

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

過濾序列雜訊以提升 ARIMA 績效之研究

—以台灣工業生產指數為例

A Study on Enhancing Performances of ARIMA by Filtering the Series
Noise: Using Taiwan Industrial Productive Index as a Case Study

研究生：邱柏誠

Graduate Student：Chiu, Po-Cheng

指導教授：林豐政 博士

Advisor：Lin, Feng-Jeng Ph. D.

中華民國一〇五年八月

August 2016

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應用經濟學碩士班邱柏誠君所提之

論文（題目）： 過濾序列雜訊以提升ARIMA績效之研究_以台灣工業生產指數為例

A Study on Enhancing Performances of ARIMA by Filtering
the Series Noise: Using Taiwan Industrial Productive Index as
a Case Study

係由本人指導撰述，同意提付審查。

指導教授



（簽章）

系所主管



（簽章）

中 華 民 國 105 年 6 月 20 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應用經濟學碩士班邱柏誠君所提之

論文（題目）： 過濾序列雜訊以提升ARIMA績效之研究_以台灣工業生產指數為例

A Study on Enhancing Performances of ARIMA by Filtering
the Series Noise: Using Taiwan Industrial Productive Index as
a Case Study

經本委員會審議，認定符合碩士資格標準。

學位考試委員

洪博仕 葉清記 林豐政

指導教授

林豐政

中華民國105年6月20日

論文名稱：過濾序列雜訊以提升 ARIMA 績效之研究－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應用經濟學碩士班

頁數：46

畢業時間：中華民國一〇五年八月

學位別：碩士

研究生：邱柏誠

指導教授：林豐政 博士

摘 要

本文主要提出一個於建立 ARIMA 模型時，過濾原資料序列雜訊以提升模型績效的程序；係針對原資料序列與其 ARIMA 預測序列間之預測誤差，應用經驗模組分解(Empirical Mode Decomposition, EMD)後的數個本質模組函數(Intrinsic Mode Function, IMF)，透過獨立成份分析(Independent Component Analysis, ICA)分解出誤差獨立成份後，藉由偏相關係數的概念，將預測誤差序列中屬於雜訊的獨立成份予以確認，還原後於原資料序列中過濾之；且再應用已過濾雜訊後的資料序列，重新建立一組混合 EMD、ICA 及 ARIMA 之混合模型；此外，再參考國內學者研究之程序，建立 EMD-ARIMA 模型，據以進行三種模型績效的評估。

研究中以台灣工業生產指數為研究對象，嘗試建構得以適切解釋該指數變化的 ARIMA 模型，並依確認後之模型進行預測工作；同時，將所蒐集之資料依移動視窗法之預測方式劃分為 15 組時間序列資料，並驗證經由研究中所提之過濾原資料序列雜訊的程序建構之 ARIMA 模型的實用性。

結果顯示：透過研究中所提之方法而重新建立的 ARIMA 模型，無論在模型適合度或預測能力的指標上，相對於直接應用原資料序列所建構之 ARIMA 模型的績效，確實較為改善。

關鍵詞：經驗模組分解、獨立成份分析、ARIMA 模型、偏相關係數

Title of Thesis : A Study on Enhancing ARIMA Performances by Filtering the Series Noise: Using Taiwan Industrial Productive Index as an Example

Name of Institute : Department of Applied Economics and Management, Master Program of Business Administration, National Ilan University

Page : 46

Graduation Time : August, 2016

Degree Conferred : Master

Student Name : Chiu, Po-Cheng

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

Abstract

In this paper, one process that try to enhance performances of ARIMA model by filtering noise from the original data was proposed. It integrated the Empirical Mode Decomposition (EMD) with Independent Component Analysis (ICA) to treat forecasting errors between the original data and forecasts of the traditional ARIMA.

Furthermore, the process identified the non-significant ICAs as noise by using partial correlation test, and reduced it from the original data. A hybrid-ARIMA model was also augmented with this process to treat de-noise data series. The model fit and forecasting performance of the hybrid-ARIMA were compared to those of the traditional ARIMA which used the original data in the empirical case of industrial productive index of Taiwan, and EMD-ARIMA from other scholar's research.

The final results of this paper show that the hybrid-ARIMA outperformed its traditional model, and the proposed hybrid forecasting approach is feasible and reliable in the empirical case.

Keyword: Empirical Mode Decomposition, Independent Component Analysis, ARIMA Model, Partial Correlation Coefficient

誌謝

研究所的兩年時間很快就過了，碩士生涯終於也進入尾聲，一路走來受到許多人的幫助。首先，當然最感謝的是指導教授豐政老師，從老師身上學習到了許多知識，不論是專業上或是生活上，到後來的論文撰寫，老師皆給予了許多幫助，也才能如此順利的完成研究。同時，亦感謝兩位口試委員，博仕老師以及清江老師，對於論文提供了諸多建議與指正，使論文得以更加完善。此外，還要感謝系上的老師們、綺貞與佶俐兩位學姊，課業上的盡心指導與鼓勵，以及學習的大小瑣事。

也要感謝銘仁學長對於題目選取以及軟體操作上的教導，感謝宇梵學姊幫助論文的校對，感謝瑋瑋學姊在課業上的幫助，以及同學郁軒、康桀、意涵、凱洧、映潔、克霖、琇涵、玉茹、瓊妙、宗益、誼娟、雯茹、岱樺、定定以及小民平時生活上的協助、訊息的交流。還有學弟妹姊縈、家鳳、俊宏、宇賢、浩哲、致宏、貝承、小蘭以及苡均，感謝你們。

特別感謝芷吟，謝謝在這幾年的時間一直有你在身邊，在煩悶的時候能有你陪著，經過了大大小小的事情，也讓我學到了許多，真的非常的感謝你。

最後要感謝最愛的家人，有了你們的支持，才能無後顧之憂的向前走、向前邁進，才能更堅定的完成研究所的學業。

邱柏誠謹誌

國立宜蘭大學應用經濟與管理學系應用經濟學碩士班

中華民國一〇五年八月

目錄

摘要.....	i
Abstract.....	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	v
圖目錄.....	vi
第壹章 緒論.....	1
第一節 研究動機與目的.....	1
第二節 研究範圍	2
第三節 研究流程	2
第貳章 文獻探討.....	4
第一節 經驗模組分解法.....	4
第二節 獨立成份分析	5
第三節 工業生產指數	7
第參章 研究設計.....	11
第一節 研究架構	11
第二節 EMD.....	13
第三節 ICA.....	18
第四節 IC 排序與篩選過程.....	21
第五節 衡量指標	22
第肆章 實證分析.....	24
第一節 時間序列資料說明.....	24
第二節 個案傳統 ARIMA 模型建立與預測	24
第三節 個案 EMD-ARIMA 模型.....	27
第四節 個案 EMD-ICA-ARIMA 模型.....	31
第五節 個案模型比較	37
第伍章 結論與建議.....	39
第一節 結論.....	39
第二節 建議.....	39
參考文獻.....	41
一、中文部份	41
二、英文部份	42
附錄.....	44

表目錄

表 2-1 EMD 相關議題文獻之彙整表	5
表 2-2 ICA 相關議題文獻之彙整表	6
表 2-3 台灣工業生產指數相關議題文獻之彙整表	9
表 2-4 台灣同時景氣指標修訂構成項目表	10
表 3-1 衡量指標種類	23
表 4-1 資料分組表	24
表 4-2 原始序列之單根檢定表	25
表 4-3 原始序列經一階差分後之單根檢定表	25
表 4-4 原始序列之 ARIMA 暫定模型殘差檢定 Q 統計量表	26
表 4-5 原始序列之 ARIMA 暫定模型適合度檢測表	26
表 4-6 個別 IMF 與原始資料 Y_t 之偏相關係數表	28
表 4-7 各階段解釋能力比較表	28
表 4-8 Y_t' 之單根檢定表	29
表 4-9 Y_t' 經一階差分後之單根檢定表	30
表 4-10 EMD-ARIMA 暫定模型殘差檢定 Q 統計量表	30
表 4-11 EMD-ARIMA 暫定模型適合度檢測表	31
表 4-12 個別 IC 與誤差序列之偏相關係數表	34
表 4-13 各階段解釋能力比較表	34
表 4-14 Y_t^c 之單根檢定表	35
表 4-15 Y_t^c 經一階差分後之單根檢定表	36
表 4-16 EMD-ICA-ARIMA 暫定模型殘差檢定 Q 統計量表	36
表 4-17 EMD-ICA-ARIMA 暫定模型適合度檢測表	37
表 4-18 三模型之預測能力比較表	38

圖目錄

圖 1-1 研究流程圖	3
圖 2-1 HHT 處理架構流程圖	4
圖 3-1 研究架構圖	12
圖 3-2 本質模組函數圖例	13
圖 3-3 IMF 篩選過程流程圖	14
圖 3-4 IMF 篩選過程流程圖(續)	15
圖 3-5 EMD 構成圖	16
圖 3-6 EMD 之各殘餘量累積圖	17
圖 4-1 台灣工業生產指數歷史走勢圖(2002/01~2015/12).....	24
圖 4-2 原始資料之 IMFs 與趨勢分量 R	27
圖 4-3 過濾雜訊後之序列資料 Y'_t	29
圖 4-4 誤差序列之 IMFs 與趨勢分量 R 序列圖	32
圖 4-5 誤差序列的 5 個 IC	33
圖 4-6 過濾雜訊後之序列資料圖	35
圖 4-7 傳統與混合模型的預測序列圖	38

第壹章 緒論

第一節 研究動機與目的

一、研究動機

在過去研究時間序列的預測議題時，於定量分析(Quantitative Analysis)的方法中，主要是透過歷史數據經由數理方法進行評估，其大致上可分為傳統的預測模型(如：迴歸分析、指數平滑、ARIMA 模型等)與機器學習(Machine Learning，如：類神經網路、支援向量機等)。其中，前者主要應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，後者則係應用近些年來發展出具學習特性的人工智慧，進而運用於時間序列的預測上。

然而，一般實案研究資料多屬量測型的資料，其無論是應用於一般傳統的時間序列預測或是電腦學習的方法，同樣於面對資料(數據)呈現不良訊號時(高頻、雜訊、非平穩)，經常無法完整準確地表達出資料之變化及其趨勢，以致在進行預測時，往往會偏離資料原本性質，導致其於建立預測模型過程中可能產生過度配適或配適不足，因而降低其之可信程度。因此，如何平穩資料(數據)以降低雜訊所帶來錯誤訊息，進而建立一個可靠、可信的預測模型，將是一個重要的研究議題。

在台灣，工業生產指數是反映工業發展速度的重要指標，其含括礦業、製造業、水電燃氣以及房屋建築業等等，且這些產業的生產總值佔整體經濟比重較大，對於經濟變化的反應亦十分敏感；因此，工業生產指數實為觀察景氣變動的重要量測指標。據此，研究中將以台灣工業生產指數為實案資料，提出結合 EMD 與 ICA 的概念，將原資料序列與其 ARIMA 預測序列間的預測誤差，透過 EMD 分解後的數個 IMF，再經由 ICA 分解出其獨立成份(ICs)，同時應用偏相關係數的概念以確認屬於雜訊的獨立成份，將其還原後由原資料序列中予以過濾之；並將其視為已過濾雜訊的時間序列資料，而重新建立 EMD-ICA-ARIMA 模型，進行模型配適與預測能力的評估，並據以探討本研究所提之過濾原資料序列雜訊的流程之適用性與可行性。

二、研究目的

因此，本研究之研究目的如后：

- (一) 將預測序列與實際值間的誤差序列，藉由 EMD 分解後之具不同特性的 IMF 分量，結合 ICA 法使其形成彼此之相互獨立的成份。
- (二) 應用偏相關係數法，將各 ICs 依重要程度排序，並嘗試過濾資料雜訊，期能降低雜訊所帶來之影響。
- (三) 應用移動視窗法之預測程序，將資料分為多組訓練組與預測組進行實案測試，據以提升研究結果的客觀性。
- (四) 對三種模型之模型適合度與預測能力進行評估。

第二節 研究範圍

本研究主要提出應用 EMD 結合 ICA 之方式，並加上偏相關係數法以過濾誤差序列中的雜訊；同時，亦以台灣工業生產指數為研究的實案資料。而資料來源取自國家發展委員會(原行政院經濟建設委員會)每月公佈之景氣同時指標，資料蒐集期間為 2002 年 1 月至 2015 年 12 月，合計 168 筆。另，研究中依移動視窗法之預測程序，將資料依訓練組年數劃分為 9 年、8 年、7 年共三期間，每一期間中又含括 2011 年至 2015 年共五年的預測年份，並以此劃分進行後續的分析與評估。

第三節 研究流程

本研究的流程如圖 1-1 所示。首先，列出本研究之研究動機，並延伸出研究目的。其次，完整瞭解目前台灣景氣同時指標之發展以及工業生產指數的定義，並針對 EMD 以及 ICA 的相關議題文獻進行回顧，以此為根據提出改良，並設計出一個過濾誤差序列雜訊之完整流程。最後，再以台灣工業生產指數為實案資料，應用研究中所設計之流程以實際過濾資料雜訊，藉以建立其之混合 ARIMA 模型，並加以分析及評估。

因此，本研究共分為五個章節，各章節內容如下：

第壹章 緒論

說明本研究的動機與目的，以確立研究範圍與研究流程。

第貳章 文獻探討

依探討主題尋找相關文獻，將國內外應用 EMD 與 ICA 於時間序列上的文獻，進行簡要的回顧與整理，並對目前台灣工業生產指數的定義詳實介紹。

第參章 研究設計與方法

描述研究中所提之結合 EMD 與 ICA，再以偏相關係數法過濾誤差序列之雜訊的程序；並應用於 ARIMA 模型的分析架構與流程。同時，亦詳細描述 EMD 法與 ICA 法的基本原理。

