**台灣工業生產指數預測:**

**ARIMA、SARIMA、SVR之模型預測效果比較**

**作者一 作者二 作者三**

**服務單位**

**E-mail**

**摘要**

工業生產指數是台灣的景氣指標之一，可以反映台灣的工業景氣狀況，在現有的研究中已有學者使用自我迴歸整合移動平均模型(ARIMA)進行工業生產指數的預測，不過ARIMA無法考量到季節性的因素，所以本研究將以季節性自我迴歸整合移動平均模型(SARIMA)進行預測，並和ARIMA的預測效果進行比較。由於有七個產業構成台灣工業生產指數，所以除了單變量的自我迴歸模型，本研究也嘗試以支援向量迴歸(SVR)進行多元迴歸預測，並和自我迴歸模型進行比較。

預測結果

**關鍵詞**：工業生產指數、ARIMA、SARIMA、SVR

1. **前言**

景氣指標及景氣對策信號自1977年開始公布，作用為衡量經濟景氣的概況，並截至2018年共經歷了8次修訂。景氣指標包含領先、同時、落後三種景氣指標，並和景氣對策信號由國發會每月同時發布。景氣指標以及景氣對策信號編制方式為將能夠代表經濟活動並可以反映景氣變化的重要總體經濟變數進行統計處理。景氣對策信號會以五種不同信號燈表示各項目的年變動率變化，用以代表景氣狀況。根據2018年8月27日正式啟用的新版景氣指標及對策信號，台灣的同時指標構成項目共有七項，分別是工業生產指數、電力總用電量、製造業銷售量指數、批發、零售及餐飲業營業額、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[1]。

在本研究中，以同時景氣指標中的工業生產指數作為研究對象。台灣的工業生產指數創始於民國42年，統計目的在於建立工礦業產銷查報制度，並按月進行編制，且每隔5年進行基期改編。調查對象依照我國行業統計類(第11次修正)選定的大分類，可分為台閩地區從事礦業及土石採取業、製造業、電力及燃氣供應業、用水供應業的4大行業之企業單位，當中的製造業根據製造業特殊複分類，又可分為金屬機電工業、資訊電子工業、化學工業、民生工業。調查項目依照台灣工業結構，選取1,595項具有重要性、代表性、領導性和策略性之產品，包含製造業1,570項、礦業及土石採取業8項、電力及燃氣供應業10項、用水供應業1項。礦業及土石採取業、電力及燃氣供應業、用水供應業的抽樣方式為進行全查，製造業則採用截斷抽樣法[2]。

在探討台灣工業生產指數和經濟變化相關的研究中，饒秀華透過分期馬可夫轉換模型(MS model)，分析台灣工業生產指數以及實質GDP的年增率[3]。林淑華使用向量自我迴歸模型、Granger因果關係檢定、衝擊反映函數，來了解物價、利率、貨幣供給額、工業生產指數、油價、匯率因素之相互影響關係[4]。而針對台灣工業生產指數的預測，在現有的研究中大多以傳統的時間序列模型來進行預測，余光倫、蘇恆信、邱柏誠等人都曾經以自我迴歸整合移動平均模型(ARIMA)進行台灣工業生產指數的預測[5，6，7]。

季節性自我迴歸整合移動平均模型(SARIMA)是ARIMA的擴展模式，增加了季節性參數的考量，本研究將探討工業生產指數的週期性變化，並比較SARIMA和傳統ARIMA的預測效果。由於台灣工業生產指數是以七種產業的生產指數構成，且支援向量迴歸(SVR)適合用於非線性的迴歸問題，並有許多研究將SVR用於具有時間性的數據預測，所以本研究除了時間序列的預測方式，也將使用SVR以七種工業生產指數作為變量，預測台灣工業生產總指數，並且嘗試透過主成份分析(PCA)、獨立成份分析(ICA)結合支援向量迴歸(SVR)，比較使用成份分析處理的資料和使用原始資料進行SVR訓練的預測績效。

1. **文獻探討**
   1. **工業生產指數**

根據經濟部統計處編印之工業產品分類共1,595項，將性質相同產品整併為723項產品群。台灣工業生產指數的計算方式採用拉氏公式(Laspeyres Formula) 基期加權總值式。計算工業生產指數之公式如下:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

算式中 : t期生產量， : t+1期生產量， : t期附加價值單價[2]。當工業生產指數年增率上升時，代表工業景氣好轉，礦業、製造業、水電燃氣業等的生產量增加，反之，當工業生產指數下降時，則代表工業景氣下滑。

* 1. **自我迴歸整合移動平均模型ARIMA**

自我迴歸整合移動平均模型(Autoregressive Integrated Moving Average Model，ARIMA) 是用於時間序列分析的預測方法。由於在自我迴歸的過程中若存在單位根(Unit Root)，會發生偽迴歸(Spurious Regression)的現象，所以需要透過單根檢定來判斷時間序列是否存在單根，常見的單根檢定有DF檢定(Dickey-Fuller Test)或是其擴展成的ADF檢定(Augmented Dickey-Fuller Test)。在ADF中，將時間序列具有單位根設為虛無假設(Null Hypothesis)，接著比較檢驗統計量(Test Statistic)和顯著水準(Significance Level)，並求出機率p值(p-value)，若p值小於顯著水準，則拒絕虛無假設，即不存在單根，該時間序列為定態(Stationary)；若p值大於顯著水準，則無法拒絕虛無假設，即存在單根，該時間序列為非定態(Non-Stationary)。如果ADF檢定結果為非定態則需要進行差分，以一時間序列{yt}為例，差分過程為:

|  |  |
| --- | --- |
|  | (2) |

式中L為後移一位的滯後運算子(Lag Operator)，若時間序列經過一階差分後，仍為非定態則進行二階差分，依此類推直到時間序列為定態，而差分的次數即為ARIMA的參數d值。當時間序列為定態時，還需要進一步確認定時間序列並非完全隨機分布，而是具有自相關性，所以需要進行白噪聲檢定，常見的檢定方式有Box-Pierce檢驗、Ljung-Box檢驗、自相關圖等。在ARIMA中，透過自我相關函數圖(Autocorrelation Function，ACF)的截尾階數取得參數q，再由偏自我相關函數(PACF)的截尾階數決定參數p。截尾的階數定義為當大於該階數後會快速趨近於0，反之，若整筆資料皆沒有趨近於0或在0區間波動，則為拖尾。

