applied Economics and Management

National Ilan University

Master Thesis

結合經驗模組分解與獨立成份分析於時間序列之預測

—以台灣工業生產指數為例

Integrating EMD and ICA for Time Series Forecasting

- Using Taiwan Industrial Productive Index as an Example

研究生：蘇恆信

Graduate Student: Su, Hang-Shin

指導教授：林豐政 博士

Advisor : Lin, Feng-Jenq Ph. D.

中華民國一〇一年六月

June 2012

國立宜蘭大學碩士學位論文
指導教授推薦函

應用經濟與管理 學系 蘇恆信 君所提之論文：

結合經驗模組分解與獨立成份分析於時間序列之預測
—以台灣工業生產指數為例

係由本人指導撰述，同意提付審查。

指導教授 林豐政 (簽章)

系所主管 吳中煥 (簽章)

中華民國 101 年 7 月 3 日

國立宜蘭大學碩士學位論文
口試委員會審定書

應用經濟與管理 學系 蘇恆信 君所提之論文：

結合經驗模組分解與獨立成份分析於時間序列之預測

—以台灣工業生產指數為例

經本委員會審議，認定符合碩士資格標準。

學位考試委員

甘清江

溫博仁

林豐政

指導教授

林豐政

中華民國

101

年

7

月

9

日

謝 誌

本篇論文之能夠順利完成特別要感謝許多老師、同學及朋友的協助，在此致上誠摯的謝意。特別感謝豐政老師兩年來的教導，無論在論文的啟發或是內容的討論，皆為論文能順利完成最為重要的一環。感謝葉清江老師所教導的課程，使得一些相關的理論方法能夠融合到論文裡，並於論文審查及口試時給予許多寶貴的意見。謝謝溫博仕老師的仔細審查並提供建議，這對於整篇文章的可讀性加分許多。本篇論文的成果也獲得 TIMS2012 商學與管理學術研討會最佳論文及 2012 全國碩博士論文獎決選青睞，在此由衷感謝各評審委員。

研究所期間謝謝綺貞及佶俐學姐的協助，讓許多煩雜的事務能順利解決；當然也要謝謝諸位老師專業的教導，實在是受益良多；感謝在學期間一起生活、學習、奮鬥、垃圾話的大家：珮伊、姿蓉、韓成、怡萱、培慈、迪麟、雅青、志仁、明翰、門門、瑤婷、俊亨、碧姿、文榮、冠銘、家福、舜翔、崧洽、明勳、彥良、凱富、韋蓓、嘉儀、思瑋、羽姿、台翎、詩韻、文常、濬翔、黑人、承祐、錫禹、偉庭、文慈、晴惠、妙瑜、阿義、芊芊、偉毓、德宏、明香、美美以及協助過我、幫助過我而我記不住的大家。最後，感謝總是默默支持我的家人及朋友。

蘇恆信 謹誌於

國立宜蘭大學應用經濟所碩士班

中華民國一〇一年七月

論文名稱：結合經驗模組分解與獨立成份分析於時間序列之預測

—以台灣工業生產指數為例

校院系所：國立宜蘭大學應用經濟與管理學系碩士班

頁數：61

畢業時間：民國一〇一年六月

學位別：碩士

研究生：蘇恆信

指導教授：林豐政 博士

摘 要

由過去研究發現，希爾伯特－黃轉換(Hilbert-Huang Transform, HHT)之方法運用於時間序列之預測中，使用經驗模組分解(Empirical Mode Decomposition, EMD)過程所分解出之數個本質模組函數(Intrinsic Mode Function, IMF)分量來選擇作為未來預測之變數，然而選取 IMF 分量之方式往往可能是由研究者本人，透過自然環境因素或文獻分析等歸納方式，選擇其 IMF 分量變數，故本研究認為在選取重要 IMF 仍有探討的空間。

據此，本研究將著重於以經驗模組分解(EMD)後的各本質模組函數(IMF)，透過獨立成份分析法(ICA)，將其獨立成份分解出來後，加入顯著性檢定的概念（擬偏 F 統計量檢定），將與原始序列最不相關之獨立成份刪除後(即保留重要之獨立成份)並還原資料，視其還原後之新序列為已去除雜訊過後之時間序列資料。本研究並以實際個案說明上述所提之選取 IMF 之方法的適用性與可行性。

研究中以台灣工業生產指數的時間序列為研究對象，嘗試建構得以適切解釋工業生產指數變化的時間序列模式。結果顯示，透過研究中所提之結合 EMD 與 ICA 以移除序列雜訊的混合模式，無論在資料的解釋能力或預測能力指標，確實要較直接應用原始序列建構的 ARIMA 模型改善許多。

關鍵字：希爾伯特－黃轉換(HHT)、經驗模組分解(EMD)、時間序列、獨立成份分析(ICA)

**Title of Thesis: Integrating EMD and ICA for Time Series Forecasting
- Using Taiwan Industrial Productive Index as an Example**

**Name of Institute: Department of Applied Economics and Management,
National Ilan University Master Thesis**

Page: 61

Graduation Time: June, 2012

Degree Conferred: Master

Student Name: Su, Hang-Shin

Advisor Name: Lin, Feng-Jenq Ph. D.

Abstract

In this study, the Empirical Mode Decomposition (EMD), a part of Hilbert-Huang Transform (HHT), was integrated independent component analysis (ICA) and the concept of significance test to analyze the time-series data. In this paper, a hybrid forecasting approach which combined empirical mode decomposition (EMD) and independent component analysis (ICA) was developed to predict the industrial productive index in Taiwan.

There are four stages in the EMD-ICA forecasting approach. The first stage, we analyzed the Taiwan industrial productive index by using EMD. After that, we sifted several intrinsic mode functions (IMF) from data. The second stage, we extracted meaningful information of IMFs by independent component analysis (ICA). In this study, we used the concept of significance test (pseudo partial F statistic) to choose the number of ICs. We used this concept for noise reduction and estimated the original data. The main contribution of the paper presented an adaptive hybrid method which was composed of EMD and ICA model. Finally, modeling a model to forecast the Taiwan industrial productive index.

Experimental results showed that the proposed hybrid model outperforms the traditional autoregressive integrated moving average (ARIMA) models.

**Keywords : Hilbert-Huang Transform, Empirical Mode Decomposition,
Time Series, Independent Component Analysis**

目錄

摘 要	I
Abstract	II
目 錄	i
表目錄	iii
圖目錄	iv
第壹章 緒論	1
第一節 研究動機與目的	1
第二節 研究範圍	3
第三節 研究流程	4
第貳章 文獻探討	6
第一節 經驗模組分解法(EMD)相關文獻	6
第二節 獨立成份分析法(ICA)相關文獻	8
第三節 工業生產指數的定義	10
第參章 研究設計與方法	15
第一節 結合 EMD 與 ICA 於預測模型的架構	15
第二節 經驗模組分解	17
第三節 獨立成份分析	24
第四節 單變量 ARIMA 模型	29
第五節 資料特性與顯著性檢定	32
第六節 預測績效衡量	34
第肆章 實證分析	36
第一節 時間序列資料說明	36
第二節 經驗模組分解法	37
第三節 獨立成份分析法	41
第四節 模型建立與預測	46

第五節 模型預測能力比較	53
第伍章 結論與建議.....	55
第一節 結論	55
第二節 建議與研究限制	56
參考文獻.....	58
中文部分	58
英文部分	59



表目錄

表 2-1	台灣同時景氣指標修訂構成項目	11
表 2-2	國外同時指標的構成項目	14
表 3-1	ACF 與 PACF 判定方法	31
表 3-2	預測能力的衡量指標	35
表 4-1	研究資料設置	36
表 4-2	研究樣本之敘述統計	37
表 4-3	各 IMF 之 Pearson 相關	38
表 4-4	個別獨立成份還原後的離差平方和比較表	45
表 4-5	各階段還原之擬偏 F 檢定結果	46
表 4-6	單根檢定之結果	47
表 4-7	Y_t^c 可能的 ARIMA 模型估計檢驗	49
表 4-8	Y_t^c 可能的 ARIMA 模型殘差檢定 Q 統計量	49
表 4-9	混合模型參數估計及檢定統計量	50
表 4-10	Y_t 可能的 ARIMA 模型估計檢驗	51
表 4-11	Y_t 可能的 ARIMA 模型殘差檢定 Q 統計量	52
表 4-12	傳統 ARIMA 模型參數估計及檢定統計量	52
表 4-13	混合 ARIMA 模型與傳統 ARIMA 模型之預測結果比較	54

圖目錄

圖 1-1	研究流程圖	5
圖 2-1	HHT 處理架構流程圖	7
圖 3-1	混合 ARIMA 預測模型架構圖	16
圖 3-2	本質模組函數圖例	17
圖 3-3	IMF 篩選過程（一）	18
圖 3-4	IMF 篩選過程（二）	19
圖 3-5	EMD 構成示意圖	21
圖 3-6	EMD 的適應性	22
圖 3-7	EMD 的完整性	23
圖 3-8	雞尾酒派對問題示意圖	25
圖 3-9	ARIMA 模型建構流程圖	30
圖 4-1	台灣工業生產指數歷史走勢圖(2000/01~2012/05)	36
圖 4-2	台灣工業生產指數之本質模組函數 IMFs 與趨勢分量 r	39
圖 4-3	原始序列、趨勢分量 r 與各殘餘量	40
圖 4-4	台灣工業生產指數的 6 個獨立成份	42
圖 4-5	原始訊號與第一步驟所還原的序列資料	44
圖 4-6	原始序列與去除雜訊序列比較圖	47
圖 4-7	Y_t^c 的 ACF 及 PACF 圖	48
圖 4-8	Y_t 之 ACF 及 PACF	51
圖 4-9	混合與傳統模型的預測序列圖	53

第壹章 緒論

第一節 研究動機與目的

一、研究動機

在過去研討時間序列預測相關議題時，於定量分析（Quantitative Analysis）的方法中，主要是透過歷史數據經由數理的方法進行評估，其大致上可分為傳統預測模型（如：迴歸分析、指數平滑、ARIMA 模型等）與電腦人工智慧學習（如：類神經網路、支援向量機等）；其中，前者主要在利用過去歷史資料特性，討論並運用不同的統計方法預估未來數據，後者則係利用近些年來發展出具學習特性的人工智慧，而將其運用在時間序列的預測上。

但無論是一般傳統時間序列預測或是電腦學習的方法，其同樣在面對資料（數據）呈現不良訊號時（高頻、雜訊、非平穩），經常無法完整準確表達出資料之變化及其趨勢，以致於在進行時間序列預測時，往往會喪失資料原本性質，因而於建立預測模型過程中會產生過度配適或配適不足，導致預測模型降低其可信的程度。基於上述原因，如何平穩資料（數據）以降低雜訊所帶來錯誤訊息，進而建立一可靠、可信的預測模型，是該領域一項重要的研究議題。

近年來 Haung et al. (1998) 提出非平穩且非線性時間序列的希爾伯特-黃轉換（Hilbert-Huang Transform, HHT）分析方法後，開始廣泛使用於自然科學的研究，如：風速、氣溫、潮汐、建築結構、生物醫學以及衛星訊號的變化等（Haung et al., 1999；Weng et al, 2006；Kamath et al, 2006；樊海濤等人, 2006；皮紅梅等人, 2007；吳順德等人, 2009）。其中的經驗模組分解法（Empirical Mode Decomposition, EMD），可將時間序列分解成數個本質模組函數（Intrinsic Mode Function, IMF），且每個本質模組函數具有不相等

的週期與頻率，並依照此特性作為平穩時間序列訊號之研究工具。其次，獨立成份分析法 (Independent component analysis; ICA) 是一訊號處理技術，方法最初的目的是用來處理未知來源分離 (Blind Source Separation, BSS) 的問題 (Karhunen et al., 1997)，近幾年有些學者開始將 ICA 應用在財務時間序列資料上 (Back & Weigend, 1997; Moody & Wu, 1997; Kiviluoto & Oja, 1998; Wu et al., 2006; 呂奇傑等人, 2007; 李泓緯, 2011)。

雖然目前已有相關文獻將 EMD 或 ICA 個別應用於預測領域上，但是完整結合此兩方法的研究則尚未見到。由於在選取 EMD 的 IMF 分量作為未來預測的變數時，往往係由研究者本人，透過自然環境的週期循環、文獻分析等主觀歸納方式以決定之，且 IMF 分量之間並不能保證彼此具獨立的貢獻性，各個 IMF 分量不但可能具相關性，且對於時間序列本身亦均隱含著解釋能力；而 ICA 則顯示出可透過重要的少數獨立成份，即可勾勒出原資料的結構與趨勢。

因此，本研究將於文中提出結合經驗模組分解法 (EMD) 與獨立成份分析法 (ICA) 的概念，而將一時間序列資料透過 EMD 分解後的數個本質模組函數 (IMF)，經由 ICA 將其獨立成份分解出來；同時，再加入顯著性檢定的概念，而將與原始序列具顯著相關的獨立成份保留（亦即剔除不重要的獨立成份）後還原資料，並視該還原後的新序列加上趨勢分量為已去除雜訊過後的時間序列資料，而進行後續混合預測模型建構的工作。

在台灣，同時指標綜合指數為衡量景氣的重要指標之一；其係由工業生產指數、電力（企業）總用電量、實質製造業銷售值、批發零售及餐飲業營業額指數、非農業部門就業人數、實質海關出口值、實質機械及電機設備進口值等七項所組成。其轉折點經常與景氣循環轉折點同步發生，故而可代表衡量當時的景氣狀況，亦即具有與景氣變動性質同步的指標。而綜合指標中的工業生產指數，則是回應經濟週期變化的重要標誌，係以工

業生產指數上升或下降的幅度，以衡量經濟復甦或者衰退的強度；無論是對公部門抑或私部門的政策決定者而言，皆是一個相當重要的決策影響因素。據此，研究中將以台灣工業生產指數為實證個案模擬，說明本研究所提之整合預測流程的適用性與可行性，以期能提供更可靠與實用的預測模型。

二、研究目的

因此，本研究的研究目的如下：

- (一) 藉由 EMD 分解後之具不同特性的本質模組函數 (IMF) 分量，結合獨立成份分析 (ICA) 法，設計出一套具檢定力、客觀且有效去除一時序資料雜訊的完整程序。
- (二) 解決當資料呈現不良訊號時，提出平穩並去除雜訊之時序資料訊號，並建立 ARIMA 混合預測模型，以期提升預測準確性。
- (三) 藉由分析歷年台灣工業生產指數資料，驗證在傳統時序資料分析中，除了經由差分轉換決定為平穩條件的另外一種方式。
- (四) 冀望透過實證方式建構台灣工業生產指數之單變量時序資料混合模型，以模擬與比較預測模型間的可靠性與實用性。

第二節 研究範圍

本研究主要為提出結合經驗模組分解法與獨立成份分析法，並提出顯著性檢定加以改良其方法，並以台灣工業生產指數為實證個案研究主題，資料來源取自行政院經濟建設委員會每月所公佈之景氣對策指標，蒐集的資料期間為 2000 年 1 月至 2012 年 5 月，合計 149 筆研究資料；其中，研究中以 2000 年 1 月至 2010 年 12 月的月資料序列為訓練樣本 (Training Set)，並據以建立預測模型；而以 2011 年 1 月至 2012 年 5 月的月資料序列為測試樣本 (Testing Set)，進行模型預測比較工作。

第三節 研究流程

本研究流程如圖 1-1 所示。首先，陳述本研究的研究動機與延伸出的研究目的；其次，針對目前台灣景氣同時指標的發展與其工業生產指數的定義作完整瞭解，並就經驗模組分解法與獨立成份分析法之應用相關議題的文獻進行回顧，據以提出加以改良，進而設計出一套移除時間序列資料雜訊的完整程序；接著，以台灣工業生產指數為實證個案，應用研究中所提的整合程序以移除資料雜訊，並據以建立其之 ARIMA 預測模型，加以分析與比較。

據此，本研究共分為五章。第一章為緒論，說明本研究之動機與目的，以確立出所要研究的範圍與研究流程。第二章為文獻回顧，依探討之主題找尋相關的文獻，並將國內外運用經驗模組分解法、獨立成份分析法於時間序列預測的文獻進行扼要回顧與整理，並針對目前台灣工業生產指數之定義作一詳述介紹。第三章為研究設計與方法，描述研究中所提出的結合經驗模組分解法及獨立成份分析法，以顯著性檢定法移除時間序列資料的雜訊，並應用於自我迴歸整合移動平均模型的分析架構與流程。同時，亦詳細敘述經驗模組分解法、獨立成份分析法、單根檢定、自我迴歸整合移動平均模型的基本原理。第四章為實證分析，以台灣工業生產指數為實證個案，配合研究目的，詳細說明分析與比較的結果。第五章則是結論與建議。