第肆章 實證分析

研究中以台灣工業生產指數為實案資料，配合研究目的，詳細地說明分析與評估的結果。

第伍章 結論與建議

將實證分析的結果進行歸納與探討，並對後續之研究進行建議。

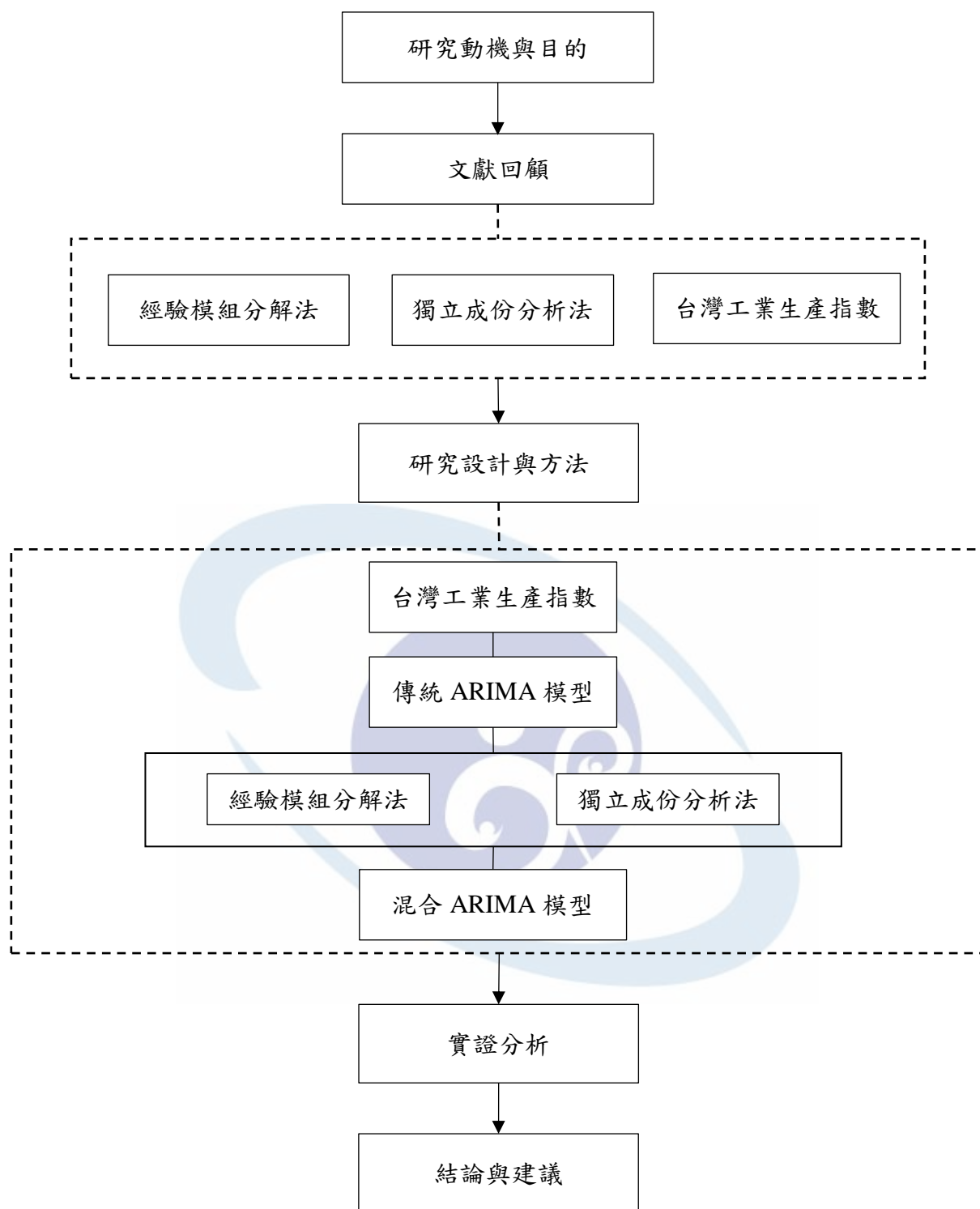


圖 1-1 研究流程圖
資料來源：本研究整理。

第貳章 文獻探討

為能提出結合 EMD 與 ICA 以過濾資料中的雜訊，並進行實案預測模型模擬之實證研究；以下將分別就 EMC 與 ICA 的相關議題文獻進行回顧。另外，亦針對實證分析中所使用之台灣工業生產指數的定義進行介紹。

第一節 經驗模組分解法

一、希爾伯特－黃轉換

Huang 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)分析方法之後，開始廣泛應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(皮紅梅、劉財、王典，2007；黃宜正、林俊賢，2013；樊海濤、何益斌、周緒紅，2006；Kamath, Lai, Zhu, & Urval, 2006；Weng, Blanco-Velasco, & Barner, 2006)，圖 2-1 即為 HHT 處理架構流程圖。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；而後，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作用，是依據訊號時間與頻率透過篩選的迭代程序以取得 IMFs。由於 IMF 會隨著訊號而改變，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)。詳細的 EMD 分解過程，將於研究方法中介紹之。

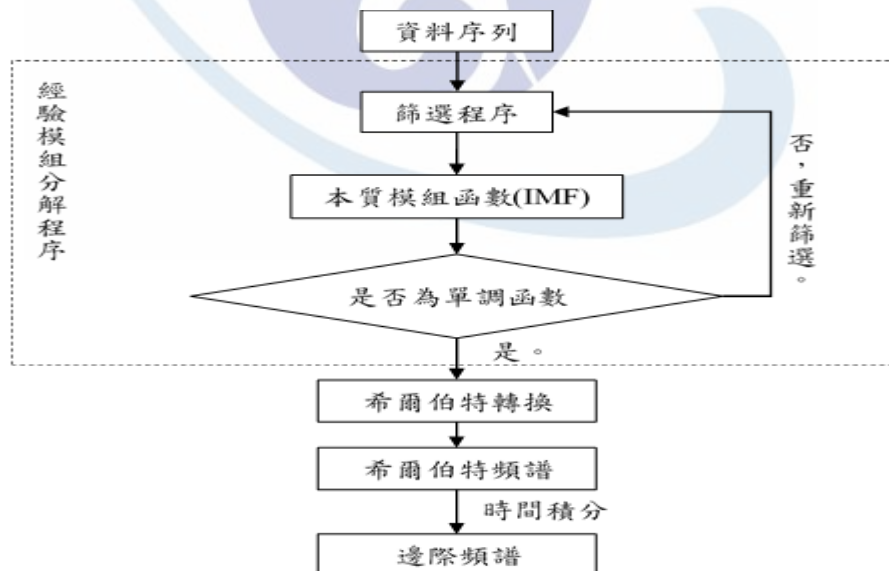


圖 2-1 HHT 處理架構流程圖

資料來源：Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering sciences*, 454(1971), 903-995.

二、運用希爾伯特－黃轉換於時間序列預測

本研究將過去學者應用 EMD 於各領域中之相關議題文獻，整理於表 2-1 中。

表 2-1 EMD 相關議題文獻之彙整表

學者(年份)	應用領域	結果
Salisbury & Wimbush (2002)	南方震盪指數	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)	台灣加權指數、道瓊工業指數、那斯達克指數及其報酬率資料	研究中透過重要 IMF 分量的組合，將可作為指數與報酬率的預測。
曹麗青、林振山 (2008)	中國三峽水庫的年平均流量	去除最高頻率之第一個 IMF 分量之後，再對水庫流量進行預測。
Yu, Wang, & Lai (2008)	西德州中級原油與布蘭特原油之期貨價格	將資料先以 EMD 分解，並應用前饋式類神經網路(FNN)預測期貨價格。
Wang, Zhao, Li, & Liu (2009)	金融領域	結合支援向量迴歸(SVR Support Vector Regression)建立預測模型，以解決該領域之序列高波動且複雜的資訊。
Chen, Lai, & Yeh (2012)	日本與港澳來台的人數	結合倒傳遞類神經網路(BPN)模型，預測日本與港澳來台的人數
王僑宏 (2013)	颱風的降雨情況	以 ARIMA、SARIMA、EMD 以及極速學習機(ELM) 等組合成複合預測模型，分別對莫拉克與賀伯颱風的降雨情況，進行長時間尺度之逕流量與短時間尺度之時雨量的預測。
Lin (2015)	工業生產指數	研究中透過增量 R^2 ，由 IMF 中選出雜訊，並改善預測模型績效。

資料來源：本研究整理。

綜合上述國內外文獻可發現：運用 EMD 技術於時間序列資料時，可有效降低訊號的非平穩性質；但多數學者於 IMF 分量的選擇上，往往偏向於透過自然的週期循環、文獻分析等主觀歸納方式以決定(選定)IMF，鮮少採行探討各個分量間可能存在的相關性，以及對於時間序列隱含的獨立解釋能力，故而值得加以探究。另，本研究亦參考 Lin (2015) 所提之程序進行分析，並與本研究之分析結果進行比較。

第二節 獨立成份分析

獨立成份分析(ICA)是一項訊號處理的技術，其某些相關理論亦包含在資訊與統計等各個領域中；方法最初目的是用來處理未知來源分離(Blind Source Separation, BSS)的

問題(Karhunen, Hyvärinen, Vigario, Hurri, & Oja, 1997)，其特色為：最初蒐集到的資料皆是以未知來源之訊號經線性混合而成，再應用 ICA 從混合訊號中分離出未知訊號。其中，用以描述 BSS 問題最著名的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)。但由於該分析法之計算極為繁複，直到 Hyvärinen (1999) 提出快速獨立成份分析(Fast Independent Component Analysis, Fast ICA)，方才使得計算效率提高，且能有效解決資料在統計上分配未知的問題，因而 ICA 才開始大量的被應用於各領域。

一、相關議題及文獻

近幾年有些學者開始將 ICA 應用在財務時間序列資料上，使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 IC 即可解釋之(Back & Weigend, 1997)。本研究將過去學者探討過之 ICA 相關議題，整理於表 2-2 中。

表 2-2 ICA 相關議題文獻之彙整表

學者(年份)	應用領域	結果
Moody & Wu (1997)	國際外匯市場	應用 ICA 將國際外匯市場的時間序列資料的雜訊分離出來，以提高未來預測的準確度。
Kiviluto & Oja (1998)	零售商店的現金流量	從 40 家零售店的現金流量找出其 IC，發現相近零售店的 IC 變動情形有較相近的趨勢。
Wu, Yu, & Li (2006)	金融指數	應用 ICA 與 GARCH 等等模型，進行各項金融指數風險值的評估；結果顯示，其在捕捉隨時間變動的波動性上優於其他比較模型。
呂奇傑、李天行、 陳學群 (2007)	日經 225 指數	應用 ICA 得到過濾雜訊的預測變數後，再以 SVR 建立其預測模型，發現具有良好的預測能力。
李泓緯 (2011)	投資決策模式	發現 ICA 以及分類決策數，於各種績效衡量方面不僅良好，且亦優於其他傳統方法。
Lin, Tung, & Huang (2012)	數據分析	應用 ICA 與 EMD，對擷取訊息加密的數據進行分析，結果發現其能有效的萃取兩個原始訊號之數據。
蘇恆信 (2012)	工業生產指數	結合 EMD 與 ICA，再應用擬偏 F 統計量，將序列中的雜訊過濾，結果發現能有效改善預測模型績效。
何川、舒勤、賀含 峰 (2014)	負荷預測	在將 ICA 特徵萃取後的數據做為神經網絡的輸入，不僅可減小神經網絡的輸入量並提高效率，還能有效的提高精確度。

資料來源：本研究整理。

由表中過去學者的文獻得知，ICA 亦可結合其他分析方法，以進一步消除時間序列資料中的雜訊，並藉以提升預測結果的品質。而在 ICA 演算法中，Fast ICA 是較為普遍常見的演算法。

二、小結

由前面的文獻回顧可以發現，雖然使用 EMD 分解出的 IMF 分量可選擇作為未來預測的變數，然而這些選取變數的方式往往係由研究者本人，透過自然環境、文獻分析等主觀歸納方式以決定之，而且 IMF 分量間並不能保證彼此具獨立的貢獻性，而 ICA 則顯示出可透過重要的少數 IC，即可勾勒出原資料的結構與趨勢。因此，研究中擬將原資料序列與其 ARIMA 預測序列間的預測誤差，透過 EMD 產生的 IMFs，再經由 ICA 而將其 IC 分解出來；同時，將確認為雜訊之序列還原後從原資料序列中予以過濾之，並將過濾後的資料序列視為已過濾雜訊的時間序列資料，據以重新建立 ARIMA 模型，而進行模型適配與預測能力的評估。

第三節 工業生產指數

台灣工業生產指數，係屬台灣景氣同時指標中之內含項；以下本研究將先介紹台灣景氣指標及工業生產指數之發展，再進行其相關文獻之探討。

一、台灣景氣指標的發展

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、同時指標以及落後指標，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、同時指標以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以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、1987 年、2007 年以及最近一

次修訂之 2013 年(如表 2-3)，先後共 4 次的修訂。近年來，由於國內外經濟環境劇烈變遷，先是 2008 年至 2009 年的金融海嘯，促使全球經濟陷入二次大戰後首波的衰退；後因各國採行積極性擴張性政策，加上以中國為主的新興國家高速的成長帶動，全球景氣於 2009 年走出谷底。然而，之後復甦腳步並不順遂，金融海嘯期間各國政府大規模的金融與財政措施，對於原本公共債務比相對偏高的先進國家，無疑是雪上加霜；因此，在 2011 年時全球經濟再度受到歐債危機干擾，逐漸地走下坡。國際大環境急速變化，使得近幾年台灣的景氣波動劇烈，於 2007 年修訂之景氣燈號構成項目與檢查值，將來未必完全適用。因此，在重新檢討之後，於 2013 年 8 月進行修訂，篩選出合適之構成項目，最後合成新的同時指標，並與目前指標進行比較。結果顯示：目前同時指標構成項目之「商業營業額指數」因略具領先性，建議改以名目之「商業營業額」替代之(徐志宏，2013；黃月盈，2013)。

二、台灣工業生產指數

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。其目的與用途在於瞭解每月工業產銷存量消長情勢，並提供產業政策及經濟建設規劃應用，與業者調節產銷存量及採取因應對策參考。另，亦可提供學術研究、經濟分析以及編制其他經濟指標之依據。

工業生產調查的週期為每月調查與編製一次，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，均是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以及行業別，而採用不同的抽出法，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，2015)。

調查的產品則依據經濟部統計處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 693 項產品群，並將其編列報表，而以 2011 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，將會產生不同的差異。而台灣工業生產指數則是採用拉式公式(Laspeyres's Formula)基期加權總值式，其計算期工業生產指數公式如下：

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

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。

三、工業生產指數相關文獻探討

以下表 2-3 將列出過去學者對於工業生產指數之相關研究。

表 2-3 台灣工業生產指數相關議題文獻之彙整表

學者(年份)	應用方法	結果
饒秀華、林修葳、黎明淵 (2001)	馬可夫轉換模型 (MS model)	採用分期 MS 模型，將 MS 模型的高、低年增率期望值與變異數設定為兩組或兩組以上，以描述開發中國家在不同經濟發展階段，總體變數之景氣循環波動情形。
林淑華 (2010)	向量自我迴歸模型 (VAR)	利用向量自我迴歸模型(VAR)來進行分析，並以 Granger 因果關係檢定及衝擊反應函數，來了解物價、利率、貨幣供給額、工業生產指數、油價及匯率因素之相互影響關係。

資料來源：本研究整理。

表 2-4 台灣同時景氣指標修訂構成項目表

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)	2013年 景氣綜合指標 (Composite Index)
1.工業生產指數變動率	1.工業生產指數變動率	1.工業生產指數變動率	1.工業生產指數	1.工業生產指數
2.製造業生產指數	2.製造業生產指數變動率	2.製造業生產指數變動率	2.電力(企業)總用電量	2.電力(企業)總用電量
3.製造業銷售值	3.製造業銷售值	3.製造業銷售值	3.製造業銷售量指數	3.製造業銷售量指數
4.票據交換變動率	4.票據交換變動率	4.票據交換金額變動率	4.商業(批發、零售及餐飲業)營業額指數	4.商業(批發、零售及餐飲業)營業額
5.國內貨運量	5.國內貨運量	5.國內貨運量	5.非農業部門就業人數	5.非農業部門就業人數
6.退票率	6.退票率	6.製造業平均每月薪資變動率	6.實質海關出口值實質	6.實質海關出口值實質
	7.製造業平均每月薪資變動率		7.機械及電機設備進口值	7.機械及電機設備進口值
	8.全體銀行等儲蓄性存款變動率			

資料來源：黃月盈 (2013)。台灣景氣同時指標之檢討與修正。《經濟研究》，行政院經濟建設委員會經濟研究處，13，43-78。

第參章 研究設計

本章節將介紹實證分析過程中所使用到的方法之基本原理；其中，包括 EMD 與 ICA。而研究中於進行 EMD 分解時，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；而 ICA 的程式碼部份，則是由阿爾托大學理工學院(Aalto University School of Science and Technology)所提供。另外，建立 ARIMA 模型時，則是使用 SAS 統計軟體。