ARIMA模型是由ARMA模型發展而來，由自我迴歸模型(Autoregressive，AR)和移動平均模型(Moving Average，MA)整合(Integration，I)而成。ARIMA模型包含三個參數(p, d, q)，其中p代表自我迴歸階數(AR(p)，q代表移動平均階數(MA(q)，d代表差分次數(I(d))。以一非定態的時間序列{yt}而言，ARIMA函式表示為:

|  |  |
| --- | --- |
|  | (3) |

式中為t-i期的自我相關係數， 為t-i期的移動平均誤差項係數，為白噪聲(White noise)。如果存在多組參數，則需透過赤池信息準則(Akaike Information Criterion，AIC)、貝葉斯信息準則(Bayesian Information Criterion，BIC)、損失函數(Loss Function)等選取配適能力較佳的模型。

ARIMA作為一個常見的時間序列預測模型，廣為應用在各個領域。李台亮以花蓮觀光旅館過往住房數作為觀測數據，透過ARIMA模型預測不同種類客源的住房數，提出業者可以實務應用的系統性預測方法[8]。翁雅君以ARIMA模型針對台北市信義區的YouBike租賃站使用狀況進行分析，並探討收費方式對單車使用狀況的影響，以及單車租賃狀態和周圍環境之間的相關性，提出YouBike的營運建議[9]。Alabdulrazzaq 等人以科威特為案例研究，使用ARIMA模型來預測COVID-19的確診以及康復病例，研究結果顯示即使影響疫情傳播的條件變動和政府不斷修正採取措施，ARIMA仍有相當高的準確性[10]。余光倫以台灣公部門所述單位統計處統計的工業生產指數以及影響台灣工業生產指數的產業，並運用ARIMA模型對這些產業進行預測，證明ARIMA模型適合用於短期的趨勢預測[5]。蘇恆信利用經驗模組分解(EMD)結合獨立成份分析(ICA)，將台灣工業生產指數的時間序列去除雜訊，並使用去除雜訊的資料進行ARIMA預測，證明該方法可以優化ARIMA模型預測結果[6]。

* 1. **季節性ARIMA(SARIMA)**

季節性自我迴歸整合移動平均模型(Seasonal ARIMA)是ARIMA的擴展模式，補足了ARIMA在處理季節性時間序列的缺點。SARIMA可以預測具有季節性或週期性的資料，以ARIMA(p, d, q)三個非季節性參數作為基礎，加入季節性參數(P, D, Q)m，形成SARIMA(p, d, q)(P, D, Q, m)模式，其中P為季節性自我迴歸階數，D為季節差分次數，Q為季節性移動平均階數，m代表時間跨距或周期的參數，其中的季節性參數和ARIMA一樣需要透過DF檢定或是ADF檢定來進行定態檢驗，若為非定態則進行季節性差分決定差分項，季節性差分的過程如下:

|  |  |
| --- | --- |
|  | (4) |

式中Lm為後移m位的滯後運算子。而SARIMA中的參數P以及Q則透過自我相關函數(ACF)、偏自我相關函數(PACF)求得。SARIMA以數學方程式表示為:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

式中與為p和P之自我迴歸函式，和為q和Q之移動平均函式，為白噪音。

SARIMA較ARIMA模型複雜，可以用來預測具有規則性週期的資料。Cong等人收集中國2005年到2018年的流感發病率月度數據，以SARIMA模型進行預測，並證明SARIMA適合對季節性流感進行短期預測[11]。Xu 等人為了預測蘇格蘭魚群洄游的位置以減少漁業損失，將目標海域分為39個區域，並使用SARIMA模型預測各個區域不同時期的海水溫度，在每個區域都得到了良好的預測效果[12]。陳一志進行台灣地區出國人次的預測，並比較灰色預測法、類神經網路、ARIMA和SARIMA四種模型的預測能力，結果顯示SARIMA在該研究中具有最好的預測效果[13]。陳景智使用單變數台灣失業率資料作為訓練資料，並以SARIMA、長短期記憶網路、時序卷積網路四種模型進預測，結果以SARIMA具有最佳的預測績效[14]。蔡宗憲等人以SARIMA模式和Holt-Winters(HW)模式進行短期運旅需求預測，並探討模式更新和Box-Cox方程式進行資料轉換對兩種模式預測的影響，研究結果證明以Box-Cox進行資料轉換可以提升預測效果，且SARIMA比HW模式有更佳的預測表現[15]。

* 1. **支援向量迴歸SVR**

支援向量迴歸(SVR)可以收集特徵空間的幾何特性以及時間序列中的特徵，解決非線性的迴歸問題。SVR的主要概念為將一個在低維度空間中的非線性迴歸問題，將資料映射到高維度的特徵空間中，使問題轉換成線性型式，SVR再以迴歸的方式，建立一個超平面，並盡可能使這個超平面接近所有的資料點，進而找到最佳的預測函式。令SVR透過線性函數f來進行迴歸，以數學式表示為:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

式中w為權重，b為偏誤(Bias)，則為一函式，作用為將資料點(x)轉換到高維度特徵空間。在將非線性資料從低維度映射到高維度空間時，需要由核函式(Kernel Function)進行轉換，常見的核函式包含多項式(Polynimail)、輻射基底函數(Radial Basis Function，RBF)、Sigmoid等。

SVR以ε-insensitive作為損失函式，並以參數ε 定義迴歸函式f(x)周圍的管狀區域，落在該管狀區域內的資料點之損失為零，而落在管狀區域外的資料點則稱為支援向量(Support Vectors)，其損失計算方式則為資料點與管狀邊界的距離，以鬆弛變數(Slack Variable)以及表示。為了避免模型過度擬合，在損失函式中加入懲罰項，即正則化係數(C)，結合結構風險最小化原則(Structural Risk Minimization Principle)，調整方程式權重向量以及偏誤項，使下列方程式最小化:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

式中 為輸入的資料樣本，為輸出值。藉由引入拉式乘數(Lagrange Multipliers)，將方程式轉為對偶問題(Double Problem)後，將方程式解開，求得最適權重量以及偏誤，便可得到迴歸超平面，如下式:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