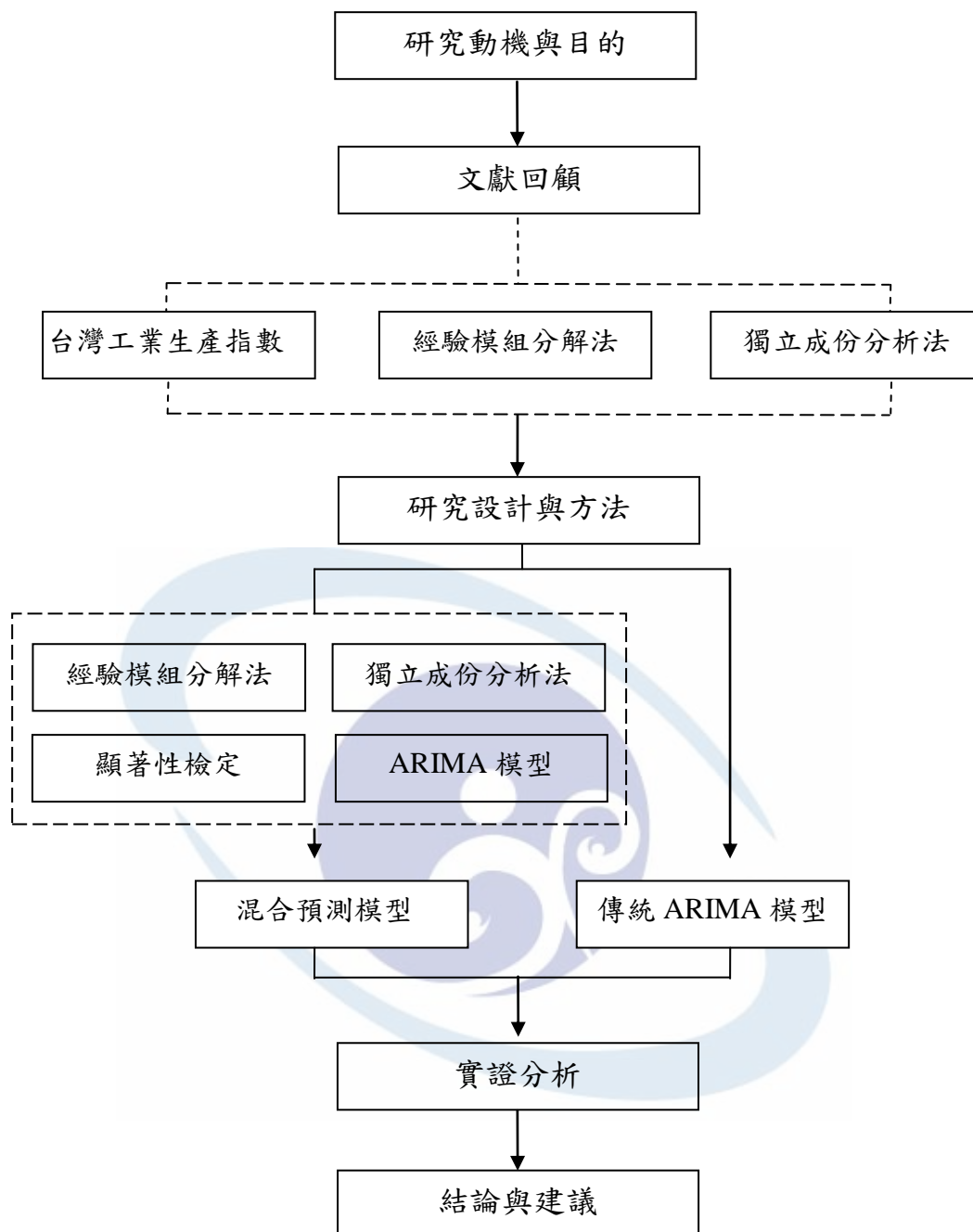


圖 1-1 研究流程圖

第貳章 文獻探討

研究中為能提出結合經驗模組分解法與獨立成份分析法以消除時間序列資料的雜訊，以進行個案預測模型模擬的實證研究；因此，本章首先分別就經驗模組分解法與獨立成份分析法之相關議題文獻進行回顧。此外，亦針對個案實證所使用的台灣景氣同時指標之工業生產指數的定義作一完整介紹。

第一節 經驗模組分解法(EMD)相關文獻

一、希爾伯特-黃轉換

非平穩且非線性時間序列的希爾伯特-黃轉換(Hilbert- Huang Transform, HHT)分析方法被提出之後(Haung et al., 1998)，開始廣泛地使用於自然科學資料的研究上，如：風速、氣溫、潮汐、建築結構、生物醫學以及衛星訊號的變化等(Haung et al., 1998；Weng et al., 2006；Kamath et al., 2006；樊海濤等人, 2006；皮紅梅等人, 2007；吳順德等人, 2009)。圖 2-1 即為 HHT 處理的分析架構流程圖；首先，是將訊號分解成有限個本質模組函數(IMFs)與一個均值趨勢分量或常數(Trend or Constant)的加總；其次，再透過 HHT 求得訊號的瞬時振幅與瞬時頻率，從中獲取訊號當中時間、頻率、能量的相關訊息。而 EMD 的篩選過程類似濾波(Filtering)的作用，是依據訊號時間與頻率透過篩選的迭代程序取得 IMFs；由於 IMF 會隨訊號而改變，實為一組適應性的基底，故而能處理不同時間訊號（包含非線性與非穩態訊號）(Haung et al., 1998)。詳細的 EMD 分解過程，將於研究設計與方法章節中介紹之。

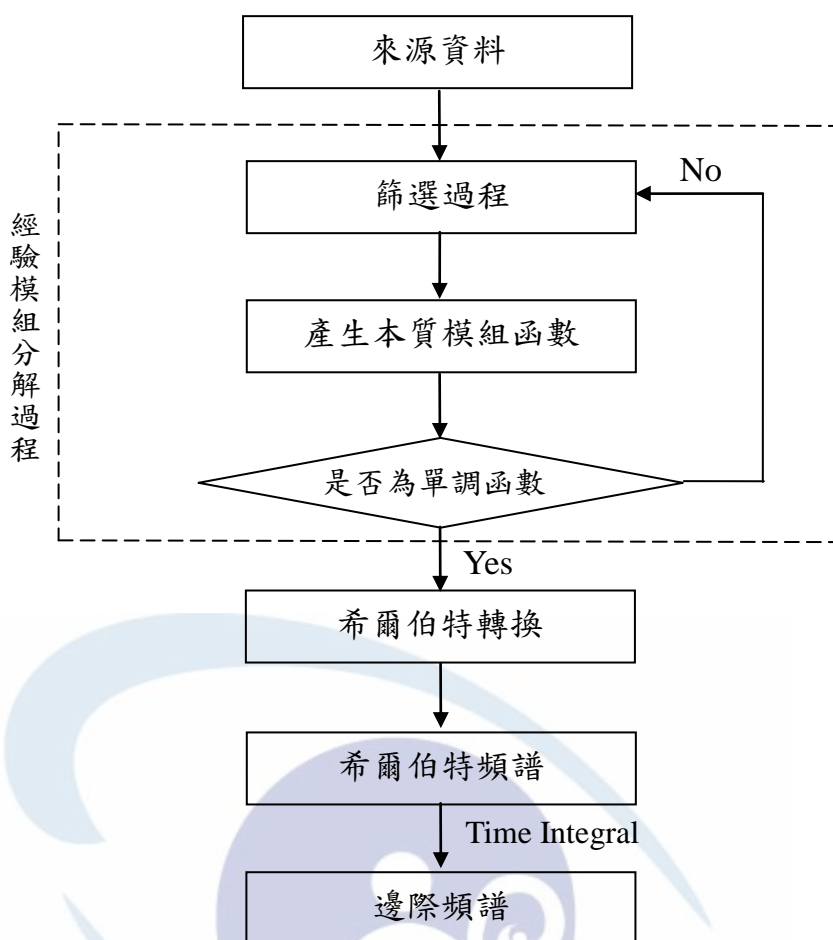


圖 2-1 HHT 處理架構流程圖

資料來源：Huang et al., 1998。

二、運用希爾伯特-黃轉換於時間序列預測

在運用希爾伯特-黃轉換於時間序列預測的文獻或議題中，Salisbury & Wimbush (2002) 以南方震盪指數進行 EMD 分析，研究中選擇第四個 IMF 分量作為預測訊號，並預測聖嬰及反聖嬰現象。楊文熙 (2003) 將蒐集得到的台灣加權指數、道瓊工業指數 (Dow Jones industrial average index)、那斯達克指數 (Nasdaq index) 及其報酬率資料進行 EMD 分析，研究結果指出透過重要分量 IMF 的組合，將可作為指數與報酬率的預測。曹麗青、林振山 (2008) 則於其研究中，去除最高頻率的第一個 IMF 分量，以預測中國三峽水庫的年平均流量值。Yu et al. (2008) 則將資料先以 EMD 分解，並應用前饋式類神經網路 (Feedforward Neural Networks, FNN) 預測美國西

德州中級原油（West Texas Intermediate, WTI）與英國布蘭特（Brent）原油的期貨價格。Wang et al.（2009）則是結合支援向量迴歸（Support Vector Regression, SVR）建立預測模型，以解決建立金融時間序列高波動且複雜的資訊。葉清江等人（2011）則運用 EMD 對台股資料做前置處理，並混合倒傳遞類神經網路（Back-Propagation Neural Network, BPN）模式，結果顯示其預測績效優於單一的 BPN 預測模型；Chen et al.（2012）同樣利用 EMD 結合 BPN 模式，預測日本及港澳來台人數。Wei & Chen（2012）則先行應用 EMD 分解後的 IMFs，並以 Pearson 及 Kendall 等級相關係數與原始序列進行檢定，並以兩檢定法的相關係數大小與顯著性，區分成重要與不重要程度兩類別，並作為建構 BPN 模式參數輸入的變數之一，以預測台北捷運木柵線的乘客流量，研究顯示，混合 EMD 與 BPN 的模型要較其他模型更有效性。

綜合以上國內外文獻可以發現，運用希爾伯特-黃轉換於時間序列預測時，能有效降低時間序列訊號之非平穩及非線性性質，然而大多數研究者在使用經驗模組方法分解出 IMF 分量選擇作為未來預測的變數時，所採行選取變數的方式往往係由研究者本人，透過自然環境的週期循環、文獻分析等主觀歸納方式以決定之，鮮少探討各個 IMF 分量間的相關性，加上 IMF 分量之間並不能保證彼此具獨立的貢獻性，各個 IMF 分量不但可能具相關性，且對於時間序列亦均隱含著解釋能力，故而值得加以探究並實證之。

第二節 獨立成份分析法(ICA)相關文獻

一、獨立成份分析（Independent Component Analysis）

獨立成份分析法（Independent Component Analysis, ICA）是一訊號處理技術，其某些相關理論係包含在資訊及統計等領域之中，方法最初的目的的是用來處理未知來源分離（Blind Source Separation, BSS）的問題（Karhunen

et al., 1997)，其中用以描述 BSS 問題的最著名例子為雞尾酒派對問題（Cocktail-party problem），其即為 ICA 所要處理的問題（Lee, 1998）。由於該分析法計算極為繁複，直到 Hyvärinen（1999）提出 FastICA，其計算效率大為提高，且能有效解決資料在統計上分配未知的問題，ICA 方才開始大量地被應用於各個領域上（Mansour & Kawamoto, 2003）。

二、運用獨立成份分析的相關議題

近幾年有些研究者開始將 ICA 應用於財務時間序列資料上，使得隱藏在資料裡的結構能夠更輕易的被觀察出來；同時，亦發現時間序列資料動態過程中，只需要幾個重要的獨立成份即可解釋之（Back & Weigend, 1997）。Moody & Wu（1997）應用 ICA 將國際外匯市場的時間序列資料之雜訊分離出來，以提高未來預測的準確度。Kiviluoto & Oja（1998）則由 40 家零售店的現金流量找出其獨立成份，且發現相近零售店的獨立成份變動情形有較相近的趨勢。呂奇傑等人（2007）則以日經 225 指數為例，應用 ICA 得到去除雜訊的預測變數後，再以 SVR 建立其預測模型，發現如此將具有良好的預測能力。Wu et al.（2006）則是將 ICA 應用於 GARCH 模型中，並進行各項金融指數風險值的評估；研究顯示，其在捕捉隨時間變動的波動性上優於其他比較之模型。李泓緯（2011）則先應用 ICA 處理美國道瓊工業指數、香港恆生指數以及台灣加權指數的雜訊，再以分類決策樹建立規則；結果顯示，其不僅各種績效衡量良好外，亦優於其他的傳統方法。

由前述獨立成份分析法的相關文獻可以發現，ICA 能有效消除時間序列中之雜訊，並且能透過重要的少數獨立成份，即可勾勒出原資料的結構與趨勢。因此，本研究亦將此方法納入研究中，同時再加入顯著性檢定的概念，以提高模式未來預測的準確度。

第三節 工業生產指數的定義

一、台灣景氣同時指標之發展

台灣景氣指標係由行政院經濟建設委員會自 1977 年起按月發布，其中包含了景氣綜合指數與景氣對策信號。目前景氣綜合指數有領先指標、同時指標及落後指標，而落後指標雖於 1984 年起暫停編製，但自 2010 年 8 月起，於現有領先指標、同時指標及景氣燈號的基礎上，一併發布落後指標，使得整體景氣指標系統更臻完備，同時亦能與國際接軌。

領先指標可預知景氣高峰谷底的發生，同時指標則為同步反映當期景氣的變化，而落後指標可於事後驗證或確認領先指標及同時指標在過去的計算是否正確。若依構成指標的因子多寡劃分，則可區分為以下兩類：

- (一) 綜合指標 (Composite Indicators)：如領先指標 (Leading Indicators)、同時指標 (Coincident Indicators) 等。
- (二) 單一指標 (Single Indicators)：如工業生產指數 (Industrial Production)、股價指數 (Stock Price Index) 等。

台灣的景氣指標自 1977 年發布迄今，歷經 1978 年、1979 年、1987 年以及最近一次修訂之 2007 年 7 月（如表 2-1），先後共 4 次的修訂。有鑒於台灣社會經濟結構改變，服務業比重日益上升，以及經濟成長趨勢明顯；因此，重新檢討同時指標後，於 2007 年 7 月參考經濟合作與發展組織 (Organisation for Economic Co-operation & Development, OECD) 在國際間推廣的景氣指標編製方法進行修訂，並重新篩選指標之構成項目，合併成新的同時指標綜合指數，以成長循環的概念明確分離出「長期趨勢 (Trend)」與「循環波動 (Cycle)」兩部份，有利清楚捕捉景氣循環波動轉折點，以期景氣指標確切反映台灣經濟景氣的脈動（許秀珊，2008；徐志宏，2009）。

表 2-1 台灣同時景氣指標修訂構成項目

1977年 景氣擴散指數 (Diffusion Index)	1978年 景氣綜合指數 (Composite Index)	1987年 景氣綜合指數 (Composite Index)	2007年 景氣綜合指數 (Composite Index)
1.工業生產指數變動率 2.製造業生產指數 3.製造業銷售值 4.票據交換變動率 5.國內貨運量 6.退票率	1.工業生產指數變動率 2.製造業生產指數變動率 3.製造業銷售值 4.票據交換變動率 5.國內貨運量 6.退票率 7.製造業平均每月薪資變動率 8.全體銀行等儲蓄性存款變動率	1.工業生產指數變動率 2.製造業生產指數變動率 3.製造業銷售值 4.票據交換金額變動率 5.國內貨運量 6.製造業平均每月薪資變動率	1.工業生產指數 2.電力（企業）總用電量 3.實質製造業銷售 4.非農業部門就業人數 5.實質海關出口值 6.實質機械及電機設備進口值 7.批發、零售及餐飲業營業額指數

資料來源：本研究整理

二、台灣工業生產指數

在台灣，工業生產指數涵蓋礦業及土石採取業、製造業、水電燃氣業及房屋建築業的實質產出，資料由 1960 年起，長期以來均可捕捉整體經濟波動，由於工業生產與整體經濟密切相關，對經濟變化十分敏感，故雖然在台灣景氣同時指標先後共經過 4 次的修訂，但工業生產指數總能確切反映台灣經濟景氣在生產面的脈動，許多國家也以此做為同時指標中隸屬於生產類別之構成項目（表 2-2）。

台灣地區工業生產統計，創始於 1953 年，其目的在於建立工礦業產銷查報制度，並按月編製工業生產指數。為使指數更具充分代表性，根據經濟的變動，每隔 5 年將更換基期與權數，增加並抽查重要及新興產品，並且編製多種的複分類指數，以提高統計用途。其目的與用途在於衡量工業產銷存量消長情勢與經濟建設成果，不僅提供給各行各業調節生產、經貿，亦提供給公部門施政計畫之參考依據。

工業生產調查的週期為每月調查與編製一次，調查對象包括：礦業及土石採取業、製造業、電力及燃氣供應業、用水供應業及建築工程業等 5 大行業，均以企業單位為調查對象。其調查項目依照台灣地區當前工業生產結構，選取具有重要性、代表性、領導性以及策略性之產品，合計選查 2,657 項產品（其中製造業 2,588 項、礦業及土石採取業 32 項、電力及燃氣供應業 9 項、用水供應及資源回收處理業 9 項以及建築工程業 19 項）。其中，礦業及土石採取業、用水供應業、建築工程業以及公營企業單位採行全部調查，其餘則視各企業對象的規模及行業別，採用不同的抽出法，其全部調查與抽查樣本家數，約計 8,000 家。礦業及土石採取業係採用經濟部礦務局資料；建築工程業則採用內政部營建署統計之各縣市核發使用執照總樓地板面積資料；而公民營製造業、電力及燃氣供應業以及用水供應業，則係由受查企業自行填報（經濟部工業生產統計年報，2012）。

調查的產品則依據 2011 年經濟部編印之工業產品分類調查（共 2,657 項），再將相同性質產品整併為 628 項產品群，並將其編列報表，並以 2006 年為固定基期，編成各類固定基期指數；計算基期選取的不同會產生不同的差異，台灣工業生產指數則是採用拉斯皮爾式（Laspeyres）基期加權總值式，其計算期工業生產指數公式如下：

$$I_{0i} = \frac{\sum Q_i P_0}{\sum Q_0 P_0} \times 100 \quad (2-1)$$

式中 Q_i ：計算期生產量， Q_0 ：基期生產量， P_0 ：基期生產淨值單價。



表 2-2 國外同時指標構成項目

國家	生產類別	銷售類別	就業及所得類別	貿易及其他類別
美國	1.工業生產指數	1.製造業及商業實質銷售值	1.非農業部門受雇員工人數 2.扣除移轉性支付之實質個人所得	
英國	1.工業生產指數	1.零售銷售	1.就業人數 2.家計單位可支配實質所得	
德國	1.工業生產指數	1.實質製造業銷售值 2.實質零售業銷售	1.就業人數	
法國	1.工業生產指數	1.零售銷售	1.受雇員工人數	1.實質進口值
日本	1.工業生產指數 2.工業大戶用電量 3.製造業產能利用率	1.工業生產財銷售指數 2.投資財銷售指數零售業銷售值年增率 3.批發業銷售值年增率 4.企業營業利潤 5.製造業中小企業銷售值	1.製造業加班工時指數 2.有效求才倍數	
韓國	1.工業生產指數 2.製造業營運比指數 3.實質營造完工金額	1.躉售及批發銷售指數	1.非農業就業人數	1.實質進口值 2.服務業活動指數 3.國內市場出貨比率
新加坡	1.工業生產指數 2.實質GDP	1.零售銷售指數	1.總就業人數	1.不含石油實質出口