第一節 研究架構

有別於以往時間序列研究中，直接對一組時間序列資料進行分析，並應用 EMD 分解後選擇的 IMF 分量以建立預測模型。研究中認為直接應用具雜訊的原始序列以建立預測模型，這些雜訊可能會影響到模型適合度，甚至是影響模型的預測能力；而且經過 EMD 分解後之各個 IMF 間亦有可能彼此之間具相關性，對時間序列具隱含的解釋能力，若直接移除一個或數個分量後建立預測模型，可能會喪失許多真實訊息，或是誤將雜訊引進預測模型當中。因此，研究中提出可針對預測誤差進行 EMD 分解，並結合 ICA，將各個 IMF 分量轉換成為數個 IC，亦即將各 IMF 分量間相似之特徵歸於同一 IC，同時保持各 IC 彼此間獨立且無關之後，再應用偏相關係數，依各個 ICs 的重要程度排序，並應用增量 R^2 之概念以選取保留重要的 IC；同時，再將其還原後與趨勢分量 R 相互結合，並視其為已過濾雜訊後之時間序列資料。因此，研究中除了參考 Lin (2015) 之研究，建立 EMD-ARIMA 模型外，另將建立結合 EMD 與 ICA 的混合 ARIMA 模型。其架構如圖 3-1 所示，流程並說明於後：

- (一) 蒐集研究樣本歷年的月份資料。
- (二) 將原始資料應用 EMD 分解成 i 個 IMF 以及趨勢分量 R
- (三) 以偏相關係數法將 i 個 IMF 依重要程度排序，並依增量 R^2 的概念保留具貢獻的重要 IMF，將剩餘不重要的 IMF 移除後，建立一組 EMD-ARIMA 模型。
- (四) 以原始資料直接建立其傳統 ARIMA 模型，另取其預測誤差以進行下一階段步驟。
- (五) 應用 EMD 將預測誤差分解成 j 個 IMF 以及趨勢分量 R_e 。
- (六) 將 IMF 經由 ICA 過程轉換為 ICs，以確認各個分量之間相互獨立且無關。
- (七) 同樣以偏相關係數法將各 ICs 依重要程度排序，並保留具貢獻的重要 IC，而還原剩餘不重要的成份後，將其從原始序列中移除之，據以成為已去除雜訊之新的時間序列資料。
- (八) 應用新的時間序列資料，建立結合 EMD-ICA 的混合 ARIMA 模型。
- (九) 進行傳統 ARIMA、EMD-ARIMA 與 EMD-ICA 混合 ARIMA 模型的適合度與預測能力比較。

其次，為能說明研究架構的可行性與可靠性，研究中將應用台灣工業生產指數的資料序列為個案，建立預測模型並驗證與比較之。

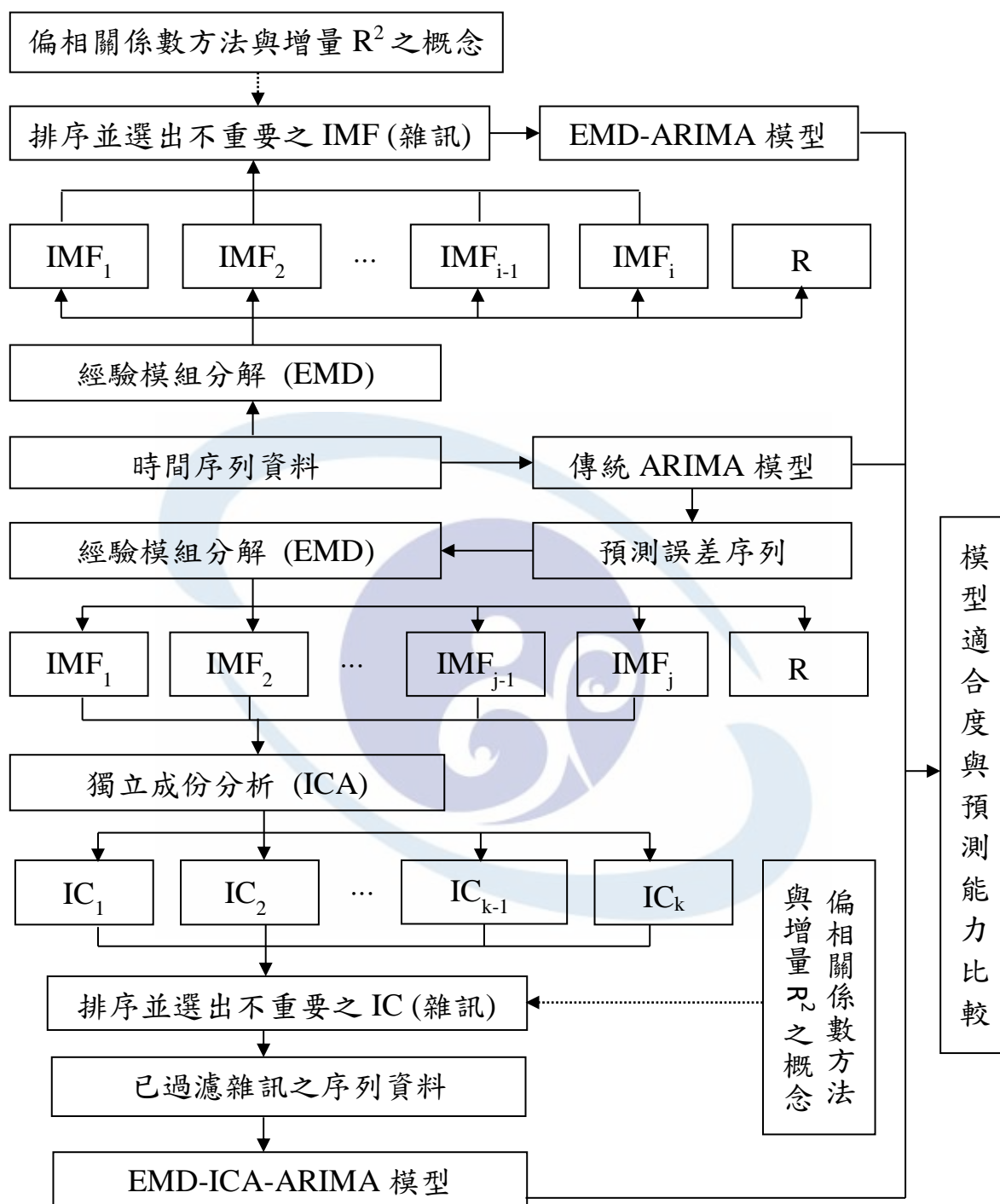


圖 3-1 研究架構
資料來源：本研究整理。

第二節 EMD

一、本質模組函數(IMF)

Huang et al. (1998)應用 EMD 將時間序列資料分解成若干個 IMFs，且不同的 IMF 間則具有不相等的週期與頻率，也具有不等的能量或權重。而 EMD 的分解需根據下列的假設：

- (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)是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)又存在，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)所提出的有關風速研究中的一個典型的 IMF。

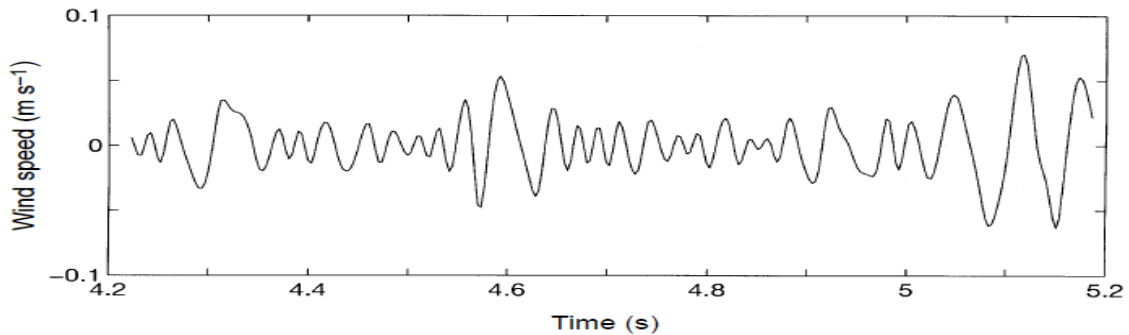


圖 3-2 本質模組函數圖例

資料來源：Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering sciences*, 454(1971), 903-995.

二、EMD 分解過程

如前所述，EMD 係藉著不斷重複的篩選過程(Sifting Process)，逐步地找出 IMF 分量。若以時間序列 $Y(t)$ 為例，則其篩選的過程如下：

步驟 1：找出時間序列 $Y(t)$ 中所有局部極大與極小值。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，並求出上、下包絡線的平均值，即可得均值包絡線 $m_1(t)$ 。

步驟 2：將原時間序列與均值包絡線相減，則可得第一個預備分量 $h_1(t)$ 。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 與步驟 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，而粗體實線為其均值包絡線 $m_1(t)$ ；圖(c)則為求得之預備分量 $h_1(t)$ 。

步驟 3：檢查預備分量 $h_1(t)$ 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，再進行第二次篩選。亦即

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

如此重複篩選 k 次

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

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

至於篩選的過程停止之標準有以下兩個：

1. 當數據通過 0 的個數與極值個數相同時，即可停止。
2. 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。

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。當 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，故而訂此為其停止的標準。

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

圖 3-4 即為 EMD 分解過程中，步驟 3 的實際流程。其中，圖(a)為 $h_2(t)$ ，而圖(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有的物理特性，因此篩選次數的準則大多是聚焦在振幅與頻率上。

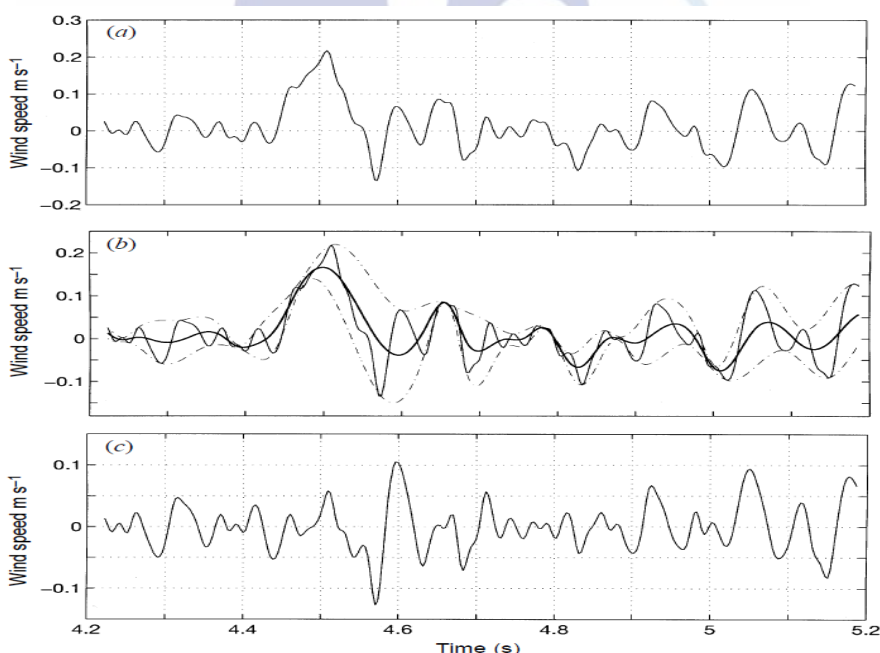


圖 3-3 IMF 篩選過程

資料來源：Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering sciences*, 454(1971), 903-995.

步驟 4：將原訊號減去第一個 IMF 分量，將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 $R_1(t)$ 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：

$$R_2(t) = R_1(t) - C_2(t), \quad (3-6)$$

$$R_3(t) = R_2(t) - C_3(t), \quad (3-7)$$

⋮

$$R_n(t) = R_{n-1}(t) - C_n(t). \quad (3-8)$$

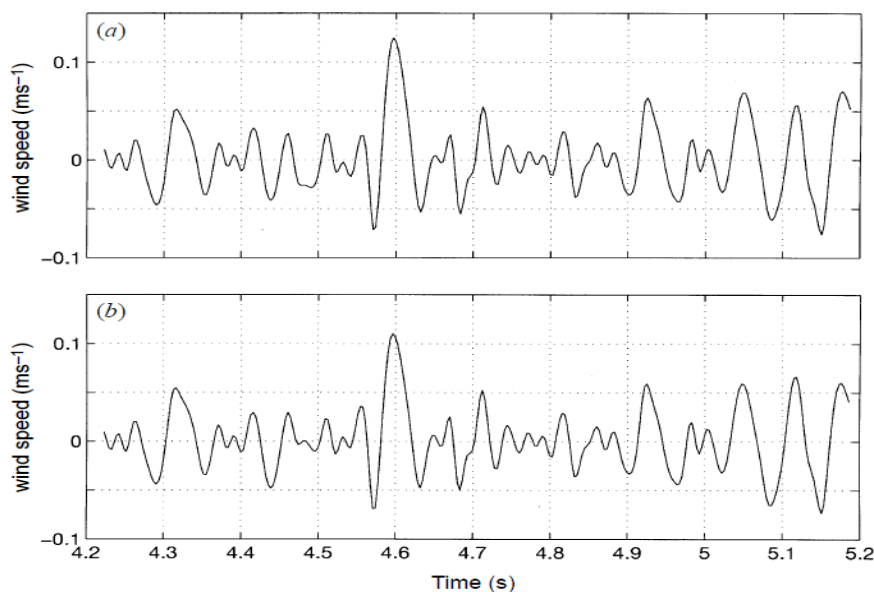


圖 3-4 IMF 篩選過程(續)

資料來源：Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering sciences*, 454(1971), 903-995.