式中以及為拉氏乘數，K為核函數，核函式將資料映射到高維的特徵空間中，取得兩個特徵向量之內積，表示為。

SVR在迴歸問題中，有良好的預測效果，經常用於財務、股票、工廠產值、自然科學實驗等各類的預測。Mohandes等人比較SVR以及多層感知器進行風速預測的效果，結果得出SVR的預測結果有較低的均方誤差(Mean Square Error)，預測趨勢的結果也較多層感知器佳[16]。Hong透過遺傳模擬退火混合演算法(Genetic Algorithm–Simulated Annealing Hybrid Algorithm，GA-SA)結合SVR形成SVRGA-SA模型，進行台灣3G手機需求預測，並且相較於ARIMA、廣義迴歸類神經網路(General Regression Neural Network)，SVRGA-SA有最佳的預測績效[17]。陳俊宏和呂奇傑以三家光學膜相關的上市公司為研究對象，透過SVR、天真預測法、逐步回迴歸法、倒傳遞類神經網路、極限學習機等五種方式方式，進行光學膜銷售額的預測，並得出SVR相較於其餘四種預測方式，在單期預測以及多期預測都有最佳預測效果[18]。

SVR可以捕捉資料在特徵空間中的幾何性質以及時間序列資料中的特徵，有許多研究證明了SVR適合用於具有時間性的預測。Tay & Cao 將SVR用於金融時間序列的預測，並和倒傳遞類神經網路進行效果比較，證明了SVR由於最小化泛化誤差 (Generalization Error)的上限、較少的自由參數的優點，所以預測效果優於倒傳遞類神經網路[19]。Brakel等人比較SVR、自我迴歸滑動平均模型(ARMA)、Elman神經網路，分別在Mackey-Glas數據集、非線性混沌數據集(Nonlinear Chaotic Data Set)、紗線紡紗過程的過濾值數據等三種數據集的預測效果，並證明了SVR對於時間序列而言為一個良好的預測模型[20]。呂奇傑等人提出利用獨立成份分析(ICA)去除資料雜訊後，再使用SVR進行預測的模式，並以日經225現貨開盤指數以及台股現貨收盤指數作為研究對象，將該模式應用在財務時間序列的預測，並證明該模式的預測表現較單純使用SVR為佳[21]。

* 1. **主成份分析PCA**

主成份分析(Principal Component Analysis，PCA) 是一種統計方法，其主要目的是透過降低多變量資料的維度進行資料分析，並使不同特徵之間可以彼此獨立，只以少數的主成份(Principal Components，PC)便可以解釋原始資料的變異特性。

執行PCA首先需要先將原始資料進行正規化處理，並建立共變異矩陣(Covariance Matrix)，以及求出特徵值(Eigenvalue)以及特徵向量(Eigenvector)，接著特徵向量依照其對應特徵值由大到小進行排列形成特徵向量矩陣(W)，作為資料維度的權重值。主成份計算可表示為:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

式中為第n個主成份，為原始數據的第k個維度， 為特徵向量(表示第k筆數據X維度對應的權重值)。萃取出變異量百分比較大的m個主成份作為新的一組數據，達到降維的效果。

PCA可以在保留資料特性的情況下，提取主成份並去除資料雜訊，以優化運算速度與結果，因此應用在許多運算的資料預處理上。Sousa等人以前饋式類神經網路預測臭氧濃度，比較以原數據輸入和PCA後的主成份輸入對預測效果的影響，最終得出以主成份輸入可以降低模型的複雜性和消除數據共線性，以優化模型預測效果[22]。林子傑以颱風期集水區測站歷史資料作為訓練資料，以倒傳遞類神經網路(BPNN) 預測集水區降雨量，探討篩選參數以及主成份對預測效果的影響，結果證明主成份分析可以提高預測之準確度[23]。李佾璋以類神經網路(ANN)以及SVR預測台灣加權股價指數，並比較以PCA後的主成份做為輸入變數對兩種預測方式的影響，其中SVR適合先執行PCA得到主成份再進行預測，ANN則輸入原資料的技術指標作為輸入變數效果較佳[24]。邱姳旋比較PCA、模糊分群(Fuzzy C-Means，FCM)在不同組合下，對倒傳遞類神經網路(BPN)進行半導體晶圓批生產週期時間預測的效果影響，並得到將三者結合的PCA-FCM-BPN具有最好的預測效果[25]。

* 1. **獨立成份分析ICA**

獨立成份分析法(Independent Component Analysis，ICA)的主要作用為進行訊號處理，起初ICA的目的是處理未知來源分離(Blind Sorse Separation，BSS)的問題[26]，在提出快速獨立成份分析法(Fast Independent Component Analysis，Fast ICA)後[27]，提高了計算效率以及解決了ICA計算複雜的問題，ICA才開始廣為使用於各個領域。ICA的基本模式是透過混合矩陣，將觀測訊號分離出獨立成份(Independent Components，ICs)。假設有一混合訊號矩陣M，其ICA模式如下示所示:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

其中A是未知混合矩陣(Mixing Matrix)，S是未知來源矩陣(Source Matrix)。在來源訊號互為統計獨立的情況下，以一個解混合矩陣(Demixing Matrix)將混合訊號矩陣X進行轉換，產生矩陣Z，表示為:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

式中的是Z的第i個列向量(Row Column)，即獨立成份(IC)，且之間需要互為統計獨立。當時，獨立成份可以用以估計潛在的來源訊號。ICA有許多方法可以用來測量獨立成份的非高斯性(Non-Gaussianity)，常見的有高階統計量法(High Order Cumulants)、共同資訊法(Mutual Information)、負熵法(Negentropy)，而FastICA即是透過極大化負熵法的概念，求得最佳混合矩陣W。