資料來源：許秀珊，2008。

第參章 研究設計與方法

除了描述研究中所提之結合 EMD 與 ICA 以消除時間序列資料雜訊，並進行個案預測模型建立的分析架構外，本章節亦將介紹實證研究過程中所使用到的方法之基本原理，其中包含經驗模組分解法 (EMD)、獨立成份分析法 (ICA) 以及自我迴歸整合移動平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) 等。而本研究在進行經驗模組分解法時，所使用的 Matlab 程式碼是由國立中央大學數據分析中心所撰寫；獨立成份分析法的 FastICA 程式部份，是由阿爾托大學理工學院 (Aalto University School of Science and Technology) 所提供。而自我迴歸整合移動平均模型則使用 Eview 分析軟體操作。

第一節 結合 EMD 與 ICA 於預測模型的架構

不同於過去時間序列研究中，透過 EMD 分解過後選擇 IMF 分量的方式；本研究認為各個 IMF 分量不但可能彼此間具相關性，同時對於時間序列亦均隱含著解釋能力；因此，若直接去除一個或數個 IMF 分量後以建立預測模型，可能會喪失許多重要的資訊。因此，研究中提出結合獨立成份分析法 (ICA)，而將各個 IMF 分量轉換成獨立成份，亦即將各個 IMF 分量相似之特徵歸類為同一個獨立成份，同時各個獨立成份彼此之間為獨立且無關；之後，再以顯著性檢定方式，選取保留重要的獨立成份並將其還原後與趨勢分量 r 相互結合，視其為已去除雜訊後的時間序列資料。因此，研究中結合 EMD 與 ICA 的混合預測模型建立，其架構如圖 3-1 所示，以下將依流程說明之：

- 一、蒐集所欲研究與建立模式的時間序列資料。
- 二、應用經驗模組分解 (EMD) 將資料為分解成 n 個本質模組函數 (IMFs) 以及趨勢分量 r 。
- 三、將 IMFs 透過 ICA 過程後，以顯著性檢定法 (研究中將提出「擬偏 F

統計量」)保留具貢獻的重要獨立成份，並還原序列後加上趨勢分量 r ，成為已去除雜訊後的時間序列資料。

四、建立結合 EMD 與 ICA 的混合預測模式。

五、進行混合 ARIMA 預測模型的預測與比較。

為能說明分析架構的可行性與可靠性，研究中將應用台灣工業生產指數資料為個案，建立混合 ARIMA 預測模型以驗證與比較之。

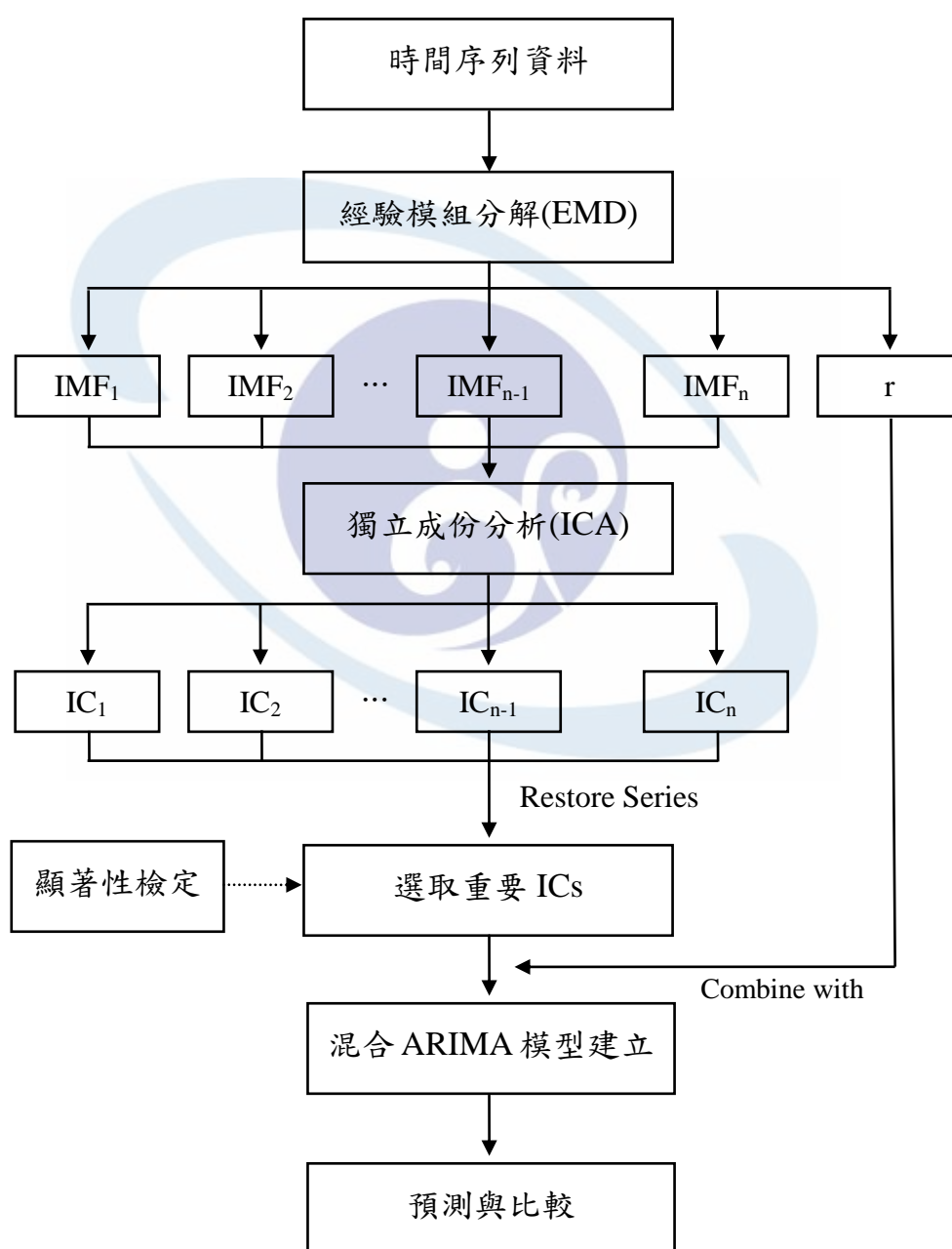


圖 3-1 混合 ARIMA 預測模型架構圖

第二節 經驗模組分解

一、本質模組函數

Huang et al. (1998) 應用經驗模組分解 (EMD) 將時間序列資料分解成數個本質模組函數 (IMF)，不同的本質模組函數間則具有不盡相同的週期與頻率，也具有不盡相等的能量或權重。EMD 的分解是依據下列的三項假設：

- (一) 整個資料必須至少有兩個以上的極值（一個極大值與一個極小值）。
- (二) 時間尺度 (Time Scale) 係由兩個極值的時間推移 (Time Lapse) 所定義。
- (三) 如果資料本身缺乏極值，但反曲點 (Inflection Point) 存在時，則可將其微分一次或多次以上，直至出現極值為止。

圖 3-2 即為 Huang et al. (1998) 所提出有關風速研究中的一個典型的 IMF。

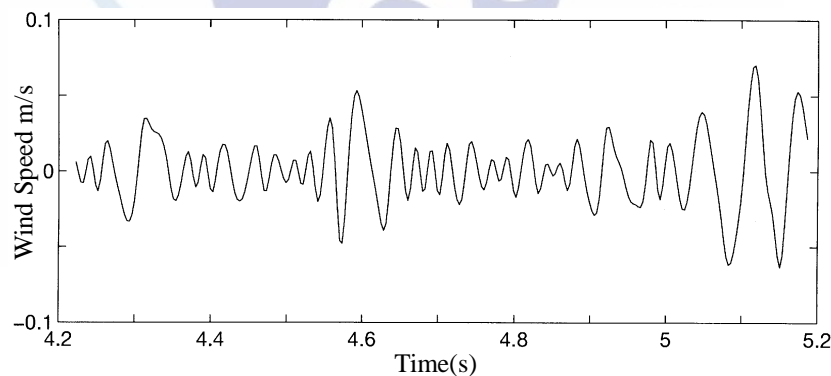


圖 3-2 本質模組函數圖例

資料來源：Huang et al., 1998。

二、經驗模組分解過程

如前所述，經驗模組分解 (EMD) 主要是藉著不斷重複的篩選過程 (sifting process)，逐步地找出 IMF 分量。現若以時間序列 $Y(t)$ 為例，則其篩選程序的過程如下：

步驟1：找出時間序列 $Y(t)$ 中的所有局部極大、極小值。其次，分別將該訊號的局部極大值與局部極小值，應用三次仿樣曲線（Cubic Spline）連接成上包絡線與下包絡線；並且求出上、下包絡線的平均值，即可得到均值包絡線 $m_1(t)$ 。

步驟2：將原時間序列與均值包絡線相減，將可得到第一個預備分量 $h_1(t)$ 。
亦即

$$h_1(t) = Y(t) - m_1(t) \quad (3-1)$$

圖 3-3 即為 EMD 分解過程中，步驟 1 及步驟 2 的實際流程。其中，圖 (a) 為原始序列 $Y(t)$ ；圖 (b) 中的虛線為上包絡線與下包絡線，粗體實線為其包絡線均值 $m_1(t)$ ；圖 (c) 則為求得之預備分量 $h_1(t)$ 。

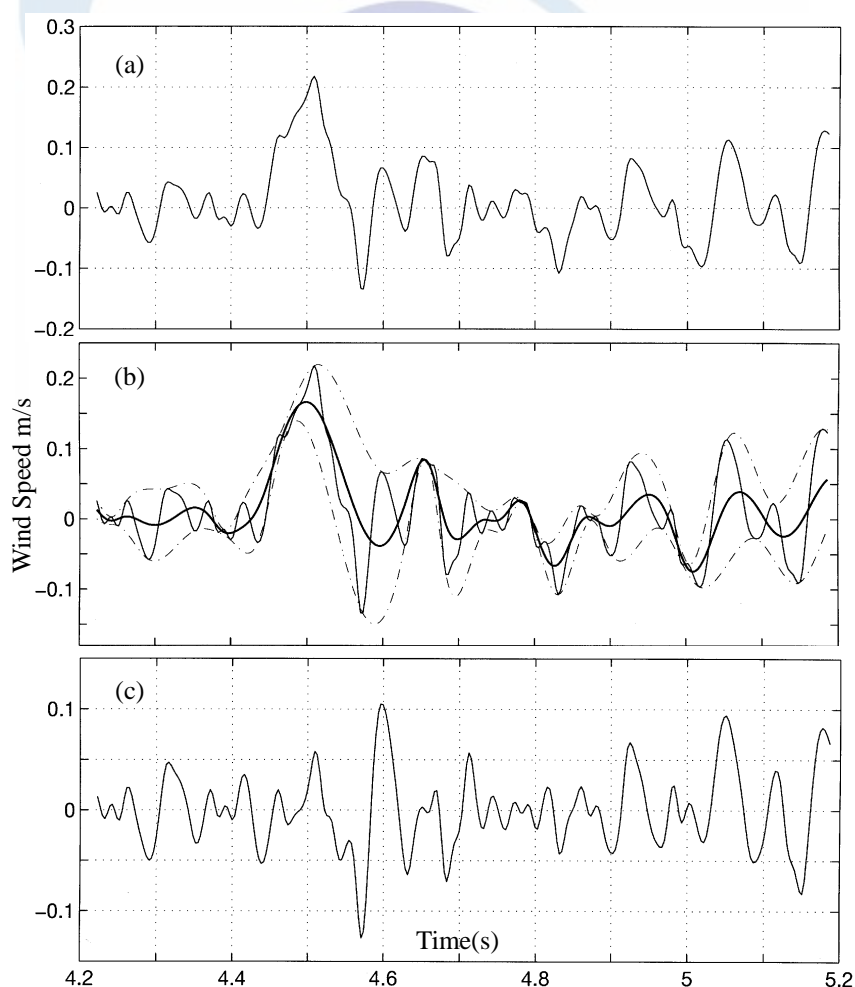


圖 3-3 IMF 篩選過程（一）

資料來源：Huang et al., 1998。

步驟3：檢查 $h_1(t)$ 是否符合IMF的條件。如果不符合，則回到步驟1，並將 $h_1(t)$

當作原訊號，進行第二次篩選工作。亦即

$$h_2(t) = h_1(t) - m_2(t) \quad (3-2)$$

如此重複篩選 k 次，亦即

$$h_k(t) = h_{k-1}(t) - m_k(t) \quad (3-3)$$

直到 $h_k(t)$ 符合IMF的條件，即可得到第一個IMF分量 $C_1(t)$ 。

圖 3-4 即為 EMD 分解過程中，步驟 3 的實際流程；其中，圖(a)為 $h_2(t)$ ；圖(b)為 $h_3(t)$ 。其次，在運算 IMF 的過程中，為避免因篩選次數過多而破壞其原物理特性，一般來說，準則的建立大多是聚焦在振幅、頻率的不相關性以及相位上。

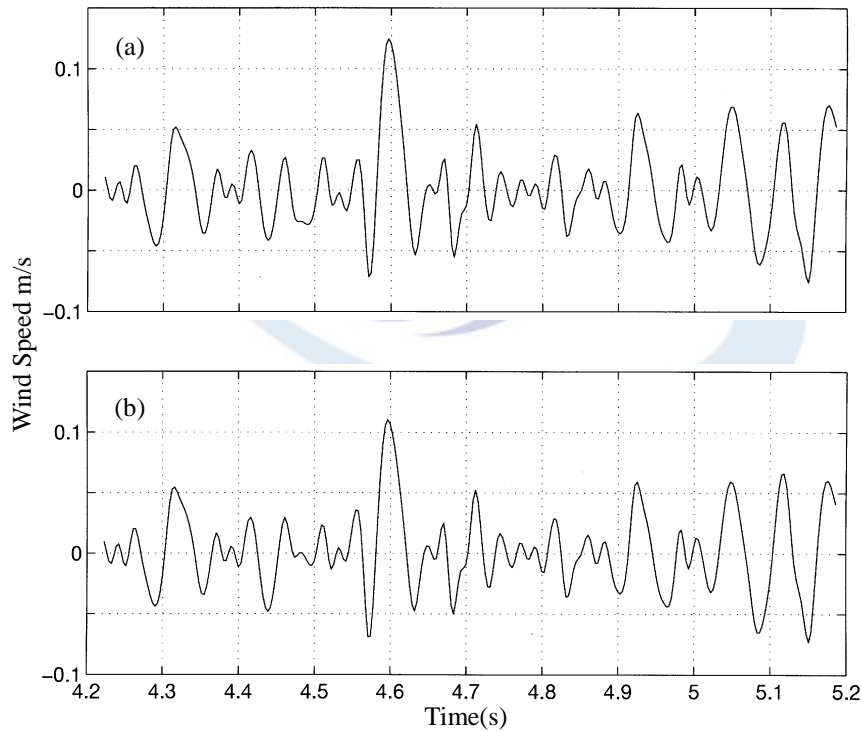


圖 3-4 IMF 篩選過程（二）

資料來源：Huang et al., 1998。

而根據 Huang et al. (1998)的建議，步驟 3 的停止標準有以下兩項：

（一）當數據通過0的個數與極值個數相同時，即可停止。

(二) 當 SD_k 小於前一次的 SD_{k-1} 時，則篩選的過程即告停止；反之，若大於則繼續。而 SD_k 其實就是標準差的運算，即函數與函數之間波動變化量大小的比較，其中 SD_k 的變化量若小於前一次時，其意味著函數會開始愈來愈收斂；典型的 SD_k 大小在 0.2 至 0.3 間，故而訂此為停止標準。而 SD_k 的計算公式如下：

$$SD_k = \frac{\sum_{t=0}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^T h_{k-1}^2(t)} \quad (3-4)$$

步驟4：將原訊號減去第一個 IMF 分量，則可得到第一個殘餘量 (Residue)，如下式：

$$R_1(t) = Y(t) - C_1(t) \quad (3-5)$$

步驟5：將殘餘量 $R_1(t)$ 當作新資料，重複步驟1到步驟4，將可得到第二個殘餘量 $R_2(t)$ 。

如此，重複 n 次，則可得

$$R_2(t) = R_1(t) - C_2(t)$$

$$R_3(t) = R_2(t) - C_3(t)$$

⋮

$$R_n(t) = R_{n-1}(t) - C_n(t) \quad (3-6)$$

當第 n 個殘餘量 $R_n(t)$ 為單調函數 (Monotonic Function)，而無法再分解 IMF 時，則 EMD 的分解過程即告完成 (圖 3-5)。

此時，若將所得到的 IMF 各個分量與最後一個殘餘量進行加總時，則將可得到原有的時間序列 $Y(t)$ 。由於各個 IMF 分量是依據訊號時間與頻率並透過篩選的迭代程序取得 (即 IMF 會隨訊號改變)，故其為一組具有適應性的時間序列資料。(Haung et al., 1998)。