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，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 $R_n(t)$ 進行加總，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透過篩選的迭代(Iterative)程序取得，故其為一組具有適應性(即 IMFs 可隨訊號改變)的時間序列資料。

圖 3-6 為 EMD 過程中各個殘餘量依序加總後所畫出之圖形；其中，圖(a)為趨勢 R ，圖(b)為趨勢 R 加上 C_8 ，圖(c)為趨勢 R 加上 C_8 及 C_7 ，以此類推，直到圖(j)為趨勢 R 加上 C_8 至 C_1 ，即為原始訊號 $Y(t)$ 。其若以數學式表之，則如下所示：

$$Y(t) = C_1(t) + C_2(t) + \cdots + C_n(t) + R_n(t) \quad (3-9)$$

式中， R 亦為原序列之趨勢序列。

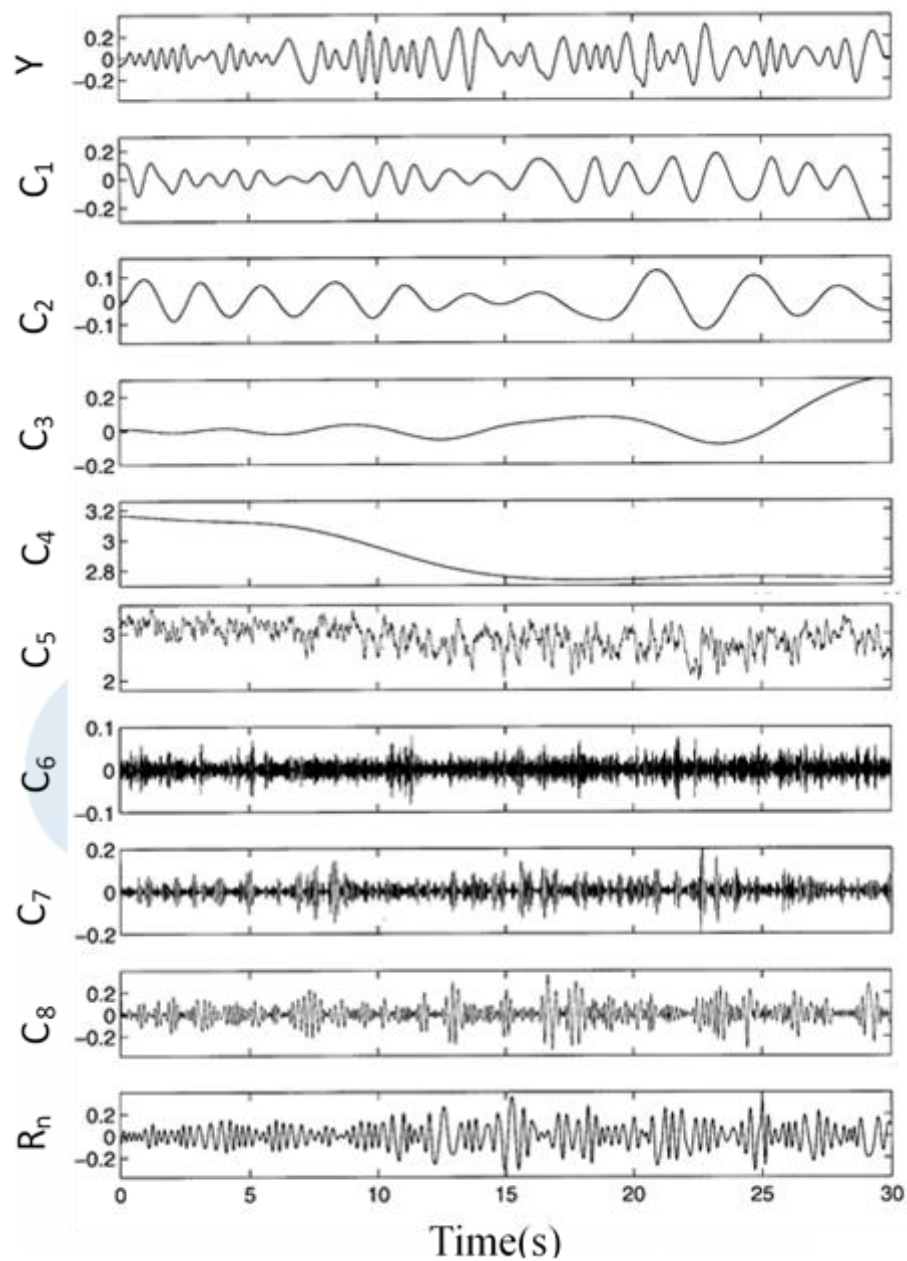


圖 3-5 EMD 構成圖

資料來源：Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering sciences*, 454(1971), 903-995.

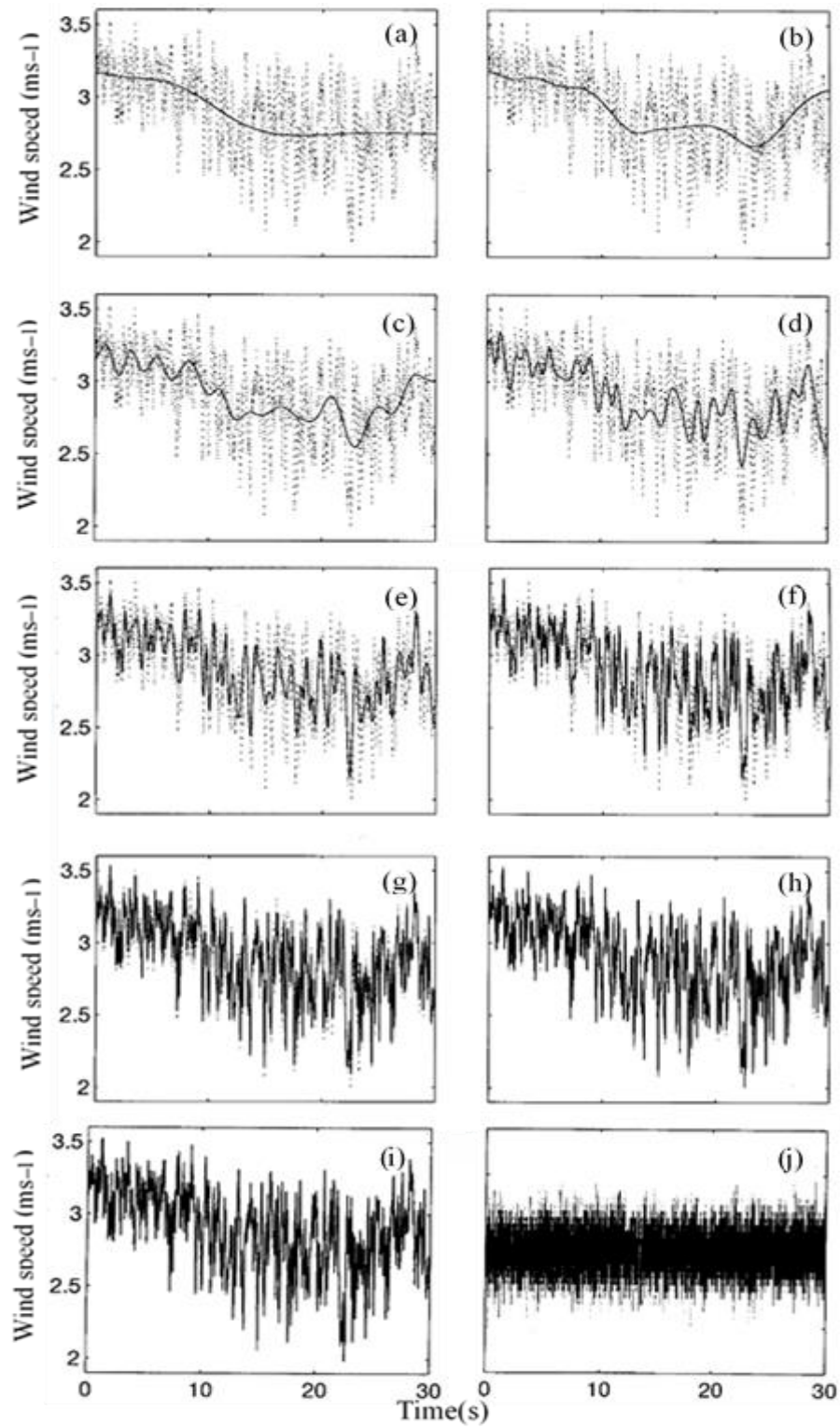


圖 3-6 EMD 之各殘餘量累積圖

資料來源：Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering sciences*, 454(1971), 903-995.

第三節 ICA

一、基本原理

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 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 IC 之間為相互獨立，但並非滿足高斯分配。

ICA 可運用於揭露隨機變數、測量值或訊號底下的隱藏因子；主要即是用來處理未知來源分離的問題，其最常被舉的例子為雞尾酒派對問題。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。但對於機器而言，則必須靠一些複雜的方法才能把混合的聲音分離，而此種分離的目的就是處理未知來源分離(BBS)。

以下透過雞尾酒派對問題來解釋 ICA 處理 BBS 的問題：

假設在一個空間內有 N 個聲音源 s ，空間內另一邊亦設置 M 支麥克風接收訊號 y ，分別放在不同的位置進行收音。則每支麥克風收到的訊號可以用以下數學式表示：

$$y_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + \cdots + a_{1N}s_N(t) \quad (3-10)$$

$$y_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + \cdots + a_{2N}s_N(t)$$

$$\vdots$$

$$y_M(t) = a_{M1}s_1(t) + a_{M2}s_2(t) + \cdots + a_{MN}s_N(t), \text{ 其中 } M \geq N$$

若以矩陣方式表示，則可假設觀測到的混合訊號矩陣為 Y ，各組來源矩陣為 S ，混合矩陣則為 A ，此三者的關係如下：

$$Y = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_M(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{bmatrix} = AS \quad (3-11)$$

其中， $y_j(t)$ ：在 t 時間的混合訊號， $j=1,2,\dots,M$

$s_i(t)$ ：在 t 時間的來源訊號或潛在因素， $i=1,2,\dots,N$

a_{11} ：混合矩陣(Mixing Matrix)， $M \times N$

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 $S = A^{-1} \cdot Y$ ；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 ；此處的矩陣 W 在 ICA 中的稱為解混合矩陣(De-mixing Matrix)，目的是用來將混合數列還原為近似於最初狀態的獨立矩陣 $X = W \cdot Y$ ，進而將獨立成份 X 用以估計未知潛在來源訊號 S 。以矩陣表示如下：

$$X = W \cdot Y \approx S \quad (3-12)$$

$$S \approx X = W \cdot Y \Rightarrow W \approx A^{-1} \quad (3-13)$$

二、ICA 的基本假設

如前所述，ICA 可用來解決未知來源分離的問題，而為了有效的分離混合訊號與辨識訊號來源，其本身有三種基本假設：

(一) 來源訊號或潛在因素必須彼此相互獨立：

因 ICA 演算法即是透過來源彼此統計上獨立之特性而發展出來，故訊號來源必須滿足此一特性。

(二) 來源訊號最多只能允許一組服從高斯分配(Gaussian Distribution)：

根據中央極限定理(Central Limit Theorem, CLT)，獨立的隨機變數之加總會更趨近高斯分配；而 ICA 即應用對中央極限定理的反向思考：「基於資料的混合訊號是原來未知訊號來源的線性組合，那假設想找出原來互為統計獨立的訊號來源時，則可尋找不具高斯分配的成份，其為訊號來源的可能性最大」(Hyvärinen & Oja, 2000)。由於 $Y=AS$ ， Y 是由 S 線性組合而成，根據 CLT， Y 會比 S 更接近高斯分配，而 ICA 希望找出一組解混合矩陣 W ，使 $X=WY$ ，當 X 為非高斯特性時， X 則非 S 之線性組合，因此可得 $WA=I$ ，表示分離成功並得到獨立成份，即 $W=A^{-1}$ 。以上，藉由最大化 X 的非高斯特性可得到原本彼此獨立之隨機變數。另外，高斯分配具有可加性，即多個高斯分配之訊號相加之後仍為高斯分配。故若混合訊號是由高斯分配的訊號混合而成，則 W 不論如何尋找，其所計算出的結果依舊是為高斯分配，也就無法分離出原始訊號。因此，來源訊號最多只能允許一組來源訊號或潛在因素為高斯分配。

(三) 觀察到的混合訊號數目，須大於或等於潛在因素數目：

由式(3-11)的角度來看，若 $M=N$ ，則 $S=A^{-1}X$ ，方程式數目等於未知訊號數目，恰可找出 S 與 A ，故可分離出獨立的訊號來源。而當 $M > N$ ，此時方程式數目較未知訊號的數目還多，此情形下可先將維度降至 N 後，再執行 ICA；但若是 $M < N$ ，方程式數目少於未知訊號數目，此情況下缺乏足夠資訊，即無法找出獨立的訊號來源。因此，觀察到的混合訊號數目，須大於或等於潛在因素數目。

三、ICA 的基本模型

經由上述說明後，可知 $Y=AS$ 即為 ICA 的基本數學模型，且須滿足三種基本假設。亦即 ICA 為了要藉由訊號來源互相獨立的這一性質，從混合訊號 Y 當中分離出來源訊號 S ，必須發展有效衡量非高斯特性的準則以作為目標函數。簡言之，ICA 最主要可分為兩部份：

(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 。

而在過去學者個案研究中，ICA 的演算法有許多選擇。以下為三種主要的 ICA 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)

上述這些求取 ICs 的演算法都不盡相同，在 Hyvärinen (1999)學者提出 FastICA 後，由於其計算效率相對較高，而成為處理 ICA 問題常見的演算法之一；因此，本研究亦採用 FastICA 演算以找出相互獨立之成份。

四、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是無法預測、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 $H(y)$ 定義如式(3-14)：

$$H(y) = - \int f(y) \log f(y) dy \quad (3-14)$$

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 $J(y)$ 定義如式(3-15)：

$$J(y) = H(y_{gauss}) - H(y) \quad (3-15)$$

式中， y_{gauss} 為與 Y 具有相同變異數的高斯隨機向量。

然而，當欲求出熵時，必須知道樣本資料的統計分配。由於當前的研究方法欲確定樣本分配將會非常困難，故而 Hyvärinen 於 1998 年根據最大熵原理，發展出式(3-16)的近似函數：

$$J(y) \propto \{E[G(y_i)] - E[G(v)]\}^2 \quad (3-16)$$

其中， v 為期望值為 0、變異數為 1，且服從高斯分配之隨機變數；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 $J(y)$ 值將會等於 0。雖然，上式 $G(y_i)$ 可以假設為任意非二次方函數，但較常被使用的則是以下兩個函數：

$$G_1(y_i) = \frac{1}{\alpha} \log[\cos(\alpha y_i)], \text{ 其中 } 1 \leq \alpha \leq 2 \quad (3-17)$$

$$G_2(y_i) = -\exp\left(\frac{-y_i^2}{2}\right) \quad (3-18)$$

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，故而研究中將以 $G_1(y_i)$ 作為輸入 FastICA 的函數。

(二) FastICA 演算法流程：

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，為了簡化 ICA 運算過程並縮短其處理時間，Hyvärinen & 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：中間化(Centering)和白化(Whitening)。係先將原始混合矩陣 Y 轉換為不具相關的資料；但由於不具相關未必代表獨立，因此必須再設法求得解混合矩陣 W ，將不具相關的資料轉換成一組近似於獨立的獨立資料 X 。因此，FastICA 基本演算法的處理步驟如下：

步驟一：隨機選取一個權重向量 w ，且 $|w| = 1$ 。

步驟二： $w^+ = E[xg(w^T x)] - E[g'(w^T x)]w$ 。

步驟三：令 $w^* = w^+ / |w^+|$ 。

步驟四：若 w 收斂則輸出 w ；反之則重回步驟二。

此外，以上步驟需再注意以下兩點：

第一點：函數 g 為式(3-17)或(3-18)中 G 的導函數(Derivative Function)。

第二點：步驟四中所提之收斂，是指新的矩陣 w^* 與舊的解混矩陣 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 IC 排序與篩選過程

在找出 IC 之後，本研究利用簡單相關及偏相關係數來決定各個成份重要程度的排序，並進一步篩選出不具貢獻，即可視為雜訊之 IC；其過程如下：

(一) 利用簡單相關及偏相關係數決定 ICs 之排序，亦即 $\{IC_{(1)}, IC_{(2)}, \dots, IC_{(p)}\}$ ；其中， p 為選取過程的階段數。

現假設誤差序列為 ε ，扣除趨勢序列後的誤差序列 $(\varepsilon - R)$ 與各 ICs 間之簡單相關為：

$$\text{Corr}(\varepsilon - R, IC_J) = r_{(\varepsilon - R), IC_J}, J \in \{1, 2, \dots, p\} \quad (3-19)$$

於此，選取相關程度最高的 IC 並視其為最具貢獻之成份。現若假設 IC5 為最高相關者，則令 $IC_{(1)} = IC5$ 。

其次，將 $IC_{(1)}$ 設為控制變數，並計算序列 $(\varepsilon - R)$ 與剩餘之 ICs 間的偏相關係數；此結果等同於計算序列 $(\varepsilon - R - IC_{(1)})$ 與剩餘 ICs 之簡單相關，亦即：

$$r_{(\varepsilon - R), IC_J | IC_{(1)}} = r_{(\varepsilon - R - IC_{(1)}), IC_J}, J \in \{1, 2, \dots, p\}, \text{ 但 } J \neq 5 \quad (3-20)$$