ICA可以從一個未知的混合訊號中，取得潛在的來源訊號。Déniz 比較PCA和ICA兩種資料前處理對於SVM在人臉辨識上的效果差異，並得出兩種方式皆有良好的辨識效果，其中ICA結合SVM有較高的辨識率，但ICA的訓練時間也遠大於PCA[28]。陳則丞應用獨立成份分析進行雜訊處理，依照ICA結果進行分群分析，並得出該方法在SVR預測財務時間序列上有較佳的表現[29]。李天行等人以ICA整合SVR、ICA整合自我迴歸(Autoregressive，AR)、單純SVR、單純AR、隨機漫步(Random Walk)五種方式建構匯率預測模式，結果顯示ICA結合SVR為最正確率最高的預測方式，並證明以ICA進行資料前處理可以改善預測績效[30]。邱柏誠以台灣工業生產指數為研究對象，並利用ICA過濾經驗模組(Empirical Mode Decomposition，EMD)分解產生數個本質模組函數中的雜訊，並結合ARIMA進行績效評估，得出該方式在模型適合度以及預測能力上皆較原資料建構的模型還要有較佳的效果[7]。

1. **研究方法**

台灣經濟部統計處每月會公布的工業生產總指數的原數值以及年增率，而本研究資料採用原指數作為預測對象，並使用自我迴歸整合移動平均模型(ARIMA)、季節性自我迴歸整合移動平均模型(SARIMA)、支援向量迴歸(SVR)進行預測，並比較上述三種模型的預測績效。

* 1. **資料集準備**

根據中華民國經濟部統計數據分析系統所公告列出的工業生產指數，當中構成台灣工業生產指數的組成依照大行業別以及製造業四大行業別進行分類可分為七類，分別為礦業及土石採取業、金屬機電工業、資訊電子工業、化學工業、民生工業、電力及燃氣供應業、用水供應業，其中礦業及土石採取業的數據於1996年1月以前為缺失狀態。台灣工業生產指數的採計又分為包含上述七種行業的總指數以及只囊括六種行業的不含土石採取業之總指數。

本研究將針對包含七種行業之總指數(總指數A)以及不含礦業及土石採取業之總指數(總指數B)，兩種指數進行預測，由於資料統計上的差異，在ARIMA和SARIMA模型的建構上，總指數A將以1997年2月至2019年4月統計共276筆總指數A的歷史數據做為模型訓練資料，並以2019年5月至2023年4月統計共48筆的總指數A歷史數據作為模型測試資料(資料集A1)；總指數B則以1983年2月至2015年10月統計共393筆總指數B的歷史數據做為模型訓練資料，並以2015年11月至2023年4月統計共90筆總指數B的歷史數據作為模型測試資料(資料集B1)。在SVR模型的建構上，總指數A將以1982年1月至2014年12月總指數A及其包含的七種行業各276筆的歷史數據做為模型訓練資料，並以2019年1月至2023年4月統計共52筆的總指數A歷史數據作為模型測試資料(資料集A2)；總指數B則以1982年1月至2014年12月總指數B及其包含的六種行業各397筆的歷史數據做為模型訓練資料，並以2015年1月至2023年4月統計共100筆總指數B的歷史數據作為模型測試資料(資料集B2)，如表()所示。

|  |  |  |
| --- | --- | --- |
| 研究對象  模型種類 | 總指數A | 總指數B |
| ARIMA & SARIMA | 資料集A1  (共315筆)   * 267筆訓練集 * 48筆測試集 | 資料集B1  (共483筆)   * 393筆訓練集 * 90筆測試集 |
| SVR | 資料集A2  (共329筆)   * 276筆訓練集 * 52筆測試集 * 包含7種指標 | 資料集B2  (共497筆)   * 397筆訓練集 * 100筆測試集 * 包含6種指標 |

在ARIMA以及SARIMA模型的訓練上，本研究將以總指數A、B的歷史數據，對應式(3)以及式(5)中的yt，個別進行模型訓練，以預測總指數A、B的未來走勢。在SVR的模型的訓練上，預測總指數A的模型將以礦業及土石採取業、金屬機電工業、資訊電子工業、化學工業、民生工業、電力及燃氣供應業、用水供應業七種行業別作為選取特徵，並以這七種行業的歷史數據作為訓練資料，對應式(8)中的xi，而總指數A則為預測對象並對應式(8)中的f(xi)；總指數B的訓練資料則由上述七種行業中除了礦業及土石採取業的其餘六種行業作為選取特徵，並以這六種行業的歷史數據做為訓練資料，進行總指數B的預測。

* 1. **研究架構與步驟**

**3.2.1 模型訓練步驟與模型種類**

在本研究預測工業生產指數的模型中，依照模型的性質分為兩類，第一類是屬於單變量自我迴歸的ARIMA和SARIMA模型，第二類是屬於多變量迴歸的SVR模型。

在第一類模型中，ARIMA和SARIMA皆是以目標預測對象的歷史數據作為特徵值的自我迴歸模型，其中ARIMA已有學者應用於台灣工業生產指數預測。本研究首先將參考現有的文獻，並調整ARIMA(p, d, q)的參數，建立ARIMA模型進行總指數A以及總指數B的預測。由於ARIMA僅考量了隨時間變化的趨勢以及隨機誤差，當資料具有季節性的週期變化時，ARIMA模型便無法考量到時間季節性的變化，SARIMA作為ARIMA的擴展型態，加入了季節性參數，較ARIMA更加適合預測具有週期性的資料。本研究將以SARIMA模型預測總指數A以及總指數B，找到SARIMA(p, d, q)(P, D, Q)s的較佳參數組合，並和ARIMA的預測效果進行比較，探討季節週期性的變化對於台灣工業生產指數預測是否具有影響。

在第二類模型中，本研究將分為兩種模式執行SVR，分別是以PCA篩選出的主成分進行SVR模型訓練的PCA-SVR模式，以及ICA分解出的獨立成分進行SVR模型訓練的ICA-SVR模式。SVR相較於第一類模型皆以預測對象自身的歷史數據進行自我迴歸，SVR可以收集多變量的特徵，形成多元迴歸，而本研究嘗試以SVR進行總指數A、總指數B的預測，將構成總指數的多個行業作為特徵值，並以這些行業的歷史數據進行模型訓練，探討以多變量進行工業生產指數預測的效果，由於選取的變量數屬於高維度，可能會有雜訊影響訓練效果或是發生危度災難的事件(Curse Of Dimensionality)，所以使用PCA(如式(9)所示)以及ICA(如式(11)所示)進行資料處理與分析，透過主成份和獨立成份的選取來萃取出重要訊息，並形成SVR的訓練特徵輸入進行模型訓練，即為PCA-SVR和ICA-SVR的模式。