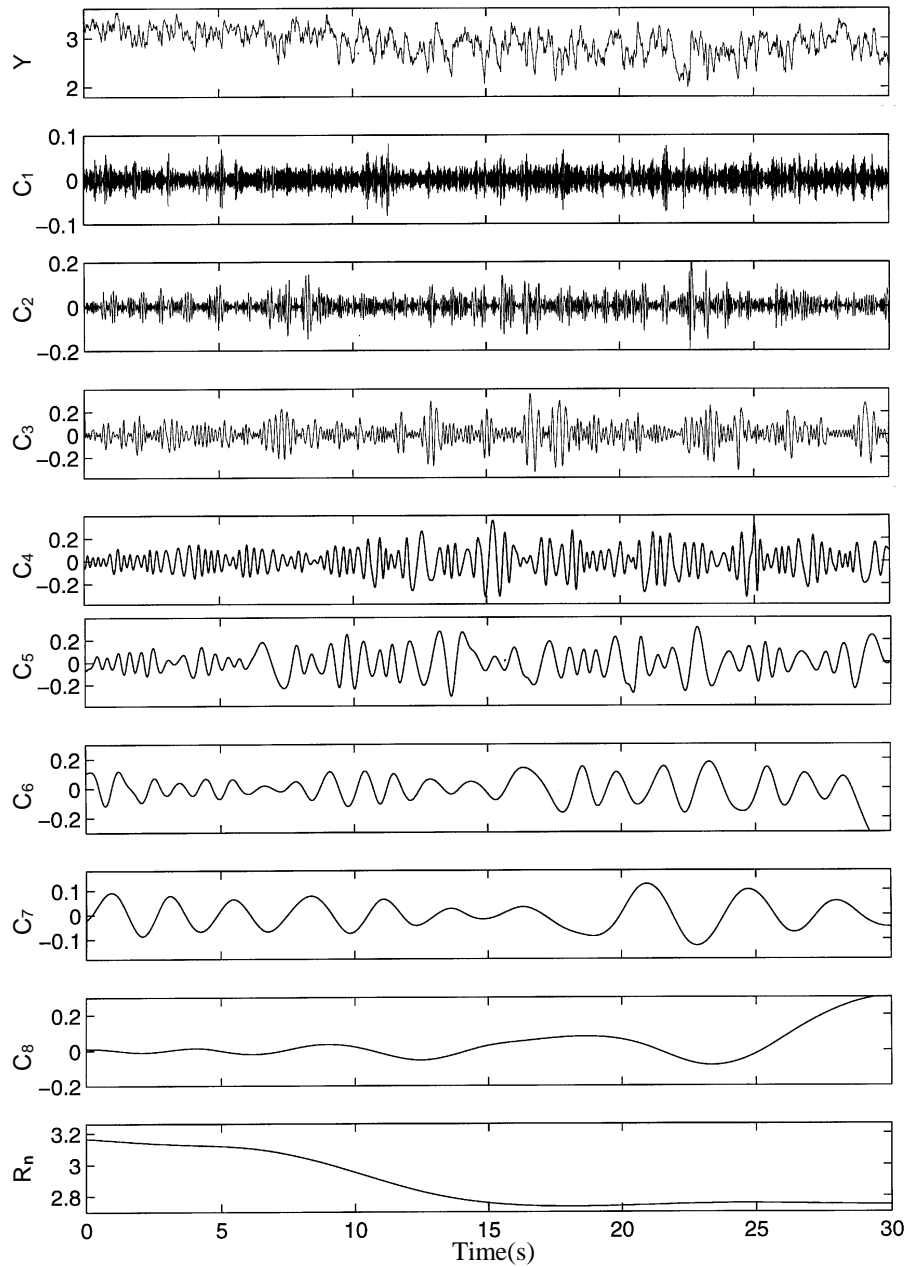


圖 3-5 EMD 構成示意圖

資料來源：Huang et al.，1998。

圖 3-6 為 EMD 過程中各個殘餘量，而 IMF 分量與原訊號的關係即是將最後一筆殘餘量與各個 IMF 分量加總後，即可得到原訊號（圖 3-7），顯示其完整性。其若以數學式表之，則如下所示：

$$Y(t) = C_1(t) + C_2(t) + \cdots + C_n(t) + R_n(t) \quad (3-7)$$

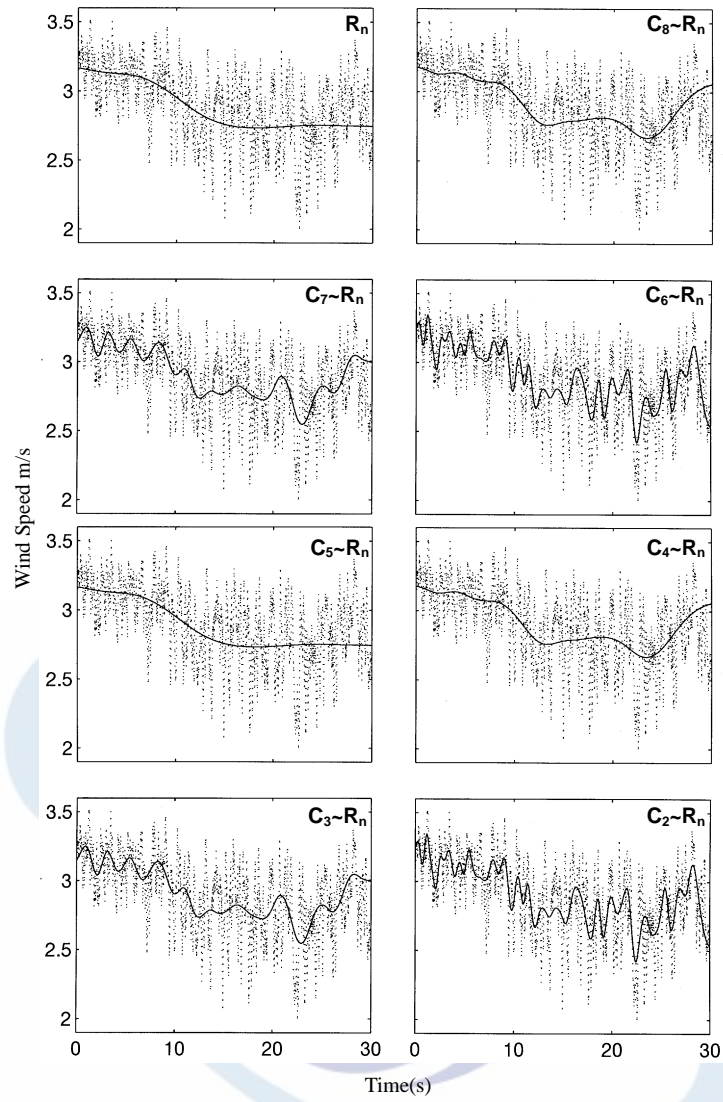


圖 3-6 EMD 的適應性

資料來源：Huang et al.，1998。

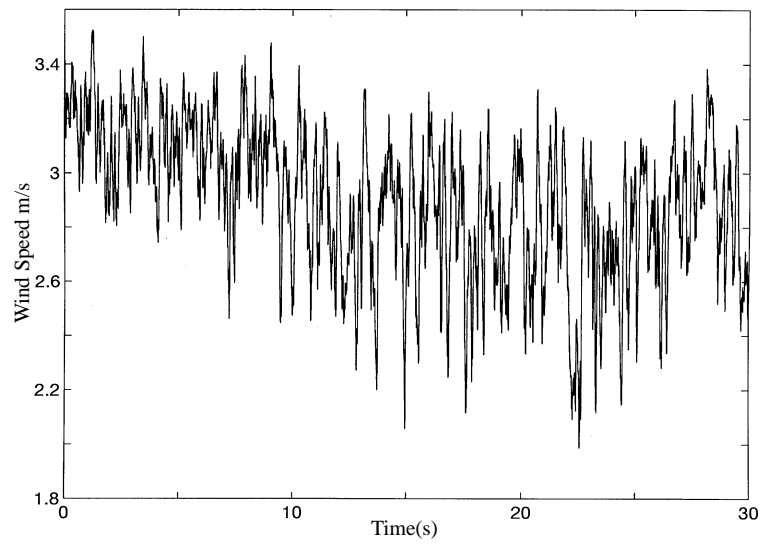


圖 3-7 EMD 的完整性

資料來源：Huang et al.，1998。



第三節 獨立成份分析

一、基本原理

獨立成份分析 (ICA) 常與主成份分析 (Principal Component Analysis, PCA) 相互進行比較，兩者主要在於假設的不同；PCA 係假設混合資料滿足高斯分配，且各個主成份的係數向量必為直交；而 ICA 的定義是假設有 n 個互相獨立的數列，在受到諸多因素的干擾而形成彼此具有相關性的混合訊號，各個獨立成份 (ICs) 為相互獨立，但並非滿足高斯分配。

獨立成份分析最常被應用於處理未知來源訊號分離 (Blind Source Separation, BSS) 的問題 (Karhunen & Hyvärinen, 1997)。此種訊號處理問題又被稱為雞尾酒派對問題 (Cocktail-Party Problem)。所謂的「雞尾酒派對問題」，是指在一個充斥各種交談與音樂的宴會中，雖然周遭夾雜不同的聲音，但人們卻仍可專注的聽到自己想聽的聲音，此種生物的特殊機制則為大腦與神經交互作用的結果；因此，ICA 一開始係發展於有關通訊與神經生理學方面；其假設所觀察到的訊號是由潛在來源訊號 (Latent Variables) 以線性方式組合而成，「未知」指的就是對於潛在來源訊號與其線性混合機制的資訊所知不多。而獨立成份分析透過假設潛在來源訊號彼此之間具有統計獨立的性質，從觀測到的訊號分離出獨立成份 (Independent Components, ICs)，並以分離出之獨立成份來估計潛在來源訊號；而這樣的統計分析方法在許多領域上皆成功的捕捉到資料的基本結構精神，所以 ICA 成為近年來非常熱門的訊號分離技術 (Hyvärinen, 1999)。

以下本研究即透過雞尾酒派對問題來說明 ICA 的最初概念：假設一個房間之內有三個人在同時說話，三位演講者的聲音訊號分別為 $s_1(t)$, $s_2(t)$, $s_3(t)$ ；而這時在房間的後面有兩支麥克風，分別放在不同的位置進行收音，可得到三個隨時間變化的聲音訊號，分別以 $x_1(t)$, $x_2(t)$, $x_3(t)$ 表示，如圖 3-8 所示，並且假設 $x_1(t)$, $x_2(t)$, $x_3(t)$ 為來源訊號 $s_1(t)$, $s_2(t)$, $s_3(t)$ 經由線

性組合而成，其數學式如下：

$$\begin{aligned}x_1(t) &= a_{11}s_1 + a_{12}s_2 + a_{13}s_3 \\x_2(t) &= a_{21}s_1 + a_{22}s_2 + a_{23}s_3 \\x_3(t) &= a_{31}s_1 + a_{32}s_2 + a_{33}s_3\end{aligned}\quad (3-8)$$

若以矩陣表示，則可假設觀測到的混合訊號矩陣為 X ，各組來源矩陣為 S ，與距離有關的混合矩陣則為 A ，則此三者的關係如下：

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = AS \quad (3-9)$$

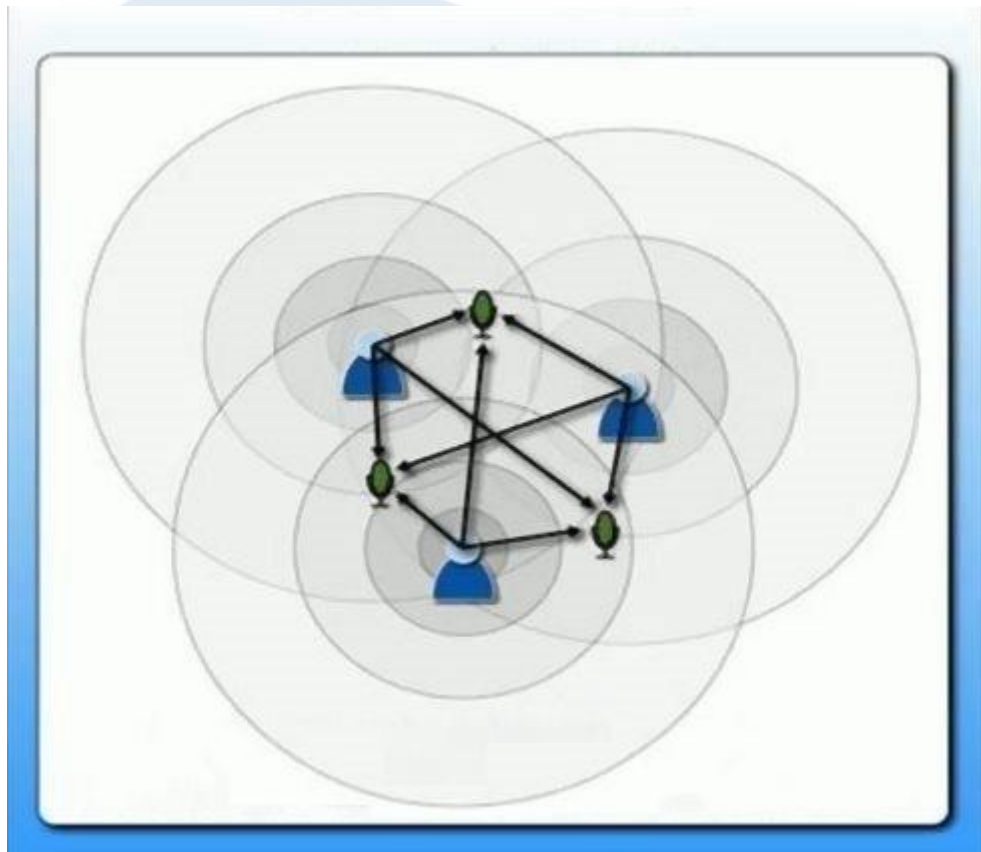


圖 3-8 雞尾酒派對問題示意圖

資料來源：本研究整理。

在假設來源聲音訊號 S 未知的情況下，如果混合矩陣 A 已知，則透過線性方程組的方式求解，將可順利的求出原先演講者聲音訊號 S ，亦即

$S = A^{-1} \cdot X$ ；但是雞尾酒派對問題在於混合矩陣 A 與來源矩陣 S 均是未知的，因而在面對這樣的難題，ICA 的基本概念在於訊號統計獨立的特徵，運用統計分析方法估計出矩陣 A' 用以近似未知的混合矩陣 A ，再利用 A' 之反矩陣求出 W ；而在 ICA 中的矩陣 W 稱為解混合矩陣 (De-mixing Matrix)，其係用來將混合數列還原為近似於最初狀態的獨立矩陣 $Y = W \cdot X$ ，進而將獨立成份 Y 用以估計未知潛在來源訊號 S 。以矩陣表示如下：

$$Y = W \cdot X \approx S \quad (3-10)$$

$$S \approx Y = W \cdot X \Rightarrow W \approx A^{-1} \quad (3-11)$$

二、ICA 基本模型

如前所述， $X=AS$ 即為 ICA 的基本數學模型。其中，混合矩陣 A 與來源矩陣 S 均屬未知；而為了有效的分離混合訊號與辨識出來源訊號 S ，則需求滿足下列兩項基本假設：

- (一) 來源訊號必須相互獨立。
- (二) 來源訊號最多只能允許一個以下服從高斯分配。

以上的假設主要是為了引用中央極限定理：「非高斯分配且互為統計獨立的隨機變數之和，會比任一組成之獨立隨機變數的分配更接近於高斯分配」。而 ICA 即透過對中央極限定理的逆向思考：「基於觀察到的混合訊號是原來未知來源訊號的線性組合，如果我們想要去找出原來互為統計獨立的來源訊號時，則可以尋找最不具備高斯分配的成份，其為來源訊號的可能性將是最大」 (Hyvärinen & Oja, 2000)。經由以上關於中央極限定理的說明後，可以瞭解 ICA 為了要藉由來源訊號互相獨立的性質，從混合訊號 X 中分離出來源訊號 S ，必須發展有效衡量非高斯特性的準則以作為目標函數。總而言之，ICA 最主要可分為兩個部分：

- (一) 利用非高斯特性量測準則發展目標函數。
- (二) 利用目標函數搜尋最佳解混合矩陣 W 。

而有關促使目標函數達到最佳化的演算法，在過去文獻中有許多選擇，以下為三種 ICA 的演算法類型：

- (一) 高階統計量法 (Cardoso & Souloumiac, 1993)
- (二) 共同資訊理論 (Bell & Sejnowski, 1995; Yang & Amari, 1997)
- (三) 獨立成份之非高斯分配特性 (Hyvärinen, 1999)

上述求取獨立成份的演算法都不盡相同，其中以學者 Hyvärinen (1999) 提出之 FastICA 演算法，由於其計算效率高，成為處理 ICA 問題最常見的演算方法之一。而本研究於後續實證中亦採用 FastICA 演算法找出相互獨立的成份。

三、FastICA 演算法

(一) 選定目標函數

在 FastICA 演算法中用以衡量非高斯特性的目標函數，是從熵(Entropy)的概念所延伸出來。熵為資訊理論(Information Theory)中用以衡量某一隨機變數資訊量的單位，其主要用意在於將一般不確定性問題中所包含「不確定」(Uncertainty) 的程度以數學定量描述之。舉例而言：投擲一枚骰子與一枚硬幣，由於骰子的可能結果多於硬幣的可能結果，因此相較起來，骰子的不確定性會高於硬幣的不確定性，此種不確定性數值即是各自計算出骰子與硬幣的熵來比較之；亦即表示該隨機變數的觀察值不確定性程度的高低，如果該隨機變數的分配越無法預測、結構性越差，則其熵也會越大。若假設隨機向量 y 其機率密度函數為 $p(y)$ ，則隨機向量 y 的熵定義如式 (3-12)：

$$H(y) = -\int p(y) \log p(y) dy \quad (3-12)$$

有了熵的初步概念後，根據熵最大化定理以及熱力學第二定律，可知在所有具有相同變異數的隨機分配中，呈現高斯分配的隨機變數具有最大的熵 (Lee, 1998)，亦即高斯分配是所有分配中最不具規則性與結構性的分