於此，再找出此階段最高相關之 IC。現若假設為 IC3，則令 $IC_{(2)} = IC3$ 。

接著，以迭代方式重複以上過程，直至所有成份排序完成為止。

(二) 應用增量 R^2 過濾不具貢獻性之成份。

在決定出 ICs 的排序之後，首先，將 $IC_{(1)}$ 帶入下式中以求算出 $R_{(1)}^2$ ；其次，再將 $IC_{(1)}$ 、 $IC_{(2)}$ 帶入式中，則可得 $R_{(2)}^2$ ；依此類推，直到最後的 $R_{(p)}^2$ 求出為止。

$$(\varepsilon - R) = \hat{\beta}_0 + \hat{\beta}_1 \sum_{j=1}^J IC_{(j)}, J \in \{1, 2, \dots, p\} \quad (3-21)$$

現假設 $R_{(0)}^2 = 0$ ，則即可定義增量 $R^2(\Delta R_{(j)}^2)$ ：

$$\Delta R_{(j)}^2 = R_{(j)}^2 - R_{(j-1)}^2, J \in \{1, 2, \dots, p\} \quad (3-22)$$

由於在絕大多數情況下， $\Delta R_{(j)}^2$ 會隨 J 的增加而以指數型式逐漸下降；因此，研究中直接以 Quenouille (1948)所提之應用標準誤 $(1/\sqrt{n})$ 為標準以檢測 $\sqrt{\Delta R_{(j)}^2}$ 。亦即在顯著水準 0.05 下以 $2/\sqrt{n}$ 為標準，選取具明顯貢獻之 IC；當若加入一個 IC 後，會使其 $\sqrt{\Delta R_{(j)}^2}$ 低於標準時，則停止選取，並視其餘未選取之成份為不具貢獻性之 ICs。

第五節 衡量指標

本研究將採用幾種常見比較不同模型間適合度及預測能力的指標進行評估工作，諸如：解釋能力(R^2)、赤池訊息準則(Akaike Information Criterion, AIC)與貝氏訊息準則(Schwarz Bayesian information criterion, SBC)，以及均方根誤差(Root Mean Square Error, RMSE)、平均絕對誤差(Mean Absolute Error, MAE)與 Theil 不等係數(Theil's Inequality Coefficient, U)。

在模型適合度方面，當模型的 R^2 值愈高、且 AIC 與 SBC 越小時，表示其模型適合度越好。而預測能力方面，當 RMSE、MAE 及 U 指標越小時，則表示所建立的模型其預測能力越佳；表 3-1 即為各指標之詳細計算公式。

表 3-1 衡量指標種類

指標類別		公式
適合度	解釋能力(R^2)	$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
	赤池訊息準則(AIC)	$AIC = 2k + n \ln \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)$
	貝氏訊息準則(SBC)	$SBC = k \ln n + n \ln \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)$
預測能力	均方根誤差(RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
	平均絕對誤差(MAE)	$MAE = \frac{\sum_{i=1}^n y_i - \hat{y}_i }{n}$
	Theil 不等係數(U)	$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2}}$

註： y_i 為樣本的實際觀察值， \hat{y}_i 為樣本的模型預測值， \bar{y} 為樣本的平均值， n 為樣本的資料個數， k 為模型中參數個數。

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

本研究以台灣工業生產指數為研究主題，資料序列取自國家發展委員會(前經建會)每月公佈之景氣對策指標，資料蒐集的期間為 2002 年 1 月至 2015 年 12 月，合計 168 筆研究資料(如圖 4-1);研究中並依移動視窗法再將資料細分，主要分為訓練組及預測組，訓練組依年度不同分為三類，分別為：九個年度、八個年度及七個年度；預測組的部分則依年份分為 5 類：2011 年、2012 年、2013 年、2014 年、2015 年；合計為 15 組序列資料。詳細資料的分組如表 4-1 所示。

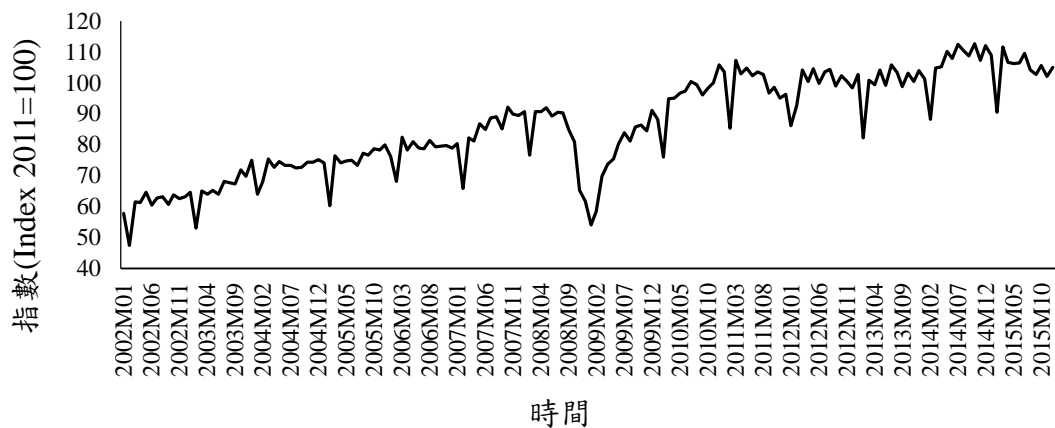


圖 4-1 台灣工業生產指數歷史走勢圖(2002/01~2015/12)

資料來源：國家發展委員會 (2016)。工業生產指數。台灣景氣指標月刊，國家發展委員會經濟發展處，40(1)，34。

表 4-1 資料分組表

組別		年份				
訓練組 (期間)	9 年	2002~2010	2003~2011	2004~2012	2005~2013	2006~2014
	8 年	2003~2010	2004~2011	2005~2012	2006~2013	2007~2014
	7 年	2004~2010	2005~2011	2006~2012	2007~2013	2008~2014
預測組		2011 年	2012 年	2013 年	2014 年	2015 年

資料來源：本研究整理。

第二節 個案傳統 ARIMA 模型建立與預測

一、單根檢定

由於本研究中，每組的樣本組之分析方法皆相同；因此，以下僅以一組資料作為過程的呈現案例，其餘樣本組則直接列出實證結果，並將詳細資料列於附錄；因此，以下取訓練組由 2003 年 1 月至 2010 年 12 月與測試組 2011 年之資料作為呈現案例。

首先，將蒐集到的 2003 年 1 月到 2010 年 12 月共 8 年 96 筆原資料(Y_t)，藉由 ADF 單根檢定的方式判斷序列是否具有單根，以決定序列資料是否具有不平穩性；若檢定結果序列具單根，則需將序列進行差分後再次進行單根檢定，直到序列不具單根為止，亦藉此判斷建構 ARIMA 模型時的差分階次(d)。

據表 4-2 之結果，原始序列 Y_t 經檢定具有單根，其一階差分後之序列再次經檢定(表 4-3)即不具單根，因此建構模型實屬於 I(1)序列。

表 4-2 原始序列之單根檢定表

擴張的 Dickey-Fuller 單根檢定							
類型	滯後	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
單一平均值	0	-13.8284	0.0459	-2.49	0.1204	3.35	0.2271
	1	-7.9337	0.2102	-1.90	0.3331	2.52	0.4366
	2	-7.1605	0.2546	-1.62	0.4705	1.87	0.5991
趨勢	0	-27.6777	0.0087	-3.83	0.0189	7.41	0.0257
	1	-16.4006	0.1210	-2.75	0.2180	3.80	0.4275
	2	-17.4752	0.0959	-2.61	0.2756	3.47	0.4915

資料來源：本研究整理。

表 4-3 原始序列經一階差分後之單根檢定表

擴張的 Dickey-Fuller 單根檢定							
類型	滯後	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
單一平均值	0	-127.874	0.0001	-14.26	<.0001	101.73	0.0010
	1	-133.252	0.0001	-8.04	<.0001	32.31	0.0010
	2	-116.951	0.0001	-5.83	<.0001	17.00	0.0010
趨勢	0	-127.890	0.0001	-14.18	<.0001	100.58	0.0010
	1	-133.309	0.0001	-8.00	<.0001	32.04	0.0010
	2	-116.970	0.0001	-5.80	<.0001	16.91	0.0010

資料來源：本研究整理。

二、暫定模型

接下來，本研究藉由檢視該序列經一階差分後之 ACF 及 PACF(如附錄中附圖 1 所示)，以判斷 AR(p)與 MA(q)之落後期數。由附圖 1 的結果可看出：ACF 於落後期數 1、12 期時較為凸出；而 PACF 於落後期數 1、11、12 期時較為凸出；藉此判斷 ARIMA 模

型的暫定模型；下一步再對此三種模型計算 Q 統計量以進行模型殘差項檢定，以判斷隨機誤差項是否滿足白噪音過程。透過 Q 統計量檢定結果，將其落後 6、12、18、24 期的 Q 統計量數整理於表 4-4 中。

表 4-4 原始序列之 ARIMA 暫定模型殘差檢定 Q 統計量表

可能模型		Q6 (P_Value)	Q12 (P_Value)	Q18 (P_Value)	Q24 (P_Value)
模型一	ARIMA[(1,12),1,0]	1.78 (0.7757)	8.54 (0.5763)	12.47 (0.7110)	22.90 (0.4076)
模型二	ARIMA(1,1,12)	1.10 (0.8938)	5.71 (0.8389)	10.54 (0.8370)	31.07 (0.0946)
模型三	ARIMA(12,1,1)	3.76 (0.4390)	12.02 (0.2836)	15.37 (0.4979)	25.43 (0.2771)

資料來源：本研究整理。

由表中訊息得知，研究中三種暫定模型之 Q 統計量皆大於顯著水準 0.05，表示接受三模型之隨機誤差項滿足白噪音過程。最後在建構可能模型之模型參數均達顯著水準 0.05 時，其模型的 R^2 、AIC、及 SBC 列於表 4-5。

表 4-5 原始序列之 ARIMA 暫定模型適合度檢測表

可能模型		R^2	AIC	SBC
模型一	ARIMA[(1,12),1,0]	0.78302	582.4733	587.5811
模型二	ARIMA(1,1,12)	0.75814	592.7877	597.8955
模型三	ARIMA(12,1,1)	0.78059	583.5337	588.6414

資料來源：本研究整理。

其中 R^2 越大表示模型適合度越佳；AIC 與 SBC 則越小適合度越佳越佳。由表中結果可知，模型一的三種指標皆優於其他兩模型，因此本研究選擇模型一來建立其傳統 ARIMA 模型。

三、模型確定

經由上述模型暫定、參數估計、檢定與修正三個步驟(Box, Jenkins, Reinsel, & Ljung, 2015)，以進行傳統 ARIMA 模型的建立；所建構之台灣工業生產指數最適模型的參數估計值(T 檢定量)與相關統計量如下所示：

$$\Delta^1 \hat{Y}_t = -0.2607 \Delta^1 Y_{t-1} + 0.49206 \Delta^1 Y_{t-12} \quad (4-1)$$

(-2.97*) (5.16*)

$$R^2 = 0.78302$$

$$AIC = 582.4733$$

$$Q(6) = 1.78$$

$$Q(18) = 12.47$$

$$SBC = 587.5811$$

$$Q(12) = 8.54$$

$$Q(24) = 22.90$$

註：1. *表示在顯著水準 0.05 下具顯著性。

2. $\Delta^1 Y_t$ 表示 Y_t 的一階差分，亦即 $\Delta^1 Y_t = Y_t - Y_{t-1}$ 。

第三節 個案 EMD-ARIMA 模型

本研究另參考 Lin (2015)所提之程序，對工業生產指數之原始序列進行 EMD 分解，由 IMFs 中選出雜訊予以過濾，並應用過濾後之序列資料配適 ARIMA 模型，再與本研究所提之程序結果比較。

一、EMD 程序

首先，將工業生產指數之序列資料，直接應用 EMD 法，分解出不同的 IMFs(圖 4-2)

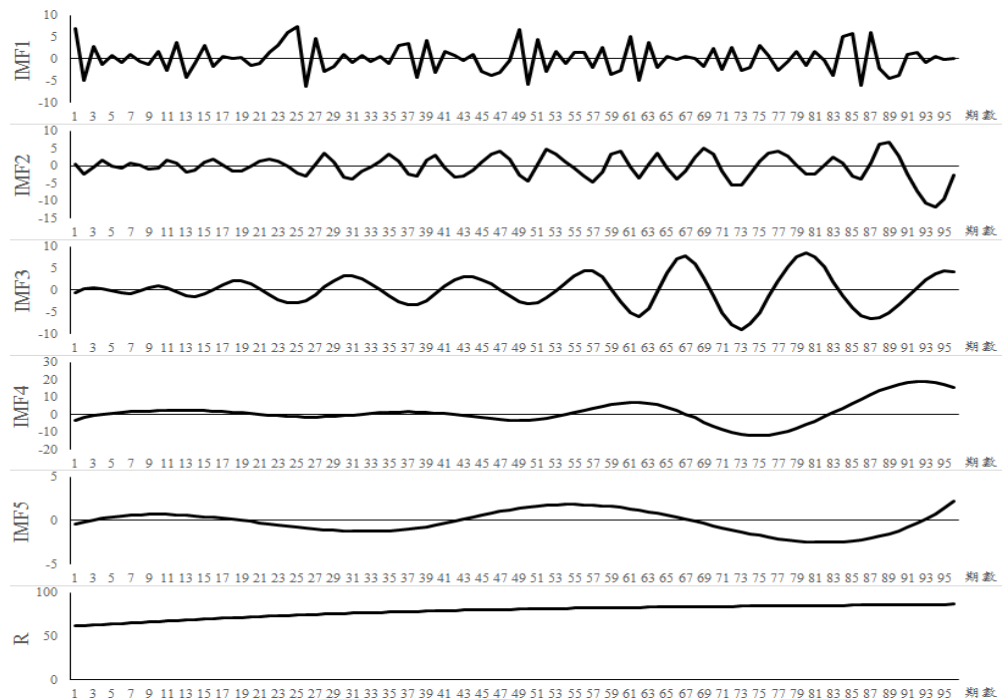


圖 4-2 原始資料之 IMFs 與趨勢分量 R

資料來源：本研究整理。

二、選擇具貢獻之 IMF 並過濾雜訊

據此，為能確認原始資料 Y_t 中的重要 IMF 以辨識出可能的資料雜訊，研究中依以下兩步驟過濾不具貢獻之 IMF：

步驟一 依重要程度排序：

先將趨勢序列 R 由原始資料 Y_t 中扣除後得到 $(Y_t - R)$ ，並計算與每一個 IMF 間之偏相關係數；其中，擁有最高相關係數者即被視為構成誤差序列之最重要的 IMF。同理，於確定排序第一的 IMF 在資料序列中予以扣除，並再次計算其與剩餘 IMFs 的偏相關係數。以此類推，直到將各個 IMFs 依重要程度排序完成為止，如表 4-6 中所示。

表 4-6 個別 IMF 與原始資料 Y_t 之偏相關係數表

階段	IMF4	IMF3	IMF2	IMF1	IMF5
1	0.7480	0.3958	0.0794	0.2926	0.3277
2	—	0.5938	0.4589	0.4832	0.2068
3	—	—	0.6717	0.6325	0.1778
4	—	—	—	0.9222	0.3442
5	—	—	—	—	1.0000

資料來源：本研究整理。

步驟二 過濾雜訊：

在經由前述判定過程後，研究中的重要獨立成份依序為 IMF4、IMF3、IMF2、IMF1、IMF5。雖說，從中確實可看出各個 IMF 的重要排序，但重點是到底要如何選擇最適之 IMF 個數，並據以將與原始資料最不具相關的 IMF 過濾之，而達到過濾雜訊之目的？選擇保留過多的 IMF，將可能會導致雜訊無法過濾乾淨；但選擇保留過少，則無法表現原始序列的結構。