**3.2.2 模型評估與挑選之步驟**

在模型的預測能力比較上，本研究將會從第一類模型以及第二類模型中，個別篩選出訓練效果較佳的模型，再針對兩類模型的預測績效進行比較。

第一類模型包含ARIMA和SARIMA，其中ARIMA共有(p, d, q)三個參數，本研究將透過赤池信息準則(AIC)、貝葉斯信息準則(BIC)、均方根誤差(Root Mean Square Error，RMSE)、平均絕對誤差(Mean Absolute Error，MAE)、平均絕對百分比誤差(Mean Absolute Percentage Error，MAPE)等五種指標進行模型訓練成果評估。而SARIMA模型包含(p, d, q)(P, D, Q, m)七個參數，並使用與ARIMA相同的五種評估指標，來比較總指數A與總指數B以不同週期參數m設定之模型訓練成效，以此來找出最理想的參數組合。第二類模型SVR將會比較選取特徵經由PCA與ICA資料篩選過後之模型訓練成效，並透過RMSE、MAE、MAPE三種指標評估模型預測成效，以此來找出最佳的特徵選取方式與SVR模型之最適參數。為了比較第一類自我迴歸模型以及第二類多元迴歸模型的預測效果，最終將針對兩類模型篩選出效果最佳的模型進行比較，使用的評估指標為RMSE、MAE、MAPE。

本研究所使用的評估指標計算方式如表()所示，其中AIC和BIC皆屬於信息準則，用以衡量擬合效果最佳的模型，透過加入懲罰項避免模型的複雜度過高以及發生過度擬合的現象，在模型選擇上，較小的AIC值和BIC值皆是代表較好的模型，當中AIC以預測角度進行模型篩選，BIC則篩選出對現有數據擬合程度較佳的模型。而RMSE、MAE、MAPE則屬於損失函數，用以衡量預測值與實際值之間的誤差，損失值越小代表模型的預測能力越好。RMSE呈現實際值與預測值的平均誤差平方合的平方根。MAE則直接表示預測值與實際值之間的絕對誤差。MAPE則是呈現損失的百分比，但資料集中的實際值不可存在0。

|  |  |
| --- | --- |
| 指標種類 | 公式 |
| AIC |  |
| BIC |  |
| RMSE |  |
| MAE |  |
| MAPE |  |
| L為概似函數，k為參數數量，n為樣本數量，為預測值，為實際值 | |

1. **研究結果分析**
   1. **第一類模型**

第一類自我迴歸模型中的ARIMA模型，在現有的研究中已有學者應用於台灣工業生產指數預測，而本研究將使用加入了季節性參數的SARIMA模型進行工業生產指數的預測，並和ARIMA的效果進行比較。第一類模型的訓練資料將會使用資料集A1與資料集B1，並相互比較訓練結果以找出最佳的模型。在模型訓練前，須要先將資料進行單根檢定與白噪聲檢定，使訓練資料符合單變量自我回歸模型的條件。

**4.1.1 資料平穩性檢定-初步檢定**

本研究將會使用ADF檢定，以此檢查資料是否為定態，初步檢查結果如表()，當中包含了資料集A1與資料集B1之ADF檢定結果。將會透過兩階段比較：第一階段，針對檢驗統計量數值比較，將會與1%、5%、10%之不同程度拒絕虛無假設的統計值進行比較，若檢定統計量小於1%、5%、10%數值，代表本研究具有足夠的把握(99%、95%或90%)來拒絕虛無假設。根據表()所示，兩資料集之檢驗統計量均沒有小於三個階段(1%、5%、10%)之統計值，由此推斷該時間序列為非定態。第二階段，針對p-value數值，若小於0.05則可以拒絕虛無假設，根據表()所示，兩資料集之p-value均為小於0.05，因此無法拒絕虛無假設，從以上結果可以得出本研究之資料需要進行序列差分史資料達到定態。

|  |  |  |
| --- | --- | --- |
| 檢定結果 | 資料集A1 | 資料集B1 |
| 檢驗統計量 | -0.875106 | -0.411091 |
| p-value | 0.796060 | 0.908230 |
| 1% | -3.451553 | -3.444076 |
| 5% | -2.870879 | -2.867593 |
| 10% | -2.571746 | -2.569994 |

**4.1.2 差分法**

透過差分法可以使資料轉為定態，圖()當中表示了資料經過差分前(顏色)與後(顏色)的圖形差異，經過差分法的數據，由圖()可看出序列已有顯示出趨近定態之表現。

**4.1.3 資料平穩性檢定-二次檢定**

經由差分法將資料轉為定態後，表()為資料集A1與B1再次進行ADF檢定之結果。第一階段之分析，如表()當中可以得出檢驗統計量分析結果兩資料集均小於1%、5%、10%之數值，因此在第一階段能夠得出本處理後之資料具備足夠的把握能夠說明兩資料集為時間序列平穩資料。第二階段，根據兩資料集之結果，p-value數值均遠小於0.05，以此拒絕虛無假設。經由序列平穩化處理過後，資料均為時間序列平穩資料。

|  |  |  |
| --- | --- | --- |
| 檢定結果 | 資料集A1 | 資料集B1 |
| 檢驗統計量 | -7.680571e+00 | -6.242601e+00 |
| p-value | 1.507817e-11 | 4.657816e-08 |
| 1% | -3.452190e+00 | -3.444491e+00 |
| 5% | -2.871158e+00 | -2.867776e+00 |
| 10% | -2.571895e+00 | -2.570091e+00 |

**4.1.4 白噪聲檢定**

此檢定將使用Ljung-Box檢定，以檢查樣本資料是否為白噪聲序列資料，若p-value小於0.05則拒絕遠假設，根據表()為兩資料集檢定結果，當中兩檢驗之p-value數值均小於0.05，因此我們能夠以95%的把握拒絕原假設，認定兩資料集數據為定態且為非白噪聲序列。

|  |  |  |
| --- | --- | --- |
| 檢定结果 | 資料集A1 | 資料集B1 |
| 統計量 | 61.592389 | 98.502071 |
| p-value | 4.224483e-15 | 3.246849e-23 |