配；換言之，若一隨機變數其分配為非高斯分配，則其熵將會小於具高斯分配的隨機變數的熵。但為了所處理的資料與高斯分布之間熵的差異達到極大化，於是將熵的定義作進一步的修改，而成為負熵（Negentropy），隨機向量 y 的負熵 $J(y)$ 定義如下：

$$J(y) = H(y_{\text{gauss}}) - H(y) \quad (3-13)$$

上式的 $J(y)$ 為高斯分布與其他分配之間熵的差異。但若欲求出熵，必須知道樣本資料的統計分配，就當前的研究方法來看，想確定樣本資料的統計分配非常困難，因此 Hyvärinen 於 1999 年根據最大熵原理，發展出下式的近似函數：

$$J(y) \propto \{E[G(y_i)] - E[G(v)]\}^2 \quad (3-14)$$

其中， v 是平均數為 0、變異數為 1，且服從高斯分配的隨機變數，而 G 則為任意非二次方函數(Non-quadratic Function)，因為當 G 為二次方函數時，其 $J(y)$ 值將等於 0。雖然，上式 $G(y_i)$ 可假設為任意非二次方函數，但 Hyvärinen 建議以下兩個非二次方函數 G ：

$$G_1(y_i) = \frac{1}{a_1} \log(\cos(a_1 y_i)) \quad , \text{ 其中 } 1 \leq a_1 \leq 2 \quad (3-15)$$

$$G_2(y_i) = -\exp\left(\frac{-y_i^2}{2}\right) \quad (3-16)$$

理論上，只要 $G(y_i)$ 為非二次方函數時，皆可用以搜尋解混合矩陣 W ；然而，近年的實證顯示 $G_1(y_i)$ 函數可廣泛適用於各領域。由於函數的不同只在於達到最佳化所花費的時間長短，且本研究的實證中亦無法確定樣本獨立成份屬何種分配，故研究中將以 $G_1(y_i)$ 作為輸入 FastICA 的函數。

（二）FastICA 演算流程

在決定好目標函數之後，則利用演算法將目標函數最佳化，進而找到最適當的解混合矩陣 W 。但是，為能簡化 ICA 運算過程並縮短其處理時間，Hyvärinen 與 Oja（2000）建議可以對欲分析的資料先進行前置處理

(Preprocessing)，主要的前置處理步驟包括兩項：資料置中(Centering) 和白化(Whitening)。透過在前置處理的過程中，已將原始混合 X 轉換為無相關的資料；但由於無相關未必代表獨立，因此必須再設法求出獨立轉換矩陣 W ，而將無相關資料轉換成一組近似於獨立的獨立資料 Y 。FastICA 基本演算法的處理步驟如下：

步驟 1：隨機選擇一個權重向量 w ，且令 $w=1$ 。

步驟 2： $w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w$ 。

步驟 3：令 $w^* = w^+ / \|w^+\|$ 。

步驟 4：假如 w 沒有收斂，則回到步驟 2。

以上步驟有幾點需要注意的：第一、函數 g 為式(3-15)或式(3-16)中 G 的導函數(Derivative)。第二、在步驟 4 中所提之收斂，是指新的向量 w^* 與舊的權重向量 w 平行；亦即 w^* 與 w 內積的絕對值為 1。為了避免每次疊代收斂於同一點上，過程中必須於每次疊代之後利用抽離式去相關(Deflation Decorrelation)或是對稱式去相關(Symmetric Decorrelation)等方式，去除獨立成份彼此之間的相關性，以確保每次估計出的獨立成份皆不相同(Hyvärinen, 1999；Hyvärinen & Oja, 2000)。

第四節 單變量 ARIMA 模型

自我迴歸整合移動平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) 是由 Box & Jenkins 在 1970 年所提出的一套時間序列分析方法，並以遞迴的方式對時間序列資料建構模型 (Box & Jenkins, 1976)；其中包含兩部分：若可藉由過去歷史資料序列中推估未來的趨勢與變化，則此序列具有自我迴歸的過程 (Autoregressive Process)；若可藉由過去歷史資料序列中的誤差項推估，則此序列具有移動平均的過程 (Moving Average Process)。而 ARIMA 模型係以的 ARMA 模式為基礎並擴展而成，其中包括「自我迴歸」、「整合」與「移動平均」三個模式，分別以 p 代表自我迴歸

階數；d 代表整合（差異）階數；q 代表移動平均之階數，而以 ARIMA(p,d,q) 表之。ARIMA 模型可應用於時間序列係一組空間或時間間隔相等的有序觀察值，亦即 $(Y_1, Y_2, Y_3, \dots, Y_{t-1}, Y_t, Y_{t+1}, \dots)$ 。現若以變數 Y_t 為非平穩的時間序列而言，則 ARIMA(p,d,q) 模型的定義如下：

$$\Delta^d y_t = a_0 + a_1 \Delta^d y_{t-1} + a_2 \Delta^d y_{t-2} + \dots + a_p \Delta^d y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots + b_q \varepsilon_{t-q} \quad (3-17)$$

式中 $\Delta^d y_t$ ，係將變數差分 d 次； a_p 為自我迴歸第 p 階數之參數估計值； b_q 為移動平均第 q 階數之參數估計值。

根據 Box & Jenkins 的方法論中，建構 ARIMA 模型的程序如圖 3-9 所示，並說明如下：

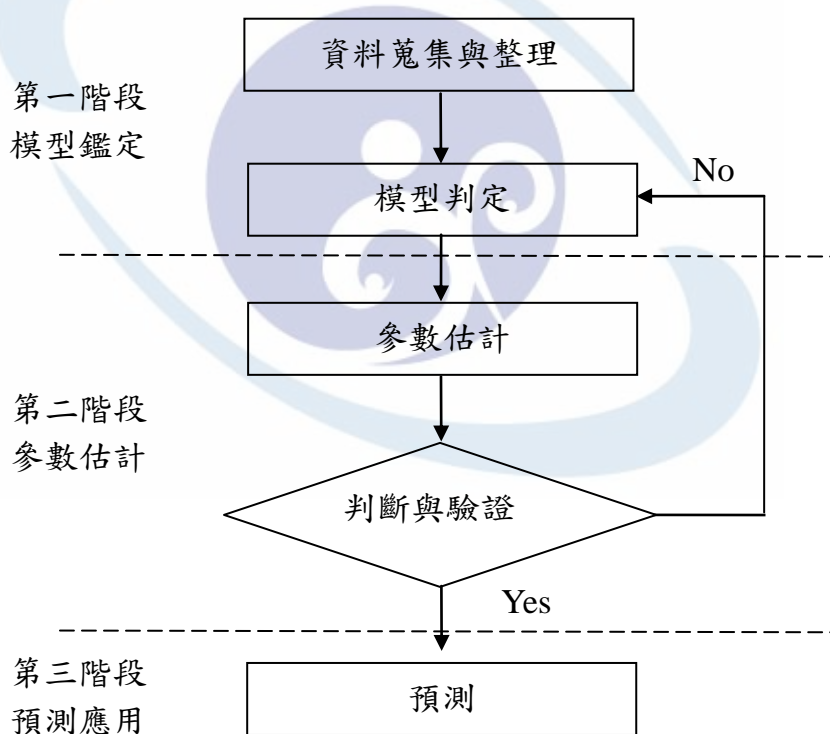


圖 3-9 ARIMA 模型建構流程圖

資料來源：Box & Jenkins，1976；本研究整理。

一、資料蒐集與整理

針對研究議題，蒐集並整理其時間序列資料。

二、模型判定

其可透過圖形、統計量、自相關函數 (Autocorrelation Function, ACF)、偏自相關函數 (Partial Autocorrelation Function, PACF) 與變數轉換、單根檢定等方法，以試驗的方式達成模型和模式成份的辨認，亦即決定 ARIMA 模型的參數值(p,d,q)。模型建構的第一步為決定差分的階次，記為 I(d)；接著，再決定自我迴歸的落後期數，記為 AR(p)；同時亦決定移動平均的落後期數，記為 MA(q)。而研判研究中的時間序列是否具有平穩性 (Stationary)，本研究應用單根檢定進行判斷。若經由單根檢定達到顯著水準，亦即落入拒絕域，則此時間序列是平穩的；反之，則為非平穩。如果檢定原時間序列為非平穩，則可經過差分使之達到平穩，並決定 ARIMA 模型差分的階次。此外，判斷模型中 AR(p)與 MA(q)可能的階次時，可利用 ACF 與 PACF，其判定方法如表 3-1 所示。

表 3-1 ACF 與 PACF 判定方法

模型	ACF	PACF
AR(p)	呈現遞減或震盪式遞減	呈現 p 期後不顯著
MA(q)	呈現 q 期後不顯著	呈現遞減或震盪式遞減
ARMA(p,q)	呈現遞減或震盪式遞減	呈現遞減或震盪式遞減

資料來源：Box & Jenkins，1976；本研究整理。

三、參數估計

模型判定時，將可同時得到數個初步模型；接著，使用最小平方法或最大概似估計法等估計方式，估計與檢定參數是否具顯著的解釋能力，同時，亦刪除不顯著的落後期數變數。

四、判斷與驗證

在模型參數估計完成之後，需診斷其隨機誤差項是否滿足常態分配且彼此互為獨立，其將可應用殘差圖、Q 統計量、或其之自相關係數與偏相關係數，以決定模式適當與否。此時，若模型適當，則可使用此模型進行下階段工作；否則，重複判定模型、估計和驗證步驟。此外，若有多個初步模型適當時，則可從中擇其最適者而為之。

五、預測與應用

完成模型建構之後，則可使用圖形、統計量與信賴區間以決定分析，並進行樣本外的預測，並可追蹤模型表現以偵測是否有失去控制的情況。

第五節 資料特性與顯著性檢定

一、單根檢定

時間序列資料可分為平穩與非平穩兩種型態，一般的平穩資料對於外部衝擊的反應只會造成暫時性的影響，其長期下來會逐漸回歸到平均值；但若時間序列屬於非平穩資料，則外部衝擊將會產生累積效果，進而使其逐漸偏離平均值。若對非平穩資料進行迴歸分析時，將會造成 Granger & Newbold (1974) 所提到的假性迴歸 (Spurious Regression) 的現象，導致估計結果不具意義；因此，在進行時間序列分析時，必須確保資料為平穩。而平穩資料會滿足下列三點特性：

$$\begin{aligned}\mu_y &= E(y_t) \\ E[(y_t - \mu_y)^2] &= E[(y_{t+i} - \mu_y)^2] \\ \text{Cov}(y_t, y_{t+j}) &= \text{Cov}(y_{t+i}, y_{t+i+j})\end{aligned}\tag{3-17}$$

式中 μ_y 為期望值，而 i, j 為任意值。

單根檢定的目的，在於判斷樣本是否為無須經過差分即可符合平穩的資料，由於 Dickey & Fuller (1979) 所提出的 DF 檢定並未考慮到殘差項的自我相關現象。因此，在本研究在建構 ARIMA 模型時，將採行考慮到殘差

項序列相關的 ADF (Augmented Dickey Fuller) 檢定，以決定差分的階次。

二、顯著性檢定

額外平方和 (Extra Sums of Squares) 的概念，為在原有某些預測變數的迴歸模型中，再加入一個或多個特定預測變數時所造成誤差平方和 (SSE) 縮減的份量，其亦可視為迴歸平方和 (SSR) 增加的份量。假設 X_1 與 X_2 為某一迴歸模型的兩個預測變數，而當 X_1 與 X_2 都在迴歸模型中時，其誤差平方和 (SSE) 必定小於模型中只有一個預測變數 (X_1 或 X_2)，而其之間的差異即為額外平方和。

現若考慮 $(k-1)$ 個預測變數的多元迴歸模型，如式(3-18)中所示，則稱其配適為完整模型 (Full Model)；現若欲檢定模式中最後一個預測變數 (X_k) 的係數 β_k 是否為 0 時，則將可得到其之縮減模型 (Reduced Model)，如式(3-19)中所示。則用以檢定係數 β_k 是否為 0 的統計量，即稱為服從於自由度為 $(df_R - df_F, df_F)$ 的偏 F 統計量 (Partial F Statistic, F')，其檢定統計量如式(3-20)中所示。

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (3-18)$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_{k-1} X_{(k-1)i} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (3-19)$$

$$F' = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}} = \frac{[SSE(R) - SSE(F)] / 1}{MSE(F)} \quad (3-20)$$

式中的 df_F 與 df_R 分別為完整模型與縮減模型的自由度， $MSE(F)$ 為完整模型 (full model) 的均方誤差 (Mean Square Error)。

據此，由於自由度的關係，研究中即應用其概念，提出如下服從於自由度為 $(1, df_F)$ 的 F 分配的「擬偏 F 統計量 (Pseudo Partial F Statistic, F^*)」，而將各個獨立成份還原後的序列與原始序列間進行差異性分析工作，以期藉由顯著性檢定方式，選定最適還原的獨立成份個數。而在將與原始數列

較不具貢獻的獨立成份剔除後（亦即保留顯著重要的獨立成份）並還原資料，並視該還原後的新序列加上趨勢分量為已去除雜訊過後的時間序列資料，而進行後續混合預測模型建構的工作。

$$F^* = \frac{\text{MSE(R)} - \text{MSE(F)}}{\text{MSE(F)}} \quad (3-21)$$

式中的 MSE(F)與 MSE(R)分別為完整模型(Full model)與縮減模型(Reduced model) 的均方誤差。

第六節 預測績效衡量

在預測模型領域中，常見用來比較不同模型預測能力的指標有：誤差均方根(Root Mean Square Error, RMSE)、平均誤差絕對值(Mean Absolute Error, MAE)、平均誤差百分比值(Mean Absolute Percentage Error, MAPE)等。其中 A_t 表示測試樣本的實際觀察值，而 P_t 表示測試樣本的模型預測值，其詳細計算公式如表 3-2 中所示；三項指標皆為當指標值愈小時，代表模型越貼近真實的時間序列資料。其中，學者 Lewis (1982) 更針對 MAPE 指標提出其判定準則：當指標值低於 10% 時，表示模型預測能力極佳；指標值介於 10% 至 20% 時，表示預測能力良好；指標值介於 20% 至 50% 時，表示預測能力合理；而當指標值高於 50% 時，則表示預測將不正確。

而本研究在台灣工業生產指數的實證個案中，將應用此三項指標以衡量研究中所提之經由結合 EMD 與 ICA 的混合預測模型於預測時的實際表現與比較情況。

表 3-2 預測能力的衡量指標

指標類別	衡量公式
MAPE	$\frac{1}{n} \times \sum_{t=1}^n \left \frac{A_t - P_t}{A_t} \right \times 100\%$
RMSE	$\sqrt{\frac{1}{n} \times \sum_{t=1}^n (A_t - P_t)^2}$
MAE	$\frac{1}{n} \times \sum_{t=1}^n \left \frac{A_t - P_t}{A_t} \right $

註：其中 n 為測試樣本資料樣本數。

資料來源：本研究整理。



第肆章 實證分析

第一節 時間序列資料說明

本研究主要以近年來台灣工業生產指數為實證個案研究主題，資料來源取自行政院經濟建設委員會每月所公佈的景氣對策指標，蒐集的資料期間為2000年1月至2012年5月止，合計149筆月研究資料（如圖4-1所示）。

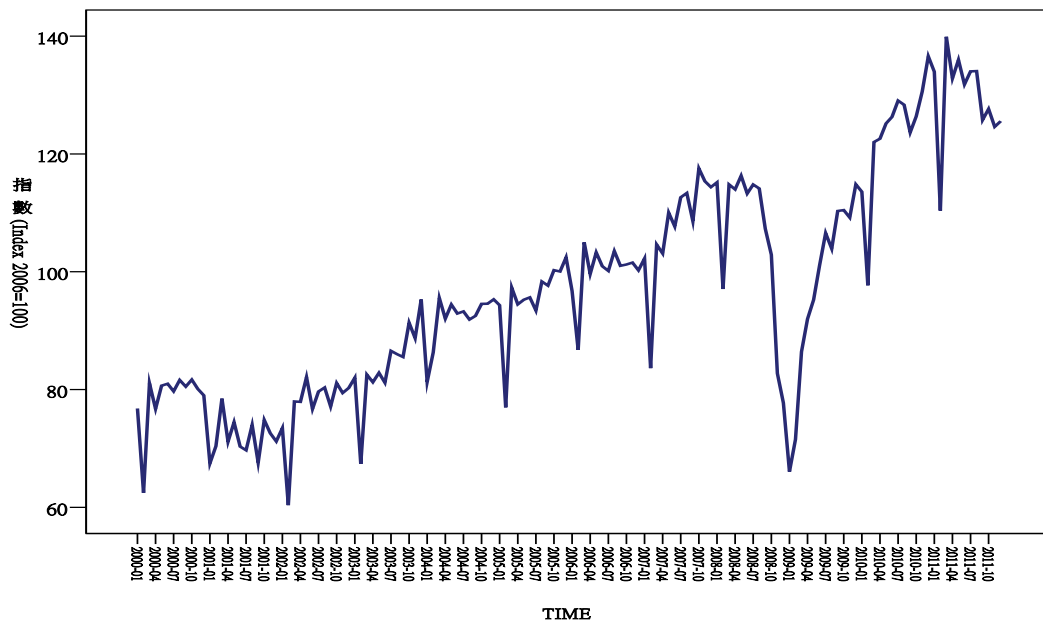


圖 4-1 台灣工業生產指數歷史走勢圖(2000/01~2012/05)

資料來源：本研究整理。

研究中係以2000年1月至2010年12月，合計132筆的資料序列為訓練樣本，並據以建立預測模型；而以2011年1月至2012年5月的資料序列為測試樣本，藉以進行模型預測與比較的工作（如表4-1所示）。

表 4-1 研究資料設置

原始資料	訓練樣本	測試樣本
2000/01~2012/05	2000/01~2010/12	2011/01~2012/05
(合計 149 筆)	(合計 132 筆)	(合計 17 筆)

資料來源：本研究整理。

第二節 經驗模組分解法

首先，將蒐集到的 2000 年 1 月至 2010 年 12 月的台灣工業生產指數，應用 EMD 分解成 6 個本質模組函數 (IMF) 分量以及最後一個趨勢分量 r ，如圖 4-2 中所示。由表 4-2、表 4-3 與圖 4-2 中顯示，各分量代表著不同尺度的變化量，且具有相關性。 IMF_1 為平均約為 2.75 月的週期震盪，標準差為 6.8983； IMF_2 為平均約為 5.87 月的週期震盪，標準差為 6.1340； IMF_3 為平均約為 10.56 個月的週期震盪，標準差為 9.1400； IMF_4 為平均約為 26.4 個月的週期震盪，標準差為 6.7255； IMF_5 為平均約為 44 月的週期震盪，標準差為 3.3651； IMF_6 則被分解成平均週期為 132 月的資料，其標準差為 2.9973；而由趨勢分量 r 可知，台灣工業生產指數研究期間係屬上升趨勢。