因此，研究中計算出 R^2 、 ΔR^2 及 $\sqrt{\Delta R^2}$ (對於逐一加入個別 IMF 後之模型解釋能力的變化)，並應用 Quenouille (1948)所提之標準誤($1/\sqrt{n}$)為標準以檢測 $\sqrt{\Delta R^2}$ ，以決定選取的重要成份個數。一旦 $\sqrt{\Delta R^2}$ 低於選取標準($2/\sqrt{96}=0.204124$)時，即視未選取之其餘 IMF 為序列中最不具貢獻之成份，並將其還原而由原資料序列中予以過濾之，同時亦視此新的資料序列為已過濾雜訊後的時間序列資料。其結果如表 4-7 所示。

表 4-7 各階段解釋能力比較表

階段	獨立成份	R^2	ΔR^2	$\sqrt{\Delta R^2}$
1	IMF4	0.5595		
2	IMF4、IMF3	0.7175	0.1584	0.39749*
3	IMF4、IMF3、IMF2	0.8392	0.1217	0.34885*
4	IMF4、IMF3、IMF2、IMF1	0.9771	0.1379	0.37134*
5	IMF4、IMF3、IMF2、IMF1、IMF5	1.0000	0.0229	0.15132

註：*表示在顯著水準 0.05 下具顯著性。

資料來源：本研究整理。

由表 4-7 可知，當模型中已存在 IMF4、IMF3 及 IMF2 時，再增加 IMF1 所還原後的序列將與原始序列 Y_t 具顯著性貢獻；但若再增加 IMF5 時，則模型的解釋能力並沒有明顯地再增加。因此，研究中的工業生產指數個案將依此一結果，選取剩餘的序列部分 IMF5 為本研究所稱之真正雜訊，尚需由原始資料序列 Y_t 中予以過濾；同時，過濾後的資料序列(Y'_t)將做為下一階段預測模型配適之已過濾雜訊的時間序列資料。

三、EMD-ARIMA 模型建立與評估

研究中將已過濾雜訊的時間序列資料 Y'_t 同樣藉由配適 ARIMA 模型，以建立結合 EMD 與 ARIMA 的模型。圖 4-3 即為過濾雜訊後的序列資料 Y'_t 。

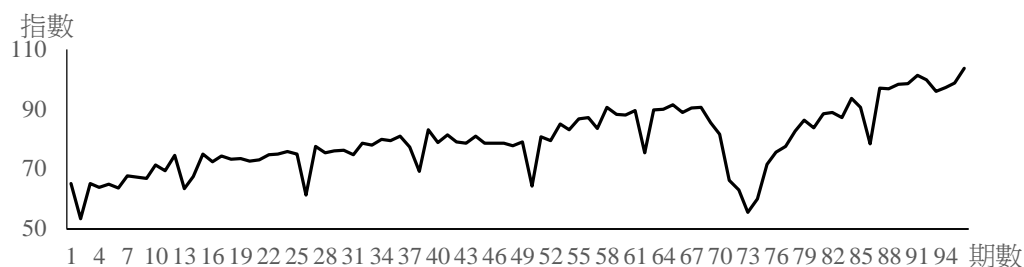


圖 4-3 過濾雜訊後之序列資料 Y'_t

資料來源：本研究整理。

(一) 單根檢定

首先，亦藉由 ADF 單根檢定的方式判斷序列是否具有單根，以決定序列資料是否具有不平穩性；據表 4-8 及表 4-9 結果顯示，過濾雜訊後之序列資料 Y'_t 仍須經一階差分，才可達到不具單根之結果，因此建構混合模型時仍屬於 I(1)序列。

表 4-8 Y'_t 之單根檢定表

擴張的 Dickey-Fuller 單根檢定							
類型	滯後	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
單一平均值	0	-14.3221	0.0402	-2.59	0.0989	3.57	0.1730
	1	-8.3136	0.1911	-2.01	0.2831	2.69	0.3945
	2	-7.4965	0.2343	-1.73	0.4154	2.02	0.5613
趨勢	0	-31.1255	0.0036	-4.16	0.0074	8.69	0.0018
	1	-18.8542	0.0709	-3.01	0.1364	4.52	0.2868
	2	-20.3917	0.0500	-2.87	0.1775	4.14	0.3603

資料來源：本研究整理。

表 4-9 Y_t' 經一階差分後之單根檢定表

擴張的 Dickey-Fuller 單根檢定							
類型	滯後	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
單一平均值	0	-128.371	0.0001	-14.38	<.0001	103.44	0.0010
	1	-135.975	0.0001	-8.14	<.0001	33.11	0.0010
	2	-122.129	0.0001	-5.93	<.0001	17.56	0.0010
趨勢	0	-128.370	0.0001	-14.30	<.0001	102.27	0.0010
	1	-135.974	0.0001	-8.09	<.0001	32.78	0.0010
	2	-122.057	0.0001	-5.89	<.0001	17.41	0.0010

資料來源：本研究整理。

(二) 暫定模型

同樣，本研究先藉由檢視該序列經一階差分後之 ACF 及 PACF(如附錄中附圖 2 所示)，以判斷 AR(p)與 MA(q)之落後期數。由附圖 2 的結果可看出：ACF 於落後期數 1、12 期時較為凸出；而 PACF 於落後期數 1、11、12 期時較為凸出；藉此判斷 ARIMA 模型的暫定模型。

接著再對此三種模型計算 Q 統計量以進行模型殘差項檢定，以判斷隨機誤差項是否滿足白噪音過程。透過 Q 統計量檢定結果，將其落後 6、12、18、及 24 期的 Q 統計量數整理於表 4-10 中。

表 4-10 EMD-ARIMA 暫定模型殘差檢定 Q 統計量表

可能模型		Q6 (P_Value)	Q12 (P_Value)	Q18 (P_Value)	Q24 (P_Value)
模型一	ARIMA[(11,12),1,0]	14.41* (0.0061)	26.33* (0.0033)	29.17* (0.0228)	41.32* (0.0076)
模型二	ARIMA(12,1,0)	14.03* (0.0154)	25.40* (0.0080)	27.99* (0.0450)	40.61* (0.0131)
模型三	ARIMA(12,1,1)	3.66 (0.4533)	12.78 (0.2361)	16.18 (0.4402)	28.23 (0.1682)

註：*表示在顯著水準 0.05 下具顯著性。

資料來源：本研究整理。

由表中訊息得知，研究中三種暫定模型之 Q 統計量中只有模型三之 P_Value 皆大於顯著水準 0.05，表示此模型接受隨機誤差項滿足白噪音過程。而在建構可能模型之模型參數均達顯著水準 0.05 時，其模型的 R^2 、AIC、及 SBC 列於表 4-11。

表 4-11 EMD-ARIMA 暫定模型適合度檢測表

	可能模型	R ²	AIC	SBC
模型一	ARIMA[(11,12),1,0]	0.82589	553.5984	561.2283
模型二	ARIMA(12,1,0)	0.80929	560.1602	565.2468
模型三	ARIMA(12,1,1)	0.83689	543.463	546.0063

資料來源：本研究整理。

其中 R² 越大表示模型適合度越佳；AIC 與 SBC 則是越小模型適合度越佳。由表中結果可知，模型三的三種指標皆優於其他兩模型。

最後本研究選擇模型三來建立其 EMD-ARIMA 模型。

(三) 確定模型

最後，在經過模型暫定、參數估計、檢定與修正三個步驟後，所建構之台灣工業生產指數最適 EMD-ARIMA 模型的參數估計值(T 檢定量)與相關統計量如下所示：

$$\Delta^1 \hat{Y}_t^c = 0.51501 \Delta^1 Y'_{t-12} + 0.25689 \varepsilon_{t-1} \quad (4-2)$$

(5.18*) (2.55*)

$$R^2 = 0.77930$$

$$AIC = 582.7189$$

$$SBC = 587.8267$$

$$Q(6) = 3.66$$

$$Q(12) = 12.78$$

$$Q(18) = 16.18$$

$$Q(24) = 28.23$$

註：*表示在顯著水準 0.05 下具顯著性。

第四節 個案 EMD-ICA-ARIMA 模型

一、EMD 程序

在此，應用前述之最適傳統 ARIMA 模型進行歷史資料的預測，並據以求得該模型的預測誤差序列 ε 。之後，再針對誤差序列 ε 進行 EMD 分解成 5 個 IMF 分量以及一個趨勢分量 R；詳細結果如圖 4-4 中所示。

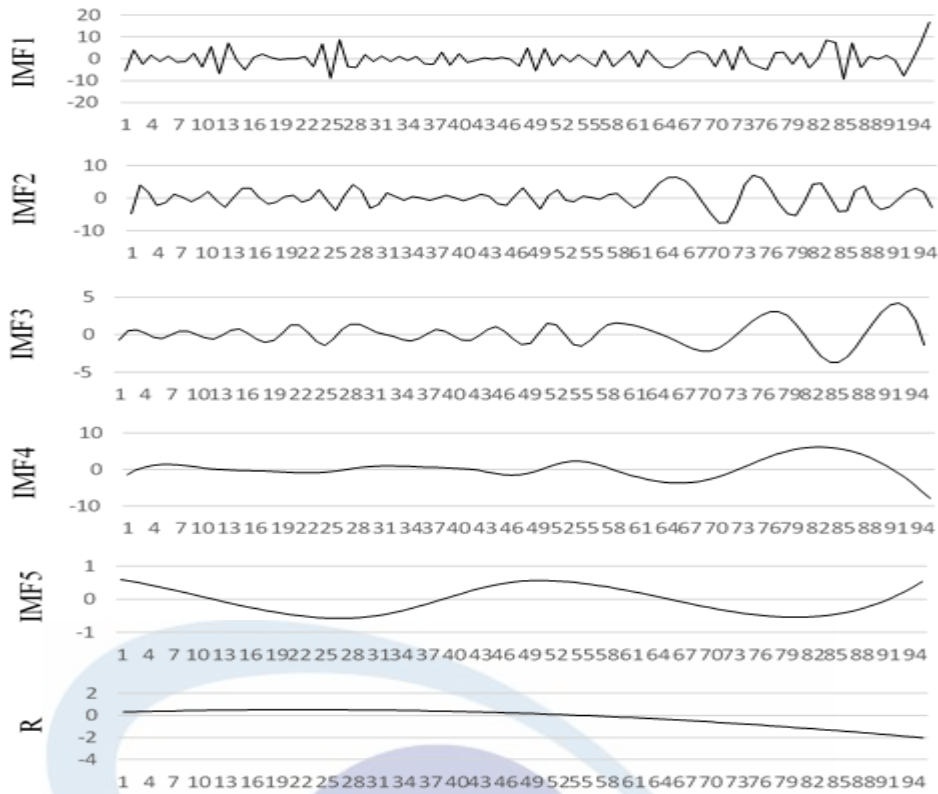


圖 4-4 誤差序列之 IMFs 與趨勢分量 R 序列圖

資料來源：本研究整理。

二、ICA 過程

(一) 分解出 ICs

本研究認為，經由 EMD 所分解的 IMF 彼此之間仍然存在著某種程度上的相關性，再加上每一個 IMF 對時間序列皆隱含著某些程度的解釋能力；若為過濾資料中的雜訊而冒然直接移除一個或數個 IMF，則將有可能會喪失許多重要的資訊。因此，研究中再將此 5 個 IMF 應用 ICA 依以下流程轉換成 5 個獨立的 IC。

現令 $IMF_1(t)$ 、 $IMF_2(t)$ 、 $IMF_3(t)$ 、 $IMF_4(t)$ 、 $IMF_5(t)$ 分別為來源訊號 $s_1(t)$ 、 $s_2(t)$ 、 $s_3(t)$ 、 $s_4(t)$ 、 $s_5(t)$ 經線性組合而成，其中 t 為 1~96。以數學式表示為式(4-3)：

$$\begin{aligned}
 IMF_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) + a_{14}s_4(t) + a_{15}s_5(t) \\
 IMF_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) + a_{24}s_4(t) + a_{25}s_5(t) \\
 IMF_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) + a_{34}s_4(t) + a_{35}s_5(t) \\
 IMF_4(t) &= a_{41}s_1(t) + a_{42}s_2(t) + a_{43}s_3(t) + a_{44}s_4(t) + a_{45}s_5(t) \\
 IMF_5(t) &= a_{51}s_1(t) + a_{52}s_2(t) + a_{53}s_3(t) + a_{54}s_4(t) + a_{55}s_5(t)
 \end{aligned} \tag{4-3}$$

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：

$$\text{IMF} = \begin{bmatrix} \text{IMF}_1(t) \\ \text{IMF}_2(t) \\ \vdots \\ \text{IMF}_5(t) \end{bmatrix}_{5 \times 1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{15} \\ a_{21} & a_{22} & \cdots & a_{25} \\ \vdots & \vdots & \ddots & \vdots \\ a_{51} & a_{52} & \cdots & a_{55} \end{bmatrix}_{5 \times 5} \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_5(t) \end{bmatrix}_{5 \times 1} = AS, t = 1 \sim 96 \quad (4-4)$$

在分析過程中，若來源矩陣 S 未知，而混合矩陣 A 已知，則可透過線性方程組的方式解出 S ，即 $S = A^{-1} \text{IMF}$ ；然而，本研究探討的問題就在於矩陣 A 及 S 皆為未知，因此在面臨這樣的問題時，研究中將應用 FastICA 演算法之分析方法估計出矩陣 A' ，用以近似混合矩陣 A ，再應用 A' 之反矩陣求出解混合矩陣 W ，最後將混合序列還原為近似於最初狀態的 IC 矩陣： $W \cdot \text{IMF}$ ，進而用以估計未知潛在來源訊號 S ；本研究所計算之解混合矩陣 W 表示如下：

$$W = \begin{bmatrix} 0.1044 & 0.0338 & -0.4077 & 0.2251 & -0.1994 \\ 0.1688 & -0.0501 & -0.0423 & -0.2283 & 0.2029 \\ 0.0218 & -0.0688 & 0.1469 & -0.1459 & -2.6819 \\ -0.1289 & -0.2360 & -0.3574 & -0.1295 & -0.0351 \\ -0.0611 & -0.2430 & -0.3906 & -0.2020 & -0.6450 \end{bmatrix} \quad (4-5)$$

在計算出 W 矩陣之後，為了將混合序列還原回近似最初狀態的 IC 矩陣，將 IMF 矩陣乘上解混合矩陣 W 後，可得到估計原始來源訊號的 IC 矩陣，即 $\text{IC} = \text{IMF} \cdot W$ 。其中，5 個獨立成份 IC，如圖 4-5 所示：

經由 ICA 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序列(各個 IC)間彼此擁有最大差異，這樣一來每個 IC 各自不具相關且彼此獨立，提供接下來研究更進一步的分析，以求得最能夠解釋原始資料的成份。