**4.1.5 模型訓練與結果分析 - ARIMA**

模型建構將會使用Python的pmdarima套件提供之 auto\_arima方法來找出ARIMA與SARIMA的最佳參數，其中針對SARIMA週期參數篩選，本研究將週期參數設置為2至48週，逐一比較週期參數的適配度，並以AIC值、BIC值當作主要評斷標準，且以MAE、MAPE、RMSE為輔，判斷模型的預測能力。

ARIMA之訓練結果分析如表()所示，表中為兩資料集進行模型訓練之後的結果，在資料集A1的訓練結果可以發現， AIC、BIC與MAPE數值均比資料集B1還低，但RMSE與MAE卻得到相對好一點的效果。由以上結果得出小結，雖然資料集A1之模型評估指標的數值並非完全優於資料集B1，但以主要比較標準AIC與BIC且輔助比較指標MAPE均優於資料集B1，以此認定模型以資料集A1當作訓練集將會有較好的模型效果。

|  |  |  |
| --- | --- | --- |
|  | 資料集A1 | 資料集B1 |
| AIC | 1377.146 | 1845.483 |
| BIC | 1398.669 | 1857.404 |
| RMSE | 5.6904104 | 4.904682 |
| MAE | 4.0088449 | 3.526743 |
| MAPE | 0.98274562 | 1.001951 |
| 參數(p, d, q) | (1, 0, 4) | (0, 0, 2) |

**4.1.5 模型訓練與結果分析 – SARIMA**

針對SARIMA之訓練結果比較，將會藉由調整週期參數來尋找對於模型最好的參數設置，週期將會從2至48週期當中尋找，並從資料集A1與資料集B1的結果當中，選出訓練效果較佳的前五種參數組合，再進行第二輪的比較。

根據資料集A1研究結果如表()所示，以主要評估指標AIC與BIC來看，週期為48與44時會有較低的AIC、BIC數值。再由輔助判斷指標判斷48與44周期的預測效果，48週期的MAPE指標數值僅低於47週期，而44週期的RMSE與MAE指標數值相較於其他週期參數設定的數值，44週期呈現出最佳的預測效果，且MAPE的數值也只大於12週期的數值。根據此研究數據判斷，44與48的週期參數設定所呈現的AIC、BIC數值相當，且於輔助指標的數值44週期皆低於48週期，顯示出較佳的預測效果，因此根據評估指標的篩選，資料集A1的週期參數設定以44週較為適合。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 週期  指標 | 48 | 44 | 47 | 46 | 12 |
| AIC | **1192.333** | **1203.208** | **1227.911** | **1234.659** | **1256.449** |
| BIC | **1212.667** | **1227.058** | **1251.666** | **1255.048** | **1281.238** |
| RMSE | **5.3866** | **4.4971** | **6.2886** | **5.3132** | **5.1208** |
| MAE | **3.9159** | **3.5767** | **4.7587** | **3.7645** | **3.6063** |
| MAPE | **2.5013** | **1.8932** | **4.1578** | **2.2941** | **1.2930** |
| 參數  (p,d,q)  (P,D,Q) | **(0, 0, 2)**  **(1, 1, [1])** | **(0, 0, 2)**  **(2, 1, [1])** | **(2, 0, 1)**  **(2, 1, [])** | **(0, 0, 2)**  **(2, 1, [])** | **(1, 0, 0)x**  **(2, 1, [1, 2])** |

註1:參數數值當中存在空的中括號[]，代表該參數若不輸入任何數值將會有較好的結果，反之，若括號當中存在多個數值，代表其數值均會有一樣的結果。

註2:表格內根據AIC與BIC表現，由左向右為數值由低到高排序。

表()為資料集B1的訓練結果，當週期設定為44時，會有最低的AIC、BIC數值，且針對其他輔助指標來看， 44週期仍然擁有最好的模型預測效果。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 週期  指標 | 44 | 48 | 32 | 42 | 47 |
| AIC | 1626.542 | 1634.827 | 1676.354 | 1714.125 | 1717.119 |
| BIC | 1653.527 | 1661.732 | 1703.576 | 1741.151 | 1744.044 |
| RMSE | 4.08072 | 4.88606 | 4.88415 | 4.94889 | 5.51968 |
| MAE | 3.2217 | 3.5516 | 3.6233 | 3.7466 | 4.1711 |
| MAPE | 1.8700 | 3.3279 | 3.4630 | 1.9468 | 3.2796 |
| 參數  (p,d,q)x  (P,D,Q) | (2, 0, 1)x  (2, 1, []) | (0, 0, 2)x  (2, 1,[1]) | (0, 0, 2)x  (2, 1,[1]) | (0, 0, 2)x  (2, 1,[1]) | (3, 0, 0)x  (2, 1, 0) |

**4.1.6 第一類模型之比較 – ARIMA、SARIMA**

根據以上研究結果分析，可以得出以下二個結論，第一點，透過評估指標比較ARIMA模型與SARIMA模型，使用不同的資料集進行模型訓練，SARIMA均呈現較好的訓練效果(如表()所示)，因此相較於ARIMA模型，SARIMA更加適合使用於工業生產指數的預測。第二點，在資料集的選擇上，當資料集A作為ARIMA與SARIMA模型的訓練資料時，主要判斷指標數值均會呈現較好的模型結果，且資料集A1、B1之間存在的差異為使否包含礦業及土石採取業，由此可初步推測該行業對於模型成效之表現具有影響力，因此在資料集的挑選上，資料集A為較適合作為第一類模型的訓練集。針對第一類模型的統整結論，當模型選用SARIMA，且使用資料集A1進行訓練，並將季節性參數之週期設定為44時，即為第一類模型中訓練效果最佳的預測模型。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 各最佳  模型  指標 | 資料集A1 | | 資料集B1 | |
| 週期:44  (SARIMA) | (ARIMA) | 週期:44  (SARIMA) | (ARIMA) |
| AIC | 1203.208 | 1377.146 | 1626.542 | 1845.483 |
| BIC | 1227.058 | 1398.669 | 1653.527 | 1857.404 |
| RMSE | 4.4971 | 5.6904 | 4.0807 | 4.9047 |
| MAE | 3.5767 | 4.0088 | 3.2217 | 3.5267 |
| MAPE | 1.8932 | 0.9827 | 1.87 | 1.0019 |
| 參數  (p,d,q)x  (P,D,Q) | (0, 0, 2)x  (2, 1, [1]) | (1, 0, 4) | (2, 0, 1)x  (2, 1, []) | (0, 0, 2) |