表 4-2 研究樣本之敘述統計

資料	樣本數	平均數	中位數	標準差	最大值	最小值	平均週期
Index	132	93.9707	94.4500	16.8534	136.5800	60.37	-
IMF1	132	-1.0806	-0.4026	6.8983	11.9179	-35.7664	2.75
IMF2	132	0.4884	0.1596	6.1340	22.2980	-19.4269	5.87
IMF3	132	2.0981	0.4667	9.1400	33.1363	-16.3199	10.56
IMF4	132	-1.0973	-0.2148	6.7255	24.7097	-17.8681	26.4
IMF5	132	-0.0439	-0.0740	1.4636	3.3651	-2.5995	44
IMF6	132	1.6471	1.8362	2.9973	5.7065	-2.9763	132
r	132	91.9588	91.5826	12.0079	113.2492	72.1393	-

資料來源：本研究整理。

表 4-3 各 IMF 之 Pearson 相關

	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆
IMF ₁	1	-	-	-	-	-
IMF ₂	-0.534*	1	-	-	-	-
IMF ₃	-0.212*	-0.216*	1	-	-	-
IMF ₄	-0.322*	0.267*	-0.228*	1	-	-
IMF ₅	-0.200*	0.160	-0.113	0.653*	1	-
IMF ₆	0.082	-0.039	-0.081	-0.037	-0.083	1

註：*表示p值小於顯著水準0.05

資料來源：本研究整理。



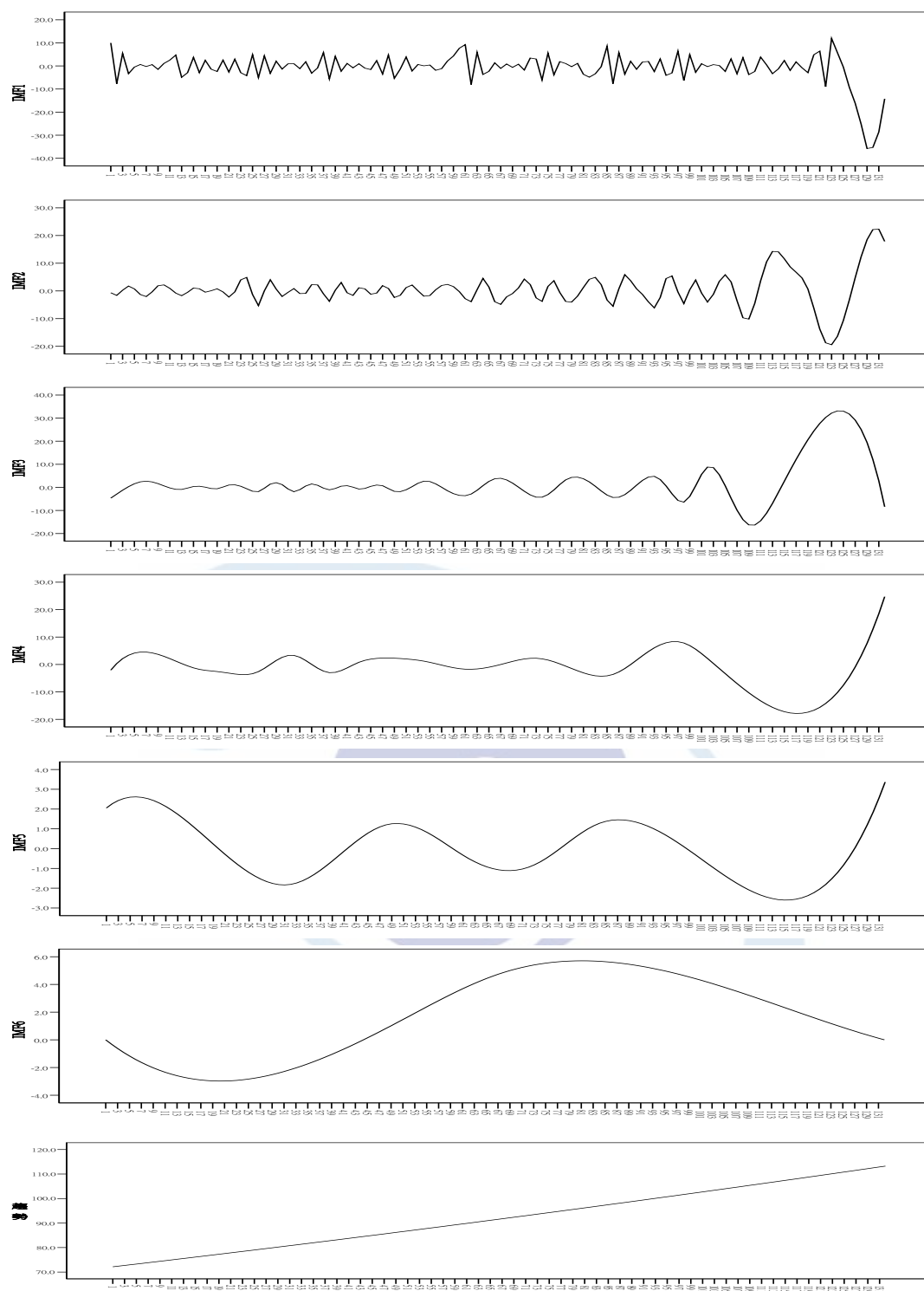


圖 4-2 台灣工業生產指數之本質模組函數 IMFs 與趨勢分量 r
(2000/01~2011/12)
資料來源：本研究整理。

其次，若以式(4-1)說明透過 EMD 篩選 IMF 分量的過程中，將原始訊號減去第一個 IMF 分量，則可得到第一個殘餘量（記為 $R_1(t)$ ），再將 $R_1(t)$ 當作新的資料，則將可得到第二個殘餘量 $R_2(t)$ 。以此類推，找出各殘餘量。各 IMF 分量、殘餘量及原始資料的關係如圖 4-3 所示。由實證結果發現，各個 IMF 分量是依據訊號時間與頻率透過篩選的迭代程序取得；因此，若樣本數的不同，則 IMF 分量的性質亦有所變動，且透過各個殘餘量則可以發現數列依不同週期的平穩化過程，而將最後一筆趨勢分量 r 與各個 IMF 分量加總後，即可得到原訊號（式(4-2)）。

$$\begin{aligned}
 R_1(t) &= Y(t) - C_1(t) \\
 R_2(t) &= R_1(t) - C_2(t) \\
 R_3(t) &= R_2(t) - C_3(t) \\
 R_4(t) &= R_3(t) - C_4(t) \\
 R_5(t) &= R_4(t) - C_5(t) \\
 R_6(t) &= R_5(t) - C_6(t)
 \end{aligned}
 \tag{4-1}$$

$$Y = C_1 + C_2 + \cdots + C_6 + r \tag{4-2}$$

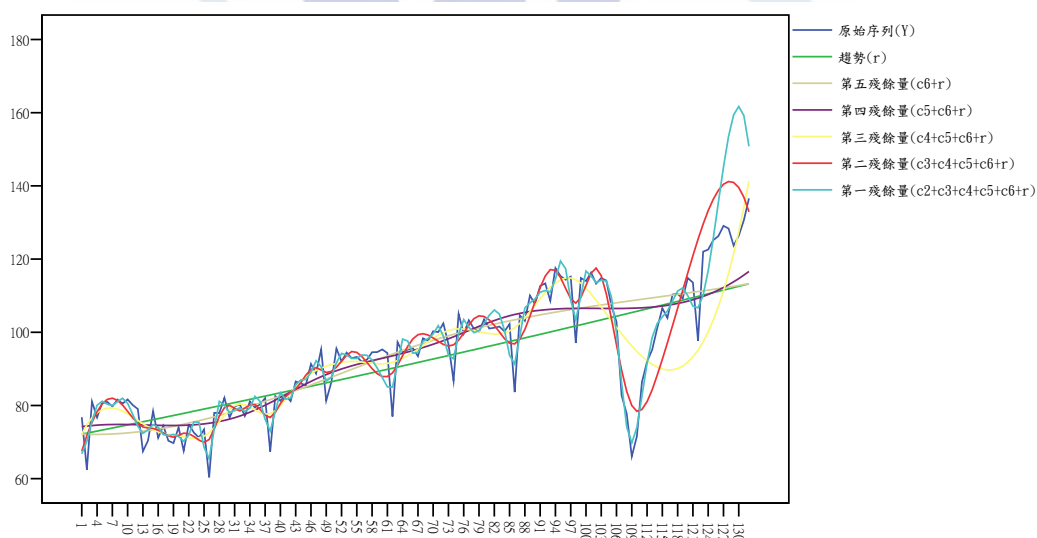


圖 4-3 原始序列、趨勢分量 r 與各殘餘量

資料來源：本研究整理。

第三節 獨立成份分析法

一、工業生產指數之獨立成份

透過前一節，說明了對原始訊號應用 EMD，將工業生產指數分解成 6 個本質模組函數（IMF）分量以及最後一個趨勢分量 r 。而本研究假設各個 IMF 分量對於工業生產指數均有隱含著解釋能力，若直接去除一個或數個 IMF 後建立預測模型，將有可能會喪失許多重要的資訊，故而研究中將 6 個 IMFs 執行 ICA 步驟。

現假設 $IMF_1(t)$, $IMF_2(t)$, $IMF_3(t)$, $IMF_4(t)$, $IMF_5(t)$, $IMF_6(t)$ 分別為來源訊號 $s_1(t)$, $s_2(t)$, $s_3(t)$, $s_4(t)$, $s_5(t)$, $s_6(t)$ 經由線性組合而成，其中 t 為 1~132。若以數學式表示（式(4-3)）：

$$\begin{aligned} imf_1(t) &= a_{11}s_{1t} + a_{12}s_{2t} + a_{13}s_{3t} + a_{14}s_{4t} + a_{15}s_{5t} + a_{16}s_{6t} \\ imf_2(t) &= a_{21}s_{1t} + a_{22}s_{2t} + a_{23}s_{3t} + a_{24}s_{4t} + a_{25}s_{5t} + a_{26}s_{6t} \\ imf_3(t) &= a_{31}s_{1t} + a_{32}s_{2t} + a_{33}s_{3t} + a_{34}s_{4t} + a_{35}s_{5t} + a_{36}s_{6t} \\ imf_4(t) &= a_{41}s_{1t} + a_{42}s_{2t} + a_{43}s_{3t} + a_{44}s_{4t} + a_{45}s_{5t} + a_{46}s_{6t} \\ imf_5(t) &= a_{51}s_{1t} + a_{52}s_{2t} + a_{53}s_{3t} + a_{54}s_{4t} + a_{55}s_{5t} + a_{56}s_{6t} \\ imf_6(t) &= a_{61}s_{1t} + a_{62}s_{2t} + a_{63}s_{3t} + a_{64}s_{4t} + a_{65}s_{5t} + a_{66}s_{6t} \end{aligned} \quad (4-3)$$

若以矩陣表示，則可假設觀測到的混合訊號矩陣為 IMF，各組來源矩陣為 S，與距離有關之混合矩陣則為 A，此三者的關係如下：

$$IMF = \begin{bmatrix} IMF_1 \\ IMF_2 \\ \vdots \\ IMF_6 \end{bmatrix}_{6 \times 132} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{16} \\ a_{21} & a_{22} & \cdots & a_{26} \\ \vdots & \vdots & \cdots & \vdots \\ a_{61} & a_{62} & \cdots & a_{66} \end{bmatrix}_{6 \times 6} \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1132} \\ s_{21} & s_{22} & \cdots & s_{2132} \\ \vdots & \vdots & \cdots & \vdots \\ s_{61} & s_{62} & \cdots & s_{6132} \end{bmatrix}_{6 \times 132} = AS$$

過程中提及到在來源訊號 S 未知的情況下，如果混合矩陣 A 已知，則透過線性方程組的方式求解，將可順利求出原先工業生產指數之訊號 S，即 $S=A^{-1} \cdot IMF$ ；但是本研究所探討的問題在於混合矩陣 A 與來源矩陣 S 均是未知的，因此面對這樣的問題，研究中將應用 FastICA 所運用統計分析方法

估計出矩陣 A' 用以近似未知的混合矩陣 A ，再利用 A' 之反矩陣求出 W 矩陣， W 矩陣是用來將混合數列還原為近似於最初狀態的獨立成份矩陣： $W \cdot IMF$ ，進而將獨立成份用以估計未知潛在來源訊號 S ，以本研究工業生產指數所計算之 W 解混合矩陣表示如下：

$$W = \begin{bmatrix} 0.0103 & -0.0122 & -2.4302 & -1.1841 & -6.3087 & -0.3435 \\ 0.0019 & 0.0831 & -2.5070 & 2.8932 & 4.1066 & -2.4114 \\ 0.0088 & 0.1140 & -1.0459 & -0.9152 & 2.2010 & 8.7245 \\ 0.1620 & 0.0882 & -0.7148 & -4.6238 & 3.6634 & -3.0895 \\ -1.0744 & -0.0802 & -0.1478 & -0.7128 & 0.5278 & -0.3959 \\ -0.0343 & 2.9593 & 0.1180 & 0.0678 & -0.3210 & -0.1884 \end{bmatrix}$$

將混合數列還原為近似於最初狀態的獨立成份矩陣為 IMF 矩陣乘上解混合矩陣 W 矩陣，並且可得到估計原始來源訊號的 IC 矩陣，即 $IC=W \cdot IMF$ 。其中，6 個獨立的 ICs ，如圖 4-4 中所示。

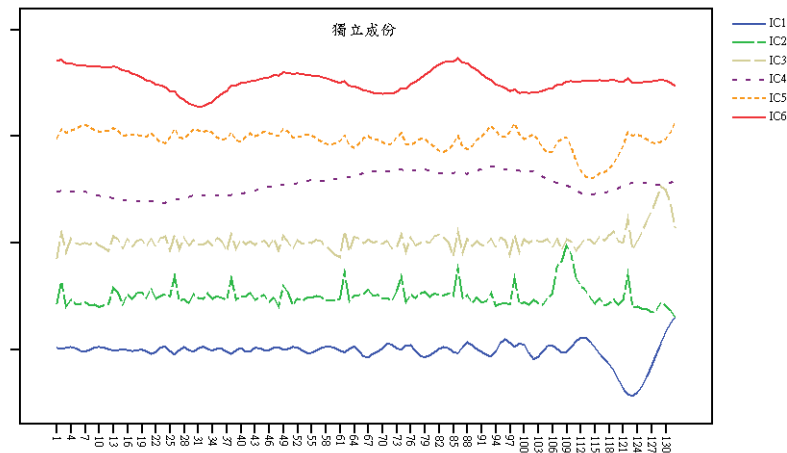


圖 4-4 台灣工業生產指數的 6 個獨立成份

資料來源：本研究整理。

經由獨立成份分析這種類似成份萃取及線性組合的概念，不僅能保有原來 IMF 分量所有的資訊，而且使得所有分離的數列 (IC_s) 間彼此之間擁有最大差異，這樣一來每個 IC 各自無相關且獨立，提供接下來更進一步研究及分析，以求得最能夠解釋原始資料的成份。

二、選取重要獨立成份

接下來分析的目的在于尋找一組變數組合，一方面要保有原始序列之結構與趨勢，另一方面則期望透過少數且重要的獨立成份即能解釋之。假設工業生產指數通常是可由少數重要的獨立成份所組成，但我們無法得知到底哪個獨立成份對於原始時間序列最具有解釋能力，故而研究中獨立成份選取的第一步驟是先將每個獨立成份各自還原之後（圖 4-5），再求得與原始資料間的離差平方和，其中擁有最小離差平方和的 IC 者將被視為建構工業生產指數最重要的獨立成份。

同理，再將確定排序第一的獨立成份，分別與其餘獨立成份兩兩相互配對後並還原之，其中擁有最小離差平方和的成份組合視為最佳組合，而其獨立成份將被視為次重要的獨立成份；以此類推，直到將所有的獨立成份依重要程度排序完成為止。而在本研究中的重要獨立成份，經前述判定過程後，依序為 IC_2 、 IC_4 、 IC_5 、 IC_3 、 IC_6 、 IC_1 ，如表 4-4 中所示。

由結果可看出，離差平方和經由每一步驟的獨立成份（IC）還原後逐漸減少，直到最後步驟的離差平方和為零。雖說，從中確實可看出各個獨立成份的重要排序，但重點是到底要如何選擇最佳的獨立成份個數，並將與原始序列最不具相關的獨立成份予以剔除，而達到去除雜訊之目的？選擇過多的獨立成份，將可能會導致雜訊無法去除乾淨；選擇過少的獨立成份，則無法表現原數列的結構與趨勢。