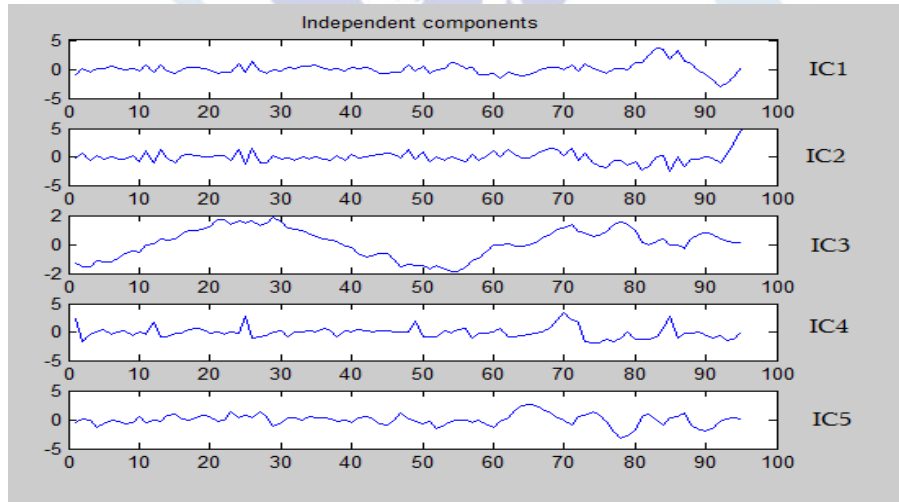


圖 4-5 誤差序列的 5 個 IC

資料來源：本研究整理。

(二) 選擇具貢獻之 IC 並過濾雜訊

據此，為能確認誤差序列 ε 中的重要 IC 以辨識出可能的資料雜訊，研究中依以下兩步驟過濾不具貢獻之 IC：

步驟一 依重要程度排序：

先將趨勢序列 R 由誤差序列 ε 中扣除後得到 $(\varepsilon - R)$ ，並計算與每一個 IC 間之偏相關係數；其中，擁有最高相關係數者即被視為構成誤差序列之最重要的 IC。同理，於確定排序第一的 IC 還原後在誤差序列中予以扣除，並再次計算其與剩餘 ICs 的偏相關係數。以此類推，直到將各個 ICs 依重要程度排序完成為止，如表 4-12 中所示。

表 4-12 個別 IC 與誤差序列之偏相關係數表

階段	IC4	IC1	IC2	IC5	IC3
1	-0.8970	0.3860	0.1778	-0.1115	-0.0484
2	—	0.8733	0.3022	-0.2522	-0.1094
3	—	—	0.8255	-0.5177	-0.2247
4	—	—	—	-0.9173	-0.3980
5	—	—	—	—	-1.0000

資料來源：本研究整理。

步驟二 過濾雜訊：

在經由前述判定過程後，研究中的重要獨立成份依序為 IC4、IC1、IC2、IC5、IC3。雖說，從中確實可看出各個 ICs 的重要排序，但重點是到底要如何選擇最適之 IC 個數，並據以將與原資料序列最不具相關的 IC 過濾之，而達到過濾雜訊之目的？選擇保留過多的獨立成份，將可能會導致雜訊無法過濾乾淨；但選擇保留過少的獨立成份，則無法表現原序列的結構。

因此，研究中計算出 R^2 、 ΔR^2 及 $\sqrt{\Delta R^2}$ （對於逐一加入個別 IC 後之模型解釋能力的變化），並應用標準誤 $(1/\sqrt{n})$ 為標準以檢測 $\sqrt{\Delta R^2}$ ，以決定選取的重要成份個數。一旦 $\sqrt{\Delta R^2}$ 低於選取標準 $(2/\sqrt{95} = 0.205196)$ 時，即視未選取之其餘 ICs 為序列中最不具貢獻之成份，並將其還原而由原資料序列中予以過濾之，同時亦視此新的資料序列為已過濾雜訊後的時間序列資料。其結果如表 4-13 所示。

表 4-13 各階段解釋能力比較表

階段	獨立成份	R^2	ΔR^2	$\sqrt{\Delta R^2}$
1	IC4	0.8046	-	-
2	IC4、IC1	0.9536	0.1490	0.3860*
3	IC4、IC1、IC2	0.9852	0.0316	0.1777
4	IC4、IC1、IC2、IC5	0.9931	0.0079	0.0888
5	IC4、IC1、IC2、IC5、IC3	1.0000	0.0069	0.0830

註：*表示在顯著水準 0.05 下具顯著性。

資料來源：本研究整理。

由表 4-13 可知，當模型中已存在 IC4 時，再增加 IC1 所還原後的序列將與原始序列 Y_t 具顯著性貢獻；但若再增加 IC2 時，則模型的解釋能力並沒有明顯地再增加。因

此，研究中的工業生產指數個案將依此一結果，選取 IC4 及 IC1 兩獨立成份所還原後的序列並加上趨勢分量 R，先將其由誤差序列中予以扣除之，而剩餘的序列部分即為本研究所稱之真正雜訊，尚需由原始資料序列 Y_t 中予以過濾；同時，過濾後的資料序列(Y_t^c)將做為下一階段預測模型配適之已過濾雜訊的時間序列資料。

三、EMD-ICA-ARIMA 模型建立

依此，研究中將已過濾雜訊的時間序列資料 Y_t^c 同樣藉由配適 ARIMA 模型，以建立一個結合 EMD、ICA 與 ARIMA 的混合模型。圖 4-6 即為過濾雜訊後的序列資料 Y_t^c ，

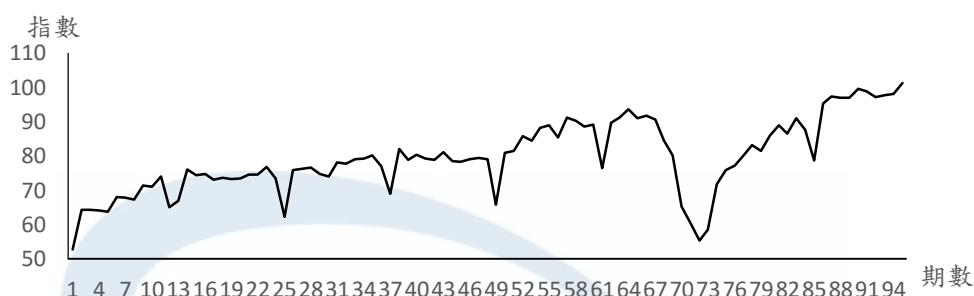


圖 4-6 過濾雜訊後之序列資料圖

資料來源：本研究整理。

(一) 單根檢定

首先，亦藉由 ADF 單根檢定的方式判斷序列是否具有單根，以決定序列資料是否具有不平穩性；據表 4-14 及表 4-15 結果顯示，過濾雜訊後之序列資料 Y_t^c 仍須經一階差分，才可達到不具單根之結果，因此建構混合模型時仍屬於 I(1)序列。

表 4-14 Y_t^c 之單根檢定表

擴張的 Dickey-Fuller 單根檢定							
類型	滯後	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
單一平均值	0	-13.9848	0.0439	-2.85	0.0560	4.50	0.0624
	1	-9.0138	0.1601	-2.00	0.2848	2.41	0.4631
	2	-8.8538	0.1666	-1.90	0.3323	2.21	0.5138
趨勢	0	-24.7038	0.0181	-3.80	0.0205	7.25	0.0300
	1	-19.0195	0.0682	-2.97	0.1454	4.43	0.3031
	2	-20.6380	0.0470	-2.90	0.1668	4.23	0.3422

資料來源：本研究整理。

表 4-15 Y_t^c 經一階差分後之單根檢定表

擴張的 Dickey-Fuller 單根檢定							
類型	滯後	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
單一平均值	0	-110.611	0.0001	-11.82	<.0001	69.81	0.0010
	1	-121.810	0.0001	-7.77	<.0001	30.16	0.0010
	2	-127.055	0.0001	-6.02	<.0001	18.11	0.0010
趨勢	0	-110.573	0.0001	-11.74	<.0001	69.06	0.0010
	1	-121.709	0.0001	-7.72	<.0001	29.85	0.0010
	2	-126.932	0.0001	-5.98	<.0001	17.93	0.0010

資料來源：本研究整理。

(二) 暫定模型

同樣，本研究先選出三種暫定模型計算 Q 統計量以進行模型殘差項檢定，以判斷隨機誤差項是否滿足白噪音過程。透過 Q 統計量檢定結果，將其落後 6、12、18、及 24 期的 Q 統計量數整理於表 4-16 中。

表 4-16 EMD-ICA-ARIMA 暫定模型殘差檢定 Q 統計量表

可能模型		Q6 (P_Value)	Q12 (P_Value)	Q18 (P_Value)	Q24 (P_Value)
模型一	ARIMA[6,1,(6,12)]	1.10 (0.7765)	8.98 (0.4393)	18.63 (0.2312)	24.97 (0.2485)
模型二	ARIMA(11,1,12)	2.04 (0.7287)	6.62 (0.7611)	11.39 (0.7847)	30.93 (0.0975)
模型三	ARIMA(12,1,0)	1.82 (0.8741)	11.95 (0.3672)	15.45 (0.5631)	20.65 (0.6024)

資料來源：本研究整理。

由表中訊息得知，研究中三種暫定模型之 Q 統計量皆大於顯著水準 0.05，表示接受三模型之隨機誤差項滿足白噪音過程。

接下來，藉由檢視該序列經一階差分後之 ACF 及 PACF(如附錄中附圖 3 所示)，以判斷 AR(p)與 MA(q)之落後期數。由附圖 3 的結果可看出：ACF 於落後期數 1、6、11、12 期時較為凸出；而 PACF 於落後期數 1、6、11、12 期時較為凸出；藉此判斷 ARIMA 模型的暫定模型。在建構可能模型之模型參數均達顯著水準 0.05 時，其模型的 R^2 、AIC、及 SBC 列於表 4-17。

表 4-17 EMD-ICA-ARIMA 暫定模型適合度檢測表

	可能模型	R ²	AIC	SBC
模型一	ARIMA[6,1,(6,12)]	0.82589	553.5984	561.2283
模型二	ARIMA(11,1,12)	0.80929	560.1602	565.2468
模型三	ARIMA(12,1,0)	0.83689	543.463	546.0063

資料來源：本研究整理。

其中 R² 越大表示模型適合度越佳；AIC 與 SBC 則是越小模型適合度越佳。由表中結果可知，模型三的一種指標皆優於其他兩模型。

最後本研究選擇模型三來建立其混合 ARIMA 模型。

(三) 模型確定

最後，在經過模型暫定、參數估計、檢定與修正三個步驟後，所建構之台灣工業生產指數最適混合 ARIMA 模型的參數估計值(T 檢定量)與相關統計量如下所示：

$$\Delta^1 \hat{Y}_t^c = 0.66543 \Delta^1 Y_{t-12}^c \quad (4-6)$$

(7.57*)

$$R^2 = 0.83689$$

$$AIC = 543.4633$$

$$SBC = 546.0063$$

$$Q(6) = 1.82$$

$$Q(12) = 11.95$$

$$Q(18) = 15.45$$

$$Q(24) = 20.65$$

註：*表示在顯著水準 0.05 下具顯著性。

第五節 個案模型比較

而由 ARIMA 的傳統、EMD 與 EMD-ICA 三個最適模型的適合度可知：透過研究中所述之結合 EMD-ICA 以過濾序列資料雜訊的混合模型，在模型的解釋能力(R²)上之表現雖與 EMD-ARIMA 相差不大，但其確實要較直接應用原始序列(Y_t)建立的 ARIMA 模型改善。其次，在對未來預測能力指標的表現上，研究中經由以 2011 年 1 月到 12 月的工業生產指數月資料序列作為測試樣本的預測，其結果如圖 4-7 與表 4-18 中所示。由表中的指標可知：無論是 RMSE、MAE 或 U 等統計量值，混合模型的預測值皆具相對較小的誤差；此亦說明此個案之 EMD-ICA-ARIMA 模型的績效皆要較採行原始序列的傳統 ARIMA 模型相對為佳。

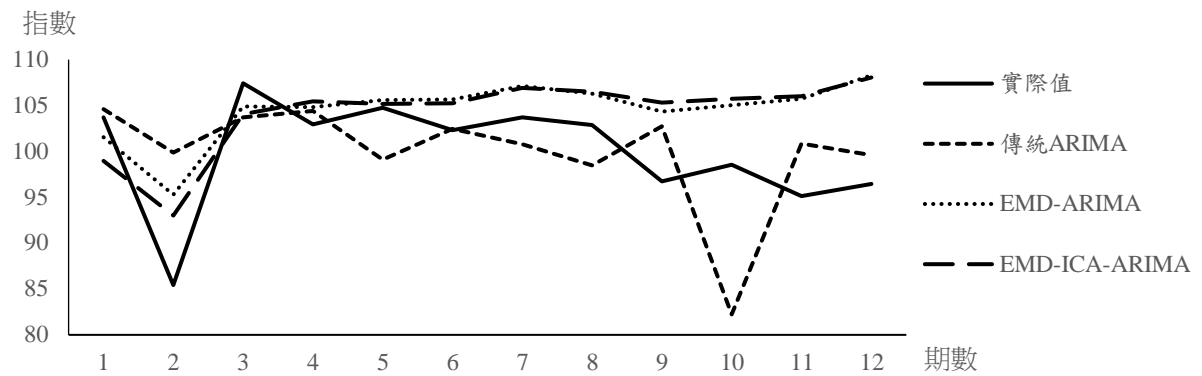


圖 4-7 傳統與混合模型的預測序列圖

資料來源：本研究整理。

表 4-18 三模型之預測能力比較表

模型	RMSE	MAE	U
傳統模型	9.0750	7.5530	0.04371
EMD-ICA-ARIMA 模型	6.4947	5.5440	0.03176
EMD-ARIMA 模型	6.4506	5.3289	0.03153

資料來源：本研究整理。

此外，由於每組個案分析過程皆相同，因此以上僅列出一筆個案作為呈現案例，其餘個案結果列於附錄(附表 1 與附表 2)中。

第五章 結論與建議

在一般實案研究中，其資料型態多屬於量測的資料；此種類型的資料無論是應用於傳統的時間序列預測模型或是電腦學習的方法，同樣需面對資料呈現不良訊號時(高頻、雜訊、非平穩)，經常無法完整準確地表達出資料本身的變化與趨勢，以致在進行預測時，往往會喪失其原本性質，導致其於建立預測模型或學習方法過程中產生過度配適或配適不足的窘境，因而降低其結果的可信度。

於此，本研究提出給合 EMD 與 ICA 的概念，將原時間序列資料經由 ARIMA 模型配適後所產生的預測誤差序列，透過 EMD 分解成為若干個 IMFs 之後，再經由 ICA 將其 IC 分解出來。其次，研究中亦提出以偏相關係數的概念，排序與原始誤差序列重要的 IC，再應用增量 R^2 的概念，將具明顯相關的重要成份保留住，同時將剩餘的成份還原後再由原時間序列資料中過濾之，進而成為本研究之已過濾雜訊的時間序列資料，亦做為下一階段模型建立與預測的序列資料。

第一節 結論

由研究中以台灣工業生產指數的月時間序列資料為個案，採行移動視窗法進行分析與評估，並將傳統 ARIMA、EMD-ARIMA 以及 EMD-ICA-ARIMA 三種模型互相比較後，綜合 15 組的結果(附表 1 與附表 2)可以發現：

- 一、研究中透過移動視窗法，將訓練組分為 9 年、8 年及 7 年。其中，訓練組之期間越短，對於所建立模型的解釋能力將越低；但訓練組期間的長短，對於模型的預測能力則較無明顯的差異性。
- 二、在 EMD-ARIMA 方面，有少數個訓練組無法挑選出雜訊，表示該組之每個 IMF 對於模型的解釋能力皆可被接受，也就是每個 IMF 皆含有重要資訊在內，或是雜訊已被分散至各 IMF 中，故而沒有任何一個 IMF 被判斷為單純的雜訊。
- 三、研究中提出在 EMD 後加入 ICA，目的就是使各 IMFs 可轉換為互不相關之 ICs，此結合 EMD 與 ICA 方式，並以偏相關及增量 R^2 概念，所得到之已過濾雜訊的資料序列，其於配適 ARIMA 模型上，除了在解釋能力的改善效果上較為不一致外，於 AIC 及 SBC 兩者統計量上，皆要較直接應用原資料序列建構的 ARIMA 模型為佳。
- 四、在預測能力的績效表現上，若除去找不到雜訊的個案，本研究提出之模型雖與 EMD-ARIMA 模型的結果相差不遠，但與傳統 ARIMA 模型比較時，則顯示出所有的指標皆相對要較為出色。