* 1. **第二類模型**

SVR模型將會配合PCA與ICA以進行以下的分析，進行PCA與ICA分析前，資料集均會先進行標準化處理，標準化處理將採用sklearn預處理套件裡所提供的標準化函式。模型將會使用資料集A2(七種指標，訓練資料為276)與資料集B2(六種指標，數量為397)進行訓練，本研究將能藉由資料集當中，不同類別數量與不同資料數量進行模型訓練，以從訓練結果分析出不同資料類別對模型產生的影響，以求得到最佳的模型效果。模型評估指標指標將會採用，均方根誤差(RMSE)、平均絕對值誤差(MAE)、平均絕對百分比誤差(MAPE)來衡量模型的訓練效果。

**4.2.1 PCA-資料集A之特徵篩選與分析**

資料集A2進行標準化處理以後，將其進行PCA之分析，從圖()所示，選取單一主成分PC1時，重要性程度已達到接近70%。當選取PC1與PC2時，重要性程度就能夠達到超過80%的效果。若選取PC1、PC2、PC3時，重要性程度將可達到約9成的效果。針對表()所示，能夠看到電力及燃氣供應業占了PC1最高的比重。從圖()所示，經由PCA所分析出的重要性程度前三高的，分別是電力及燃氣供應鏈業、礦業及土石採取業、用水供應業。

(圖片)

**4.2.2 PCA資料集B之特徵篩選與分析**

資料集B2進行標準化處理以後，將其進行PCA之分析，從圖()所示，單一選取PC1時，重要性程度已接近80%。當選取PC1與PC2時，重要性程度就能夠將近9成的效果。針對表()可以看到，在PC1當中，所占重要性程度最高的為電力及燃氣供應業。從圖()所示，可以看到重要性前二高的分別是電力及燃氣供應業與化學工業。

(圖片)

**4.2.3 SVR - PCA模型預測效果**

SVR模型配合PCA之訓練結果，資料集A2如表()與圖()、資料集B2如表()與圖()，當中包含選取指標類別數量的結果差異，從資料集A2研究結果所示，當主成分數量選取7個成分，也就是選取所有成分時，在每一個判斷指標上均會有最佳的表現，且在選擇數量5以後才會開始有數值效果提升的現象，當我使用資料集A2進行訓練時，至少選取5種以上的特徵，將會有較好的效果產生。

針對資料集B2之研究結果所示，當主成分數量選取全部時仍會有最好的效果， 且在主成分數量選擇上，當我選擇至少4種特徵時才會有效果提升的現象出現，因此當我使用資料集B進行訓練時選取至少4種特徵，將會有較好的效果產生。

|  |  |  |  |
| --- | --- | --- | --- |
| (A2)  主成分數量 | RMSE | MAE | MAPE |
| 1 | 3.753504 | 14.08879716 | 0.1476148883 |
| 2 | 4.123241 | 17.00112319 | 0.175721453 |
| 3 | 3.788718 | 14.35438261 | 0.148102028 |
| 4 | 3.798697 | 14.43009661 | 0.149817334 |
| 5 | 3.640276 | 13.25160582 | 0.136723962 |
| 6 | 3.384559 | 11.45524253 | 0.117506565 |
| 7 | 3.000901 | 9.005409531 | 0.091691397 |

(圖片)

|  |  |  |  |
| --- | --- | --- | --- |
| (B2)  主成分數量 | RMSE | MAE | MAPE |
| 1 | 3.511746363 | 12.33236 | 0.133264 |
| 2 | 3.941298108 | 15.53383 | 0.168121 |
| 3 | 3.519585862 | 12.38749 | 0.129599 |
| 4 | 3.334334346 | 11.11779 | 0.116470 |
| 5 | 3.194135401 | 10.20250 | 0.107118 |
| 6 | 2.927799947 | 8.572013 | 0.089725 |

(圖片)

**4.2.4 SVR - ICA模型預測效果**

SVR模型結合ICA進行特徵分解後的訓練結果，資料集A2如表()與圖()、資料集B2如表()與圖()，當中包含不同成分數量的結果差異。根據資料集A2訓練結果，可以發現當選擇成分數量為3時，將會有最佳的RMSE、MAE、MAPE數值產生，且從表()可以發現，當成分數量大於且包含4時，模型成效將會有下降的發生，因此當使用資料集A2訓練時，ICA數量選擇小於且包含3時將會有較理想的效果產生。

針對資料集B2之比較結果所示，當我成分數量選擇為1或2時，RMSE、MAE、MAPE均會有較好的效果呈現，反之，當我數量選擇大於且包含3時將會有模型成效下降的效果產生，因此當使用資料集B2訓練時，選擇1或2時，將會有較理想的效果產生。

|  |  |  |  |
| --- | --- | --- | --- |
| (A2)  成分數量 | RMSE | MAE | MAPE |
| 1 | 4.207183 | 17.700388 | 0.189774 |
| 2 | 4.141761 | 17.154186 | 0.18652 |
| 3 | 4.079945 | 16.645952 | 0.17942 |
| 4 | 4.800954 | 23.049163 | 0.249045 |
| 5 | 5.041118 | 25.412866 | 0.274285 |
| 6 | 5.517577 | 30.443651 | 0.328314 |
| 7 | 5.706606 | 32.565356 | 0.351449 |

(圖片)

|  |  |  |  |
| --- | --- | --- | --- |
| (B2)  成分數量 | RMSE | MAE | MAPE |
| 1 | 4.98264 | 24.826701 | 0.290126 |
| 2 | 5.196112 | 26.999578 | 0.317405 |
| 3 | 5.754442 | 33.113605 | 0.389774 |
| 4 | 6.322055 | 39.968381 | 0.468843 |
| 5 | 6.567239 | 43.128626 | 0.506157 |
| 6 | 6.695746 | 44.833015 | 0.526332 |

(圖片)