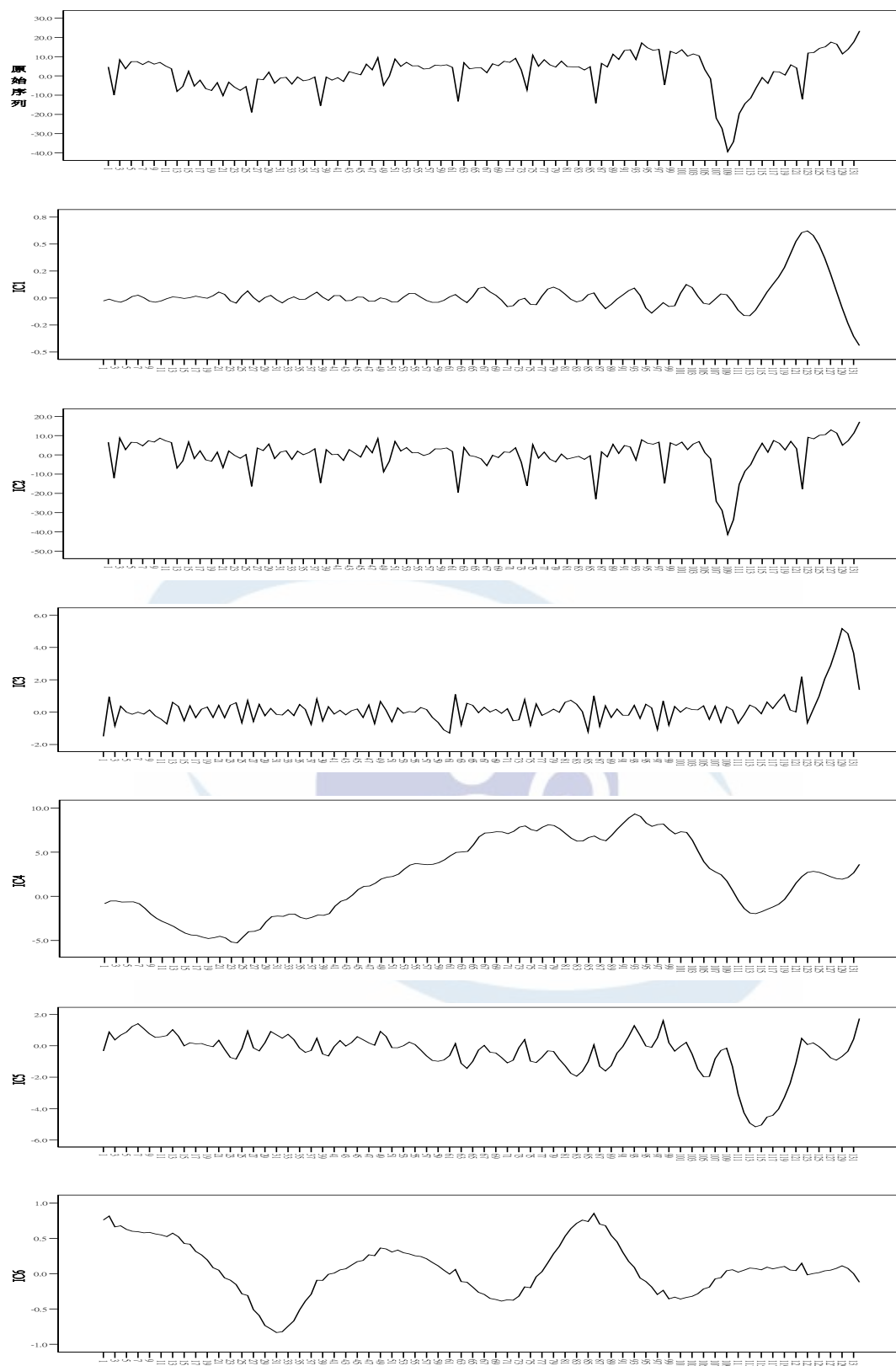


圖 4-5 原始訊號與第一步驟所還原的序列資料

資料來源：本研究整理。

表 4-4 個別獨立成份還原後的離差平方和比較表

比較次數	還原之獨立成份	離差平方和	比較次數	還原之獨立成份	離差平方和
第一步驟	IC ₁	13235.50	第三步驟	IC ₂ 、IC ₄ 、IC ₁	373.34
	IC₂	3254.55		IC ₂ 、IC ₄ 、IC ₃	264.42
	IC ₃	13016.75		IC₂、IC₄、IC₅	158.26
	IC ₄	10315.50		IC ₂ 、IC ₄ 、IC ₆	360.16
	IC ₅	13292.18	第四步驟	IC ₂ 、IC ₄ 、IC ₅ 、IC ₁	154.02
	IC ₆	13195.84		IC₂、IC₄、IC₅、IC₁、IC₃	22.08
第二步驟	IC ₂ 、IC ₁	3241.88	第五步驟	IC ₂ 、IC ₄ 、IC ₅ 、IC ₆	135.29
	IC ₂ 、IC ₃	3028.12		IC ₂ 、IC ₄ 、IC ₅ 、IC ₃ 、IC ₁	18.94
	IC₂、IC₄	375.21		IC₂、IC₄、IC₅、IC₃、IC₆	2.79
	IC ₂ 、IC ₅	3286.23	第六步驟	IC ₂ 、IC ₄ 、IC ₅ 、IC ₃ 、IC ₆ 、IC ₁	0
	IC ₂ 、IC ₆	3203.43			

資料來源：本研究整理。

額外平方和 (Extra Sums of Squares) 的概念，係在原有某些預測變數的迴歸模型中，再加入一個或多個特定預測變數時所造成誤差平方和 (SSE) 縮減的份量，其亦可視為迴歸平方和 (SSR) 增加的份量。據此，本研究即應用此概念而提出「擬偏 F 統計量 (Pseudo Partial F Statistic, F^*)」，將各個獨立成份還原後的序列與原始序列間進行差異性分析工作，以期藉由顯著性檢定方式，選定最適還原的獨立成份個數。而在將與原始數列較不具貢獻的獨立成份剔除後（亦即保留顯著重要的獨立成份）並還原資料，則視該還原後之新序列為已去除雜訊過後的時間序列資料。

研究結果如表 4-5 中所示，其顯示出當模型中已存在 IC₂ 時，再增加 IC₄ 所還原後的序列將與原始序列具顯著性貢獻（式(4-4)）；但一旦再增加 IC₅ 時，則模型的解釋能力並沒有明顯地增加（式(4-5)）。因此，研究中的工業

生產指數個案將依此顯著性檢定之結果，選取 IC_2 、 IC_4 兩獨立成份所還原後的序列並加上趨勢分量 r ，做為下一階段預測模型配適之已去除雜訊的時間序列資料（不含趨勢分量 r ）。

$$F_1^* = \frac{MSE(IC_2) - MSE(IC_2, IC_4)}{MSE(IC_2, IC_4)} = 7.606 \quad (4-4)$$

$$F_2^* = \frac{MSE(IC_2, IC_4) - MSE(IC_2, IC_4, IC_5)}{MSE(IC_2, IC_4, IC_5)} = 1.354 \quad (4-5)$$

表 4-5 各階段還原之擬偏 F 檢定結果

階段	獨立成份	模型均方誤差	F^*
1	IC_2	25.035	—
2	IC_2 、 IC_4	2.909	7.606*
3	IC_2 、 IC_4 、 IC_5	1.236	1.354

註：* 表示在顯著水準 0.05 下具顯著性

資料來源：本研究整理。

第四節 模型建立與預測

首先，本研究將 2000 年至 2010 年台灣工業生產指數的月資料序列 (Y_t)，經由 EMD 分解出的 IMF 分量，且應用獨立成份分析法將其轉換成獨立成份 (ICs)，再經由研究中所提之顯著性檢定方式，選取具顯著貢獻的重要獨立成份以還原序列。因此，將此還原後的序列加上趨勢分量 r ，視為已去除雜訊的時間序列資料 (Y_t^c)。而原始的時間序列與經由本研究透過 EMD 與 ICA 經由顯著性檢定篩選機制之已去除雜訊的時間序列資料如圖 4-6 中所示。

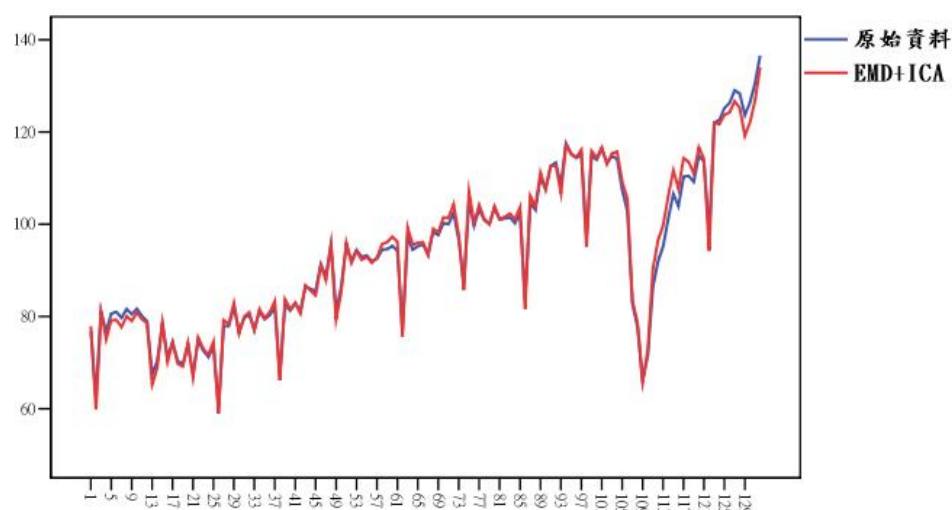


圖 4-6 原始序列與去除雜訊序列比較圖

資料來源：本研究整理。

一、序列的整合階次比較

本研究欲瞭解透過研究中所提之整合方法以去除雜訊的時間序列資料 (Y_t^c) 與原始時間序列 (Y_t) 兩變數數列是否具平穩性，將藉由 ADF 單根檢定的方式判定序列是否具有單根，以決定序列資料是否具有不平穩性；若檢定結果序列具有單根，則需將序列進行差分後再進行單根檢定，直到序列不具單根為止，亦藉此判斷建構 ARIMA 模型時差分階次(d)。

首先，判定 Y_t^c 及 Y_t 兩變數及其差分項的單根檢定結果明細，則列示於表4-6。依據表4-4結果顯示，去除雜訊的時間序列資料 Y_t^c 經檢定已不具單根，因此在建構時間序列模型時屬於I(0)序列。而原始序列 Y_t 經檢定具單根，其一階差分後之序列經檢定則有不具單根之結果，故判斷其經差分一次過後之序列將不具單根，因此在建構原始時間序列模型時屬於I(1)序列。

表 4-6 單根檢定之結果

	未經差分	一階差分
Y_t^c	-3.6746*	-18.2817*
Y_t	-2.9149	-17.3844*

註：*表示p值小於顯著水準0.05

因此，經由結合 EMD 與 ICA 所建立之已去除雜訊的時間序列資料，其在建構 ARIMA 模型過程中，本研究確認其已屬平穩的時間序列，此說明了透過 EMD 與 ICA 將能有效使得時間序列資料呈現平穩結果。

二、混合ARIMA模型之建立

2000 年至 2010 年之已去除雜訊時間序列資料 Y_t^c 經判定為 $I(0)$ ，顯示其無需再經由差分即已達平穩性。因此，研究中直接檢視該序列未經差分過後之 ACF 及 PACF（如圖 4-7），藉以判斷 $AR(p)$ 與 $MA(q)$ 的落後期數。

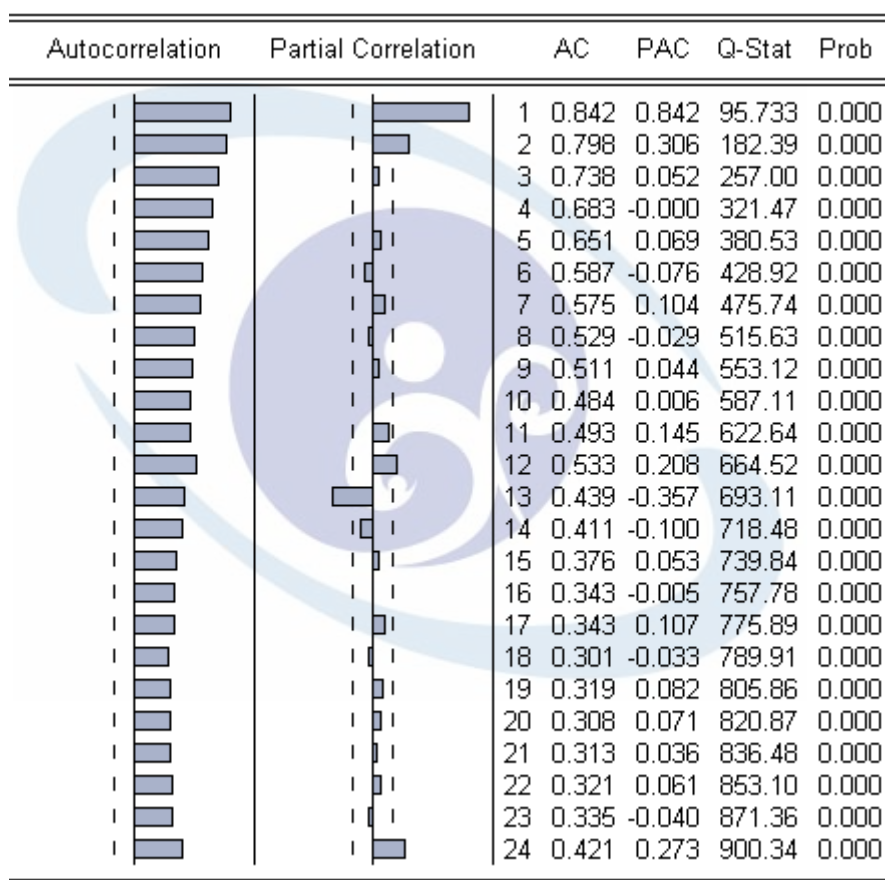


圖 4-7 Y_t^c 的 ACF 及 PACF 圖

資料來源：本研究整理。

由 ACF 可以看出，於落後期數呈現遞減的現象，而 PACF 於落後期數 1、2、12、13、24 期有較凸出，依此判斷序列 ARIMA 可能的模型，在建構可能之模型參數均達顯著水準 0.05，其模型的 \bar{R}^2 、 R^2 、Akaike Information Criterion（以下簡稱 AIC）及 Schwarz Bayesian Criterion（以下簡稱 SBC）

列示於表 4-7，其中 \bar{R}^2 及 R^2 越大顯示模型配適能力越佳；AIC 與 SBC 越小則模型配適能力越佳。由表中訊息可知，模型三之 \bar{R}^2 、 R^2 、AIC 及 SBC 皆優於其他兩模型，若以 \bar{R}^2 、 R^2 判斷，則模型一優於模型二；若以 AIC、SBC 判斷，則又以模型二較佳，故而較不易明顯判斷兩模型何者較優。

表 4-7 Y_t^c 可能的 ARIMA 模型估計檢驗

	可能模型	\bar{R}^2	R^2	AIC	SBC
模型一	ARMA[(1,2),0]	0.7857	0.7890	6.9592	7.0254
模型二	ARMA(1,1)	0.7832	0.7865	6.9954	7.0612
模型三	ARMA[(1,2,12,13),0]	0.8321	0.8363	6.7110	6.8044

資料來源：本研究整理。

其次，本研究同時以三個模型進行 Q 統計量以進行模型的殘差項檢定，以判斷隨機誤差項是否滿足常態分配，將通過 Q 統計量檢定結果的模型，節錄其落後 6、12、18、24 期的 Q 統計量數值整理於表 4-8 中。由表中訊息可知，模型一及模型二在落後期數 12 以後其 P-Value 皆小於 0.05，代表在給定顯著水準 0.05 下，拒絕兩模型之隨機誤差項滿足常態分配，故而最後本研究選擇模型三為最適模型做為混合 ARIMA 模型之建立。

表 4-8 Y_t^c 可能的 ARIMA 模型殘差檢定 Q 統計量

	可能模型	Q6	Q12	Q18	Q24
模型一	ARMA[(1,2),0]	3.343	40.032*	46.490*	77.343*
模型二	ARMA(1,1)	6.868	37.296*	46.086*	77.331*
模型三	ARMA[(1,2,12,13),0]	1.697	6.818	11.976	15.073

註：*表示p值小於顯著水準0.05

資料來源：本研究整理。

藉由配適 ARIMA 模型，將可建構一結合 EMD、ICA 與 ARIMA 的混合模型。在經由上述 ARIMA 模式建立的模式暫定、參數之估計、檢定與修正三個步驟（Box & Jenkins, 1976）後，所建構之台灣工業生產指數最適混合模型的參數估計值與相關統計量如表 4-9 所示：

表 4-9 混合模型參數估計及檢定統計量

變數	係數	T統計量	P值
AR(1)	0.6117	7.6466	< 0.001*
AR(2)	0.2737	3.6243	< 0.001*
AR(12)	0.5161	6.6871	< 0.001*
AR(13)	-0.3911	-4.7593	< 0.001*
R^2	0.8363	AIC	6.7110
\bar{R}^2	0.8321	SBC	6.8044
D.W.	2.0549		

註：*表示p值小於顯著水準0.05

資料來源：本研究整理。

而混合模型方程式表達如式(4-6)：

$$\hat{Y}_t^c = 0.6117Y_{t-1}^c + 0.2737Y_{t-2}^c + 0.5161Y_{t-12}^c - 0.3911Y_{t-13}^c \quad (4-6)$$

三、傳統ARIMA模型之建立

基於客觀的比較準則，研究中亦應用工業生產指數的原始月資料序列（ Y_t ）依據步驟以建構出最適的傳統 ARIMA 模型。首先，2000 年至 2010 年之原始時間序列資料（ Y_t ）經判定為 I(1)；接下來，檢視該序列經由一階差分過後之 ACF 及 PACF（如圖 4-8 所示），藉以判斷 AR(p)與 MA(q)之落後期數。