由以上四點可知，經由研究中所提之過濾原資料序列雜訊的方式，所建構之 EMD-ICA-ARIMA 模型的流程，在實務上確具適用性與可行性。

第二節 建議

在實務上，預測分析的方法仍有非常多，對於每個不同的領域，各自會有相對應最好的方法。本研究提出的程序，於經濟領域之中，仍有待改善之處；於此，本研究提出以下幾點建議：

- 一、由於 EMD-ARIMA 之程序會有數個年度找不出雜訊，可能是在 EMD 過程中，雜訊已被分散至各 IMFs 中，建議後續研究者可再透過其他方法，試圖將雜訊集中後再進一步分析，期能將雜訊過濾完全。
- 二、本研究採用台灣工業生產指數進行個案分析，所被關心之重點或許僅在於未來指數的走向；然實務上仍可將研究中提出之程序套用於不同領域之序列資料中，以期嘗試出更進一步的效果。
- 三、在經濟相關領域中，如前所述，預測方法亦非常多種；且隨著技術發展使趨於成熟與各種套裝軟體的開發、精進，各種預測技術皆持續進步、更新。而本研究中的預測模型僅採用相對較常見的 ARIMA 模型，建議後續研究者仍可將此概念應用於不同預測方法中，甚至可採行，應用多種模型的組合預測方式，使其結果能更具客觀性與穩定性。



參考文獻

一、中文部份

- 王僑宏 (2013)。應用希爾伯特-黃轉換法與極速學習機於逕流量及颱風降雨之預測研究。碩士論文，淡江大學水資源及環境工程學研究所。
- 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。
- 何川、舒勤、賀含峰 (2014)。ICA特徵提取與BP神經網路在負荷預測中的應用。電力系統及自動化學報，26(8)。
- 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。
- 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。碩士論文，國立台北科技大學經營管理系。
- 林淑華 (2010)。總體經濟因子與台美匯率關係性之探討。碩士論文，樹德科技大學金融與風險管理系。
- 徐志宏 (2013)。台灣景氣對策信號編製之檢討。經濟研究，13，1-42。
- 國家發展委員會 (2016)。工業生產指數。台灣景氣指標月刊，國家發展委員會經濟發展處，40(1)，34。
- 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。
- 黃月盈 (2013)。台灣景氣同時指標之檢討與修正。經濟研究，13，43-78。
- 黃宜正、林俊賢 (2013)。應用經驗模態與希爾伯特黃轉換探討耳脈搏訊號與心電圖訊息之比較。技術學刊，28(1)，1-7。
- 楊文熙 (2003)。股票變化之穩健預測。碩士論文，國立中央大學統計研究所。
- 經濟部 (2015)。工業生產統計年報。台北：經濟部。
- 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。
- 蘇恆信 (2012)。結合經驗模組分解與獨立成分分析於時間序列之預測—以台灣工業生產指數為例。碩士論文，國立宜蘭大學應用經濟與管理學系。
- 饒秀華、林修葳、黎明淵 (2001)。藉由分期MS模型分析台灣經濟景氣狀態。經濟論文，29(3)，297-319。

二、英文部份

- 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*, 8(4), 473-484.
- Bell, A. J., & Sejnowski, T. J. (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129-1159.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*, 4th ed. NY: Wiley.
- Cardoso, J. F., & Soudoumiac, A. (1993). Blind beamforming for non-gaussian signals. *IEEE Signal Processing Society*, 140(6), 771-774.
- 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.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H., Zheng, 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. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903-995.
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3), 626-634.
- Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5), 411-430.
- Kamath, V., Lai, Y. C., Zhu, L., & Urval, S. (2006). Empirical mode decomposition and blind source separation methods for antijamming with GPS signals. *IEEE/ION Position Location and Navigation Symposium*, 1, 335-341.
- Karhunen, J., Hyvärinen, A., Vigario, R., Hurri, J., & Oja, E. (1997). Applications of neural blind separation to signal and image processing. *Acoustics, Speech, and Signal Processing*, 1, 131-134.
- Kiviluto, K., & Oja, E. (1998). Independent component analysis for parallel financial time series. *International Conference on Neural Information Processing*, 2, 895-898.
- Lee, T. W. (1998). *Independent component analysis: Theory and applications*. Boston, Mass: Kluwer Academic Publishers.
- Lin, S. L., Tung, P. C., & Huang, N. E. (2012). Application of ICA-EEMD to secure communications in chaotic system. *International Journal of Modern Physics C*, 23(4), 1250028-1-1250028-11.
- Lin, F. J. (2015). Adding EMD Process and Filtering Analysis to Enhance Performances of

- ARIMA Model When Time Series Is Measurement Data. *The Romanian Journal of Economic Forecasting*, 18(2), 92-104.
- Moody, J. E., & Wu, L. (1997). What is the “True Price”? State space models for high frequency FX data. *Computational Intelligence for Financial Engineering*, 697-704.
- Quenouille, M. H. (1948). Some results in the testing of serial correlation coefficients. *Biometrika*, 35(3-4), 261-367.
- 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.
- Wang, W., Zhao, H., Li, Q., & Liu, Z. (2009). A novel hybrid intelligent model for financial time series forecasting and its application. *Business Intelligence and Financial Engineering*, 279-282.
- Weng, B., Blanco-Velasco, M. and Barner, K. E. (2006). ECG denoising based on the empirical mode decomposition. *Conference Proceedings IEEE Engineering in Medicine and Biology Society*, 1, 1-4.
- Wu, E. H., Yu, P. L., & 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.
- Yang, H. H., & Amari, S. I. (1997). Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information. *Neural Computation*, 9(7), 1457-1489.
- Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30(5), 2623-2635.

附錄

自相關														偏自相關													
序位	共變異數	相關	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1				
0	37.725264	1.00000											*****		0												
1	-13.496322	-.35775											*****	.													
2	3.433068	0.09100											**	.													
3	0.279044	0.00740												.													
4	-3.063311	-.08120												**	.												
5	1.916584	0.05080												.	*	.											
6	-6.085639	-.16131												.	***	.											
7	3.991079	0.10579												.	**	.											
8	-4.917071	-.13034												.	***	.											
9	2.675254	0.07091												.	*	.											
10	-1.204825	-.03194												.	*	.											
11	-7.336651	-.19448												.	****	.											
12	17.384100	0.46081												.	*****	.											
13	-7.340408	-.19458												.	****	.											
14	-0.308899	-.00819												.		.											
15	0.263052	0.00697												.		.											

1	-.35775												*****	.										
2	-.04241													.	*									
3	0.83005													.	*	.								
4	-.07458													.	*									
5	-.00675													.		.								
6	-.016265													.	***	.								
7	-.00357													.		.								
8	-.010828													.	**	.								
9	-.00709													.		.								
10	-.03662													.	*									
11	-.025119													.	*****	.								
12	0.34714													.	*****	.								
13	0.10855													.	**	.								
14	-.013539													.	***	.								
15	-.06641													.	*									

附圖 1 2003 至 2010 年，原始資料之 ACF 與 PACF
資料來源：本研究整理。

自相關														偏自相關													
序位	共變異數	相關	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1				
0	37.557982	1.00000																									
1	-13.658925	-0.36368																									
2	3.311555	0.08817																									
3	0.166170	0.00442																									
4	-3.203991	-0.08531																									
5	1.755741	0.04675																									
6	-6.217183	-0.16554																									
7	3.857717	0.10271																									
8	-5.041404	-0.13423																									
9	2.561021	0.06819																									
10	-1.172100	-0.03121																									
11	-7.376364	-0.19640																									
12	17.292149	0.46041																									
13	-7.370599	-0.19625																									
14	-0.329215	-0.00877																									
15	0.243096	0.00647																									

附圖 2 2003 至 2010 年，EMD 過濾雜訊後序列資料之 ACF 與 PACF
資料來源：本研究整理。

自相關														偏自相關													
序位	共變異數	相關	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1				
0	29.766331	1.00000												0													
1	-5.619800	-0.18880												0.103142													
2	-0.340506	-0.01144												0.106755													
3	-0.092262	-0.00310												0.106768													
4	-1.999374	-0.06717												0.106769													
5	1.600929	0.05378												0.107218													
6	-5.652049	-0.18988												0.107504													
7	2.378234	0.07990												0.111015													
8	-2.506373	-0.08420												0.111625													
9	0.860173	0.02890												0.112299													
10	-1.598266	-0.05369												0.112378													
11	-7.071418	-0.23756												0.112650													
12	17.037231	0.57237												0.117860													
13	-5.061902	-0.17005												0.144434													
14	-1.420704	-0.04773												0.146548													
15	-1.236277	-0.04153												0.146714													

附圖 3 2003 至 2010 年，EMD-ICA 過濾雜訊後序列資料之 ACF 與 PACF
資料來源：本研究整理。

附表 1 各年份三種模型之適合度比較

期 間	年份		模型	R ²	AIC	SBC
	訓練組	預測組				
9 年	2002~2010 年	2011 年	EMD-ICA-ARIMA	0.82081	643.5052	648.8321
			EMD-ARIMA	0.83463	646.2897	651.6354
			傳統 ARIMA	0.82922	648.3705	653.7162
	2003~2011 年	2012 年	EMD-ICA-ARIMA	0.84893	638.5573	641.2207
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.83361	653.4967	658.8424
	2004~2012 年	2013 年	EMD-ICA-ARIMA	0.81686	651.9925	654.656
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.81814	661.3682	666.7139
	2005~2013 年	2014 年	EMD-ICA-ARIMA	0.79451	668.3674	673.6943
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.77794	683.3672	688.7128
	2006~2014 年	2015 年	EMD-ICA-ARIMA	0.80409	667.9673	673.2942
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.81022	671.9563	677.3019
8 年	2003~2010 年	2011 年	EMD-ICA-ARIMA	0.83689	543.4632	546.0063
			EMD-ARIMA	0.77931	582.7189	587.8267
			傳統 ARIMA	0.78302	582.4733	587.5811
	2004~2011 年	2012 年	EMD-ICA-ARIMA	0.82512	570.1601	575.2467
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.80317	587.6096	595.2712
	2005~2012 年	2013 年	EMD-ICA-ARIMA	0.79774	587.8378	590.3811
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.79384	596.2755	601.3832
	2006~2013 年	2014 年	EMD-ICA-ARIMA	0.79224	586.5127	589.056
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.77226	603.6188	608.7266
	2007~2014 年	2015 年	EMD-ICA-ARIMA	0.77411	603.5572	608.6438
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.77547	609.0529	614.1606
7 年	2004~2010 年	2011 年	EMD-ICA-ARIMA	0.73367	511.0288	518.2489
			EMD-ARIMA	0.73295	513.1823	520.4388
			傳統 ARIMA	0.74854	513.4015	520.658
	2005~2011 年	2012 年	EMD-ICA-ARIMA	0.79932	504.4235	506.8302
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.78724	515.8164	518.2353
	2006~2012 年	2013 年	EMD-ICA-ARIMA	0.77722	520.0264	524.8398
			EMD-ARIMA	0.78246	516.6077	521.4454
			傳統 ARIMA	0.77507	525.4576	530.2953
	2007~2013 年	2014 年	EMD-ICA-ARIMA	0.72342	533.704	538.5175
			EMD-ARIMA	-	-	-
			傳統 ARIMA	0.71867	543.8059	548.6436
	2008~2014 年	2015 年	EMD-ICA-ARIMA	0.77695	530.6312	535.4446
			EMD-ARIMA	0.76814	536.3276	541.1652
			傳統 ARIMA	0.76654	537.2961	542.1337

資料來源：本研究整理。

附表 2 各年份三種模型之預測能力比較

期 間	年份		模型	RMSE	MAE	U
	訓練組	預測組				
9 年	2002~2010 年	2011 年	EMD-ICA-ARIMA	8.8121	7.2851	0.042434
			EMD-ARIMA	6.0627	4.7219	0.029757
			傳統 ARIMA	9.2534	7.7126	0.044517
	2003~2011 年	2012 年	EMD-ICA-ARIMA	6.7943	6.4541	0.034823
			EMD-ARIMA	-	-	-
			傳統 ARIMA	7.7897	7.7266	0.040834
	2004~2012 年	2013 年	EMD-ICA-ARIMA	5.3467	3.4474	0.026482
			EMD-ARIMA	-	-	-
			傳統 ARIMA	6.0926	3.7054	0.030140
	2005~2013 年	2014 年	EMD-ICA-ARIMA	5.0509	4.8441	0.023936
			EMD-ARIMA	-	-	-
			傳統 ARIMA	6.5385	6.2247	0.031205
	2006~2014 年	2015 年	EMD-ICA-ARIMA	7.9045	7.0865	0.040121
			EMD-ARIMA	-	-	-
			傳統 ARIMA	8.9809	8.0949	0.041133
8 年	2003~2010 年	2011 年	EMD-ICA-ARIMA	6.4947	5.5445	0.031764
			EMD-ARIMA	6.4506	5.3289	0.031582
			傳統 ARIMA	9.0756	7.5532	0.043716
	2004~2011 年	2012 年	EMD-ICA-ARIMA	5.7541	5.2195	0.029332
			EMD-ARIMA	-	-	-
			傳統 ARIMA	6.6866	6.3592	0.034335
	2005~2012 年	2013 年	EMD-ICA-ARIMA	5.3983	3.3882	0.026844
			EMD-ARIMA	-	-	-
			傳統 ARIMA	6.0569	3.6779	0.029963
	2006~2013 年	2014 年	EMD-ICA-ARIMA	4.7813	4.5510	0.022695
			EMD-ARIMA	-	-	-
			傳統 ARIMA	5.1400	4.9238	0.024364
	2007~2014 年	2015 年	EMD-ICA-ARIMA	8.1224	6.3806	0.037534
			EMD-ARIMA	-	-	-
			傳統 ARIMA	8.8292	7.9356	0.040469
7 年	2004~2010 年	2011 年	EMD-ICA-ARIMA	5.7048	3.9161	0.027999
			EMD-ARIMA	4.1852	3.2902	0.020719
			傳統 ARIMA	6.0821	5.0360	0.029663
	2005~2011 年	2012 年	EMD-ICA-ARIMA	6.9462	6.7347	0.035716
			EMD-ARIMA	-	-	-
			傳統 ARIMA	9.1207	8.9248	0.047445
	2006~2012 年	2013 年	EMD-ICA-ARIMA	5.3855	3.4729	0.026899
			EMD-ARIMA	6.0142	3.9090	0.0296
			傳統 ARIMA	6.2379	4.0273	0.030775
	2007~2013 年	2014 年	EMD-ICA-ARIMA	5.4906	4.9425	0.025983
			EMD-ARIMA	-	-	-
			傳統 ARIMA	6.4986	6.1711	0.031934
	2008~2014 年	2015 年	EMD-ICA-ARIMA	8.6122	7.7883	0.039504
			EMD-ARIMA	4.3391	3.8159	0.020441
			傳統 ARIMA	8.7286	7.8247	0.040029

資料來源：本研究整理。