**4.2.5 第二類模型之比較 – SVR**

根據以上SVR研究結果以進行分析，表()為SVR結合PCA與SVR結合ICA之較好模型結果之整理，當中將挑選各模型之效果較好的2個結果。根據結果，可以得出以下二點小結論，第一點，針對SVR配合PCA與ICA的挑選上，以模型之判斷指標好壞來看，當選擇PCA配合SVR時，將會有較好的模型成效。第二點，在訓練集的挑選上在PCA訓練時，選擇資料集B2將會有較好的結果，但當我使用ICA時，資料集A2則會有較好的訓練效果。統整上述，今天使用SVR-PCA且選用資料集B2，將會有整體SVR模型訓練之最好成效。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 資料集  指標 | PCA | | | | ICA | | | |
| (A2) | (A2) | (B2) | (B2) | (A2) | (A2) | (B2) | (B2) |
| 成分數量 | 6 | 5 | 6 | 5 | 3 | 2 | 2 | 1 |
| RMSE | **3.38456** | **3.64028** | **2.9278** | **3.194135** | **4.07995** | **4.14176** | **5.19611** | **4.9826** |
| MAE | **11.455243** | **13.25161** | **8.572013** | **10.20250** | **16.64595** | **17.1542** | **26.9996** | **24.8267** |
| MAPE | **0.117507** | **0.136724** | **0.089725** | **0.107118** | **0.17942** | **0.18652** | **0.31741** | **0.29013** |

* 1. **模型比較統整**

統整以上所有模型訓練之比較，將會針對第一類模型與第二類模型進行模型成效比較，整理於下表()，當中為兩類模型當中最好的參數設定與模型成效。

|  |  |  |
| --- | --- | --- |
| 模型&  資料集  指標 | SARIMA | SVR-PCA |
| 資料集A1 | 資料集B2 |
| RMSE | 4.4971 | 2.9278 |
| MAE | 3.5767 | 8.572013 |
| MAPE | 1.8932 | 0.089725 |
| 參數 | * (p,d,q)x(P,D,Q)   = (0, 0, 2)x(2, 1, [1])   * 週期: 44 | * 成分數量選擇: 6 |

1. **結論**

**參考文獻**

1. 景氣指標查詢系統，檢自https://index.ndc.gov.tw/n/zh\_tw (July 12, 2023)
2. 經濟部統計處工業產銷存動態調查簡介，檢自https://www.moea.gov.tw/Mns/dos/content/ Content.aspx?menu\_id=6820(June 28, 2023)
3. 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。
4. 林淑華(2010)，總體經濟因子與台美匯率關聯性之探討，樹德科技大學。
5. 余光倫(2022)，應用 ARIMA 模型對影響台灣工業生產指數之產業預測與分析，逢甲大學。
6. 蘇恆信(2012)，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，宜蘭大學。
7. 邱柏誠(2016)，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，宜蘭大學。
8. 李台亮(2013)，旅館需求預測之研究-以花蓮C飯店為例，東華大學。
9. 翁雅君(2015)，透過時間序列方法分析YouBike使用狀況:以信義區為例，雲林科技大學。
10. Haneen Alabdulrazzaq, Mohammed N. Alenezi, Yasmeen Rawajfih, Bareeq A. Alghannam, Abeer A. Al-Hassan, & Fawaz S. Al-Anzi(2021). On the accuracy of ARIMA based prediction of COVID-19 spread, 27.
11. J Cong, M Ren, S Xie, & P Wang(2019). Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018, 16(23).
12. Feng Xu, Yu-Ang Du, Hong Chen, & Jia-Ming Zhu(2021). Prediction of Fish Migration Caused by Ocean Warming Based on SARIMA Model.
13. 陳一志(2007)，台灣地區出國人次之預測-灰色預測法、類神經網路、ARIMA、SARIMA模型之應用。
14. 陳景智(2021)，以季節性差分整合移動平均自迴歸模型與類神經網路預測台灣失業率，清華大學。
15. 蔡宗憲、李治綱(2012)，一個混合時間序列法與指數平滑法的預測流程：列車旅運需求預測之應用，運輸學刊，24(1)，95~112。
16. M.A. Mohandes, T.O. Halawani, S. Rehman, Ahmed A. Hussain(2004). Support vector machines for wind speed prediction, Renewable Energy, 29(6), 939-947.
17. Wei-Chiang Hong, Yucheng Dong, Li-Yueh Chen, Chien-Yuan Lai(2010). Taiwanese 3G mobile phone demand forecasting by SVR with hybrid evolutionary algorithms, Expert Systems with Applications, 37(6), 4452-4462.
18. 陳俊宏、呂奇傑(2015)，應用與比較多種預測技術於光學膜產業銷售額預測，數據分析，10(5)，105-127。
19. Francis E.H. Tay, Lijuan Cao(2001). Application of support vector machines in financial time series forecasting, Omega, 24(4), 309-317.
20. U Thissen, R van Brakel, A.P de Weijer, W.J Melssen, L.M.C Buydens(2003). Using support vector machines for time series prediction, Chemometrics and Intelligent Laboratory Systems, 69(1-2), 35-49.
21. 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。
22. S.I.V. Sousa, F.G. Martins, M.C.M. Alvim-Ferraz, M.C. Pereira(2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Environmental Modelling & Software, 22, 97-103.
23. 林子傑(2022)，結合類神經網路及主成份分析於颱風時期大氣參數對降雨量推估之研究，淡江大學。
24. 李佾璋(2010)，類神經網路與支援向量迴歸實作之投資決策模擬-以台灣加權股價指數為例，天主教輔仁大學。
25. 邱姳旋(2012)，晶圓批生產週期時間預測:結合主成份分析、模糊分群與倒傳遞網路，逢甲大學。
26. Juha Karhunen, Aapo Hyvarinen, Ricardo Vagcirao, Jarmo Hurri and Erkki Oja(1997). APPLICATIONS OF NEURAL BLIND SEPARATION TO SIGNAL AND IMAGE PROCESSING, IEEE,
27. Aapo Hyvärinen(1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis, IEEE Trans. on Neural Networks, 10(3), 626-634.
28. O. Deniz , M. Castrillon, M. Hern andez(2003), Face recognition using independent component analysis and support vector machines, Pattern Recognition Letters, 24(13), 2153-2157.
29. 陳則丞(2012)，以獨立成份分析為基之支援向量迴歸模式預測時間系列股價，中山大學。
30. 李天行、呂奇傑、劉彥均(2012)，整合獨立成份分析特徵萃取技術與支援向量迴歸建構匯率預測模式，清雲學報，32(2)， 15-37。