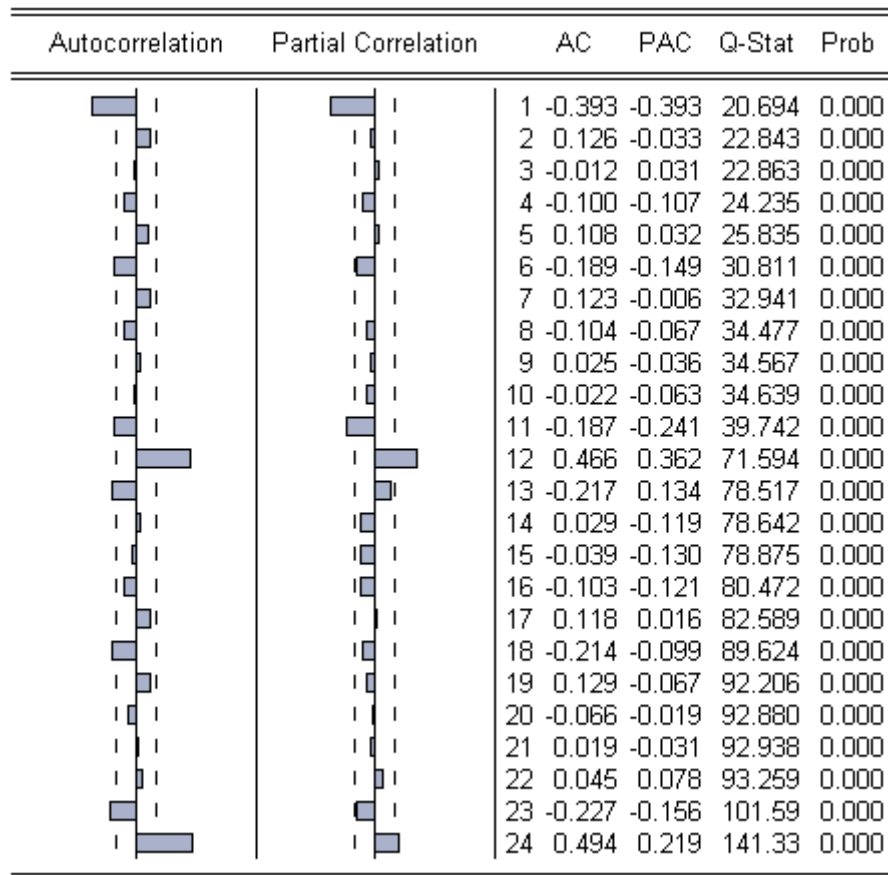


圖 4-8 Y_t 之 ACF 及 PACF

資料來源：本研究整理。

由 ACF 可以看出，於落後期數呈現震盪式遞減的現象，且於落後期數 1、6、11、12、24 期有較凸出；而 PACF 於落後期數 1、6、11、12、24 期有較凸出；依此判斷序列 ARIMA 可能的模型，在建構可能之模型參數均達顯著水準 0.05，其模型之 \bar{R}^2 、 R^2 、AIC 及 SBC 列示於表 4-10 中。由表中訊息可知，模型二之 \bar{R}^2 、 R^2 、AIC 及 SBC 皆優於模型一，若以 \bar{R}^2 、 R^2 判斷，則模型一優於模型二。

表 4-10 Y_t 可能的 ARIMA 模型估計檢驗

	可能模型	\bar{R}^2	R^2	AIC	SBC
模型一	ARIMA[0,1,(1,11,12)]	0.2535	0.2668	6.6083	6.6741
模型二	ARIMA[(12),1,(1,11,12)]	0.5867	0.5972	6.0211	6.1145

資料來源：本研究整理。

其次，本研究亦同時將兩個模型進行 Q 統計量以進行模型的殘差項檢定，以判斷隨機誤差項是否滿足常態分配，將通過 Q 統計量檢定結果的模型，節錄其落後 6、12、18、24 期的 Q 統計量數值整理於表 4-11 中。由表中訊息可知，模型一在落後期數 24 以後其 P-Value 小於 0.05，代表在給定顯著水準 0.05 下，拒絕該模型之隨機誤差項滿足常態分配，故而最後本研究選擇模型二為最適模型做為本研究傳統 ARIMA 模型之建立。

表 4-11 Y_t 可能的 ARIMA 模型殘差檢定 Q 統計量

可能模型		Q6	Q12	Q18	Q24
模型一	ARIMA[0,1,(1,11,12)]	6.8090	10.558	19.238	37.516*
模型二	ARIMA[(12),1,(1,11,12)]	5.9086	12.938	14.743	20.585

註：*表示p值小於顯著水準0.05

資料來源：本研究整理。

藉由配適 ARIMA 模型，將可建構一傳統 ARIMA 模型。在經由上述 ARIMA 模式建立的模型暫定、參數之估計、檢定與修正三個步驟後，所建構之台灣工業生產指數最適傳統模型的參數估計值（T 檢定量）與相關統計量如表 4-12 所示：

表 4-12 傳統 ARIMA 模型參數估計及檢定統計量

變數	係數	T統計量	P值
AR(12)	0.9968	36.2267	< 0.001*
MA(1)	-0.1403	-3.5958	< 0.001*
MA(11)	0.1293	3.2539	0.002*
MA(12)	-0.9082	-30.8586	< 0.001*
R^2	0.5972	AIC	6.0211
\bar{R}^2	0.5867	SBC	6.1145
D.W.	2.1416		

註：*表示p值小於顯著水準0.05

資料來源：本研究整理。

而模型方程式表達如式(4-7)：

$$\Delta^1 \hat{Y}_t = 0.9968 \Delta^1 Y_{t-12} - 0.1403 \epsilon_{t-1} + 0.1293 \epsilon_{t-11} - 0.9082 \epsilon_{t-12} \quad (4-7)$$

式中的 $\Delta^1 Y_t$ 係將變數 Y_t 變數差分一次，即 $\Delta^1 Y_t = Y_t - Y_{t-1}$ 。

第五節 模型預測能力比較

由本研究之台灣工業生產指數訓練樣本內的兩個模型配適度可知，透過研究中所提之結合 EMD 與 ICA 以去除序列雜訊的混合模式，在資料的解釋能力上 (R^2) 確實要較直接應用原始序列建構的 ARIMA 模型改善許多。而在預測能力指標的表現上，經由測試樣本的預測結果如表 4-13 與圖 4-9 中所示；無論是 RMSE、MAE 或 MAPE，混合模型的預測值皆具相對最小誤差，亦即混合 ARIMA 模型的預測效能皆要較原始序列的傳統 ARIMA 模型為佳；同時，其混合模型的 MAPE=5.5859%，其亦符合 Lewis (1982) 所提出之 MAPE 值低於 10% 的預測能力極佳層次。

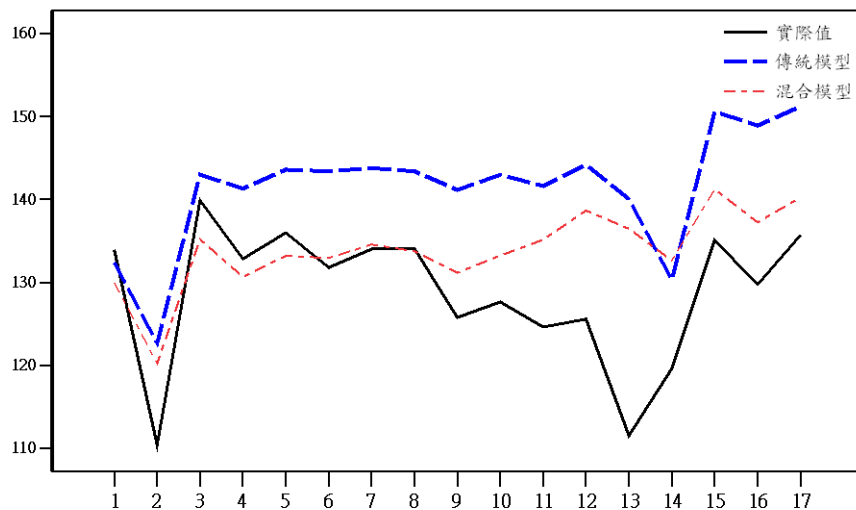


圖 4-9 混合與傳統模型的預測序列圖

資料來源：本研究整理。

表 4-13 混合 ARIMA 模型與傳統 ARIMA 模型之預測結果比較

模型	RMSE	MAE	MAPE
混合模型	9.0730	6.8410	5.5859%
傳統模型	14.3375	12.9110	10.2627%

資料來源：本研究整理。



第伍章 結論與建議

第一節 結論

時間序列的分析與預測，在發展過程中，傳統的相關理論已有相當程度的發展；自 1980 年代初期開始，由於電腦變得容易取得，加以其運算處理能力的快速發展，使得相關的統計理論與概念可透過迭代運算方式付諸實現；其中，人工智慧（AI）的電腦學習相關理論與方法開始進入時間序列分析的領域內，也促使許多的研究者致力於其之預測過程，期望能夠降低預測誤差，並提高預測的準確度。

Haung et al. (1998) 所提出之解決非平穩且非線性時間序列的希爾伯特-黃轉換（HHT），其經驗模組分解（EMD）近年來已開始被應用於時間序列的研究。同時，獨立成份分析法原是一種未知來源訊號分離的技術，近年來愈來愈多的研究者亦將其廣泛地應用於處理時間序列分析領域中。

基於兩種方法的特性，本研究提出結合經驗模組分解與獨立成份分析法的分析架構，將原始時間序列資料透過經驗模組分解後產生的數個本質協振函數，再透過獨立成份分析法，並加入顯著性檢定的方式，找出構成時間序列資料結構與趨勢的重要成份，進而將這些重要成份透過適當的預測模型，建立一套標準的混合模型之建構流程，以降低預測誤差，提高準確度，以期能對時間序列的分析與預測有更進一步的突破與貢獻。研究中並以台灣工業生產指數為實案進行分析。以下簡述研究過程與實證結論：

不同於過去研究所選擇其 IMF 分量的方式，本研究假設各個 IMF 分量對於工業生產指數均隱含著解釋能力，若直接剔除一個或數個 IMF 後所建立的預測模型，有可能會喪失許多重要的資訊；因此結合 ICA 而將各個 IMF 轉換成彼此之間無相關的獨立成份，並透過研究中所提的「擬偏 F 檢定量」，保留與原始序列具明顯相關的重要分量，將其還原後與趨勢分量 r 相互結合，而成為研究中去除雜訊的時間序列資料，做為預測模型的配適資料。

而在研究中實際預測模型的建立上，則是以上述方法建立混合 ARIMA 模型，並與原始數列直接建立的 ARIMA 模型進行比較。因此，在以台灣工業生產指數的實案資料模擬後的結果可以發現：

- 一、應用研究中所提的 EMD 與 ICA 結合方式所得到之去除雜訊過後的時間序列，其不但在配適 ARIMA 模型上要較直接應用原始序列建構的 ARIMA 模型為佳；同時，於預測能力的表現上亦相當出色。足見結合 EMD 與 ICA，並以顯著性檢定方式決定保留重要分量的預測模型建構流程，極具適用性與可靠性。
- 二、值得一提的是，經由結合 EMD 與 ICA 所建立之去除雜訊的時間序列資料，其於建構 ARIMA 模型過程中，以 ADF 單根檢定確認出其已屬平穩的時間序列；此說明了透過 EMD 與 ICA 將能有效促使時間序列資料呈現平穩結果。此一結果亦可提供傳統時間序列分析時，必須經由差分轉換為平穩的另外一種途徑選擇。

第二節 建議與研究限制

經濟相關領域的預測方式有千百種，通常大多所在意的是預測數據是否與實際結果相符。然而，預測的本質理應是從整個預測過程是否符合邏輯性來看，無論是何種預測方法，往往皆有其局限性，但亦有其可取之處；就以常見的計量方程式而言，在進行模型的預估時，其預測議題的機率分配為何？於已公佈的資訊中，哪些變數是可以放入議題中？又有哪些變數是不可忽略的？不同於自然科學領域，其可在實驗室中將許多的外生變數固定，而經濟預測則面對的是一個開放的環境，這些都是預測時所需要克服，也是經濟預測所需面臨的難題。

以本研究的台灣工業生產指數實案預測為例，所被關心的主軸也許只有一項，就是預測未來指數的走向；想當然爾，未來的真實情況是無法得知的，故而研究中以樣本外的估計（Out of Sample Evaluation）來檢驗模型預

測的能力。但在實際的研究上，雖說擁有一定的樣本期間資料，但眾所周知最新的資料通常隱含著最重要的訊息；換言之，先進行樣本內的估計以作為樣本外的預測，則樣本內的資料長度要多新？需多長？實屬不易。若以台灣 GDP 為例，90 年代之前超過 10% 的成長率，90 年代後該現象則未曾發生，亦即在估計時間序列模型時，若以高成長率的樣本內模型估計低成長率的樣本外資料，往往會發生高估情形。總而言之，本研究實證模型的建立，係來自於樣本內資訊；而樣本外的預測，則僅是以樣本內建立的模型評估之。

其次，除了預測未來的走向外，很少提及模型建立過程的問題，這反而是非常重要的。預測實證結果更進一步的用意，應該是在找出改進預測表現的方法，藉由預測的一連串流程，瞭解問題，並提升預測能力；而本研究的研究過程與觀點，也就是模型的設定與檢驗。也許不同的參數設定、不同變數的選擇，找出來的顯著性可能就不一樣，實證的結果也就不一樣，這也是為什麼可以有各式各樣的理論，建構各式各樣的模型，去解釋同一件議題。因此，本研究藉由 EMD、ICA、顯著性檢定、ARIMA 等研究方法，運用這樣的概念與想法，實際建立模型以解釋未知的資料。期望研究中所設計出的一套具檢定力、客觀且有效移除時間序列資料雜訊的完整程序的解釋能力能被接受；也期望透過本研究實際個案的混合模型建構流程中，是否仍有需被預測或改善之處，以調整模型的建立，改善預測的表現。

參考文獻

中文部分

1. 皮紅梅、劉財、王典(2007)。利用 Hilbert- Huang 變換提取地震信號瞬間參數。石油地球物理勘探，42(6)，418-424。
2. 李泓緯(2011)。以空間性及時間性獨立成份分析與分類迴歸樹為基礎的投資決策模式（未出版之碩士論文）。國立台北科技大學，台北。
3. 呂奇傑、李天行、陳學群(2007)。應用獨立成份分析與支援向量迴歸於財務時間序列預測。資訊管理學報，14(4)，161-183。
4. 吳順德、陳思予、陳虹伯(2009)。經驗模態分解法之研究趨勢探討。臺北科技大學學報，42(1)，31-50。
5. 徐志宏(2009)。台灣景氣落後指標初探。經濟研究，10，35-70。
6. 許秀珊(2008)。新編台灣景氣同時指標之研究。經濟研究，8，51-87。
7. 楊文熙(2003)。股票變化之穩健預測（未出版之碩士論文）。國立中央大學，桃園。
8. 曹麗青、林振山(2008)。基於 EMD 的 HHT 變換技術在長江三峽水庫年平均流量預報中的應用。水文，28(6)，21-24。
9. 經濟部（2012）。工業生產統計年報。台北：經濟部。
10. 葉清江、齊德彰、郭定峪(2011)。結合經驗模態分解法與類神經網路在股價預測之應用。2011 第 14 屆科際整合管理研討會，125-138。
11. 樊海濤、何益斌、周緒紅(2006)。基於 Hilbert-Huang 變換的結構損傷診斷方法研究。建築結構學報，27(6)，114-122。

英文部分

1. Back, A. D., & Weigend, A. S. (1997). A First Application of independent Component Analysis to Extracting Structure from Stock Returns. *International Journal of Neural Systems*, 473-484.
2. Bell, A. J., & Sejnowski, T. J. (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129-1159.
3. Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis forecasting and control*, 2th ed. San Francisco: Holden Day.
4. Cardoso, J. F., & Soudoumiac, A. (1993). Blind beamforming for non Gaussian signals. *Radar and Signal Processing, IEE Proceedings F*, 140(6), 362-370.
5. Chen, C. F., Lai M. C., & Yeh, C. C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, 26, 281-287.
6. Dickey, D., & Fuller, W. (1979). Distribution of the estimators for autoregressive times series with a unit root. *Journal of the American Statistical Association*, 74, 427-431.
7. Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2), 111-120.
8. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, H. Q., Yen, N. C., Tung, C. C., Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *The Royal Society*, A454, 903-995.
9. Huang, N. E., Shen, Z., & Long, S. R. (1999). A new view of nonlinear water waves: the Hilbert spectrum. *Annual Review of Fluid Mechanics*, 31, 417-457.
10. Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent

- component analysis. *Transactions on Neural Networks*, 10, 626–639.
11. Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13, 411-430.
 12. Kamath, V., Lai, Y. C., & Zhu, L. (2006). Empirical Mode Decomposition and Blind Source Separation Methods for Antijamming with GPS Signals, *Position, Location, And Navigation Symposium, 2006 IEEE/ION*, 335-341.
 13. Karhunen, J., Hyvärinen, A., Vigario, R., Hurri, J., & Oja, E. (1997). Applications of neural blind separation to signal and image processing, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, 131-134.
 14. Kiviluoto, K., & Oja, E. (1998). Independent component analysis for parallel financial timeseries, *Proceedings of International Conference on Neural Information Processing*, 2, 895-898.
 15. Lee, T. W. (1998). *Independent Component Analysis; Theory and Applications*, Boston : Kluwer Academic Publishers.
 16. Lewis, C. D. (1982). *Industrial and Business Forecasting Method*, London: Butterworth Scientific.
 17. Mansour, A., & Kawamoto, M. (2003). A Classification of ICA Algorithms According to their Applications & Performances, *Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 86(3), 620-633.
 18. Moody, J. E., & Wu, L. (1997). What is the “true price”?-State space models for high frequency FX data, *Proceedings of the Conference on Computational Intelligence for Financial Engineering*, 150-156.
 19. Salisbury, J. I., & Wimbush, M. (2002). Using modern time series analysis techniques to predict ENSO events from the SOI time series. *Nonlinear Processes in Geophysics*, 9 , 341-345.
 20. Wang, W., Zhao, H., Li, Q., & Liu, Z. (2009). A Novel Hybrid Intelligent Model for Financial Time Series Forecasting and its Application.

International Conference on Business Intelligence and Financial Engineering, 279-282.

21. Wei, Y., & Chen, M. C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks, *Transportation Research Part C: Emerging Technologies*, 21(1), 148-162.
22. Weng, B., Blanco-Velasco, M., & Barner, K. E. (2006). ECG Denoising Based on the Empirical Mode Decomposition. *Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*, 1-4.
23. Wu, E. H. C., Yu, P. L. H., & Li, W. K. (2006). Value at Risk Estimation using independent component analysis-generalized autoregressive conditional heteroscedasticity (ICA-GARCH) models. *International Journal of Neural Systems*. 16(5), 371-382.
24. Yang, H. H., & Amari, S. (1997). Adaptive online learning algorithm for blind separation: maximum entropy and minimum mutual information. *Neural Computation*, 9, 1457–1482.
25. Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting Crude Oil Price with and EMD-based Neural Network Ensemble Learning Paradigm. *Energy Economics*, 2623-2635.