**台灣工業生產指數預測:**

**ARIMA、SARIMA、SVR之模型預測效果比較**

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，景氣指標包含領先、同時、落後三種景氣指標，本研究以同時景氣指標中的工業生產指數作為研究對象，並比較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)三種模型預測的效果。

將本研究所使用的模型依照模型性質分為兩類，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。第一類模型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。本研究將使用加入了季節性參數的SARIMA模型進行預測，並和ARIMA的效果進行比較。在第二類模型中，本研究將分為三個模式執行SVR，分別為以原始數據做為訓練資料的單純SVR模式，結合主成份分析(PCA)的PCA-SVR模式、結合獨立成份分析(ICA)的ICA-SVR模式。

本研究之結果分析如下，針對第一類單變量自我回歸模型之資料分析與模型訓練過程，首先，資料集將會分為資料集A(包含六種指標，資料數量為497)、資料集B(包含七種指標，資料數量為329筆)，指標數量差別於礦業及土石採取業採取與否，為了符合該模型之資料條件，需要經過資料平穩性與白噪聲檢定，資料集經過檢定與預處理過後，均符合單變量自我回歸模型條件。

模型訓練，建構ARIMA模型將會使用pmdarima之 auto\_arima方法來找出ARIMA與SARIMA之(p,d,q)(P,D,Q)最佳參數，且針對季節性ARIMA部分，週期參數之篩選，將會透過從週期2至48週來進行模型成效的比對，從AIC值、BIC值、R平方、RMSE來判斷模型的好壞。

模型訓練成效，針對ARIMA之兩個資料集訓練比較，將會主要以AIC、BIC指標作為主要評斷依據，並且以R平方、RMSE為輔。由表(1)所示，可以得出模型的好壞，在資料集B訓練過後的結果所示，在AIC與BIC以及R平方的數值上均勝過於資料集A之數值，但在RMSE部分，資料集A擁有較好的結果。這部分認定以資料集B訓練ARIMA模型將會得出較理想的結果。

表 1. ARIMA之結果比較

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | AIC | BIC | R平方 | RMSE | 參數(p, d, q) |
| 資料集A | 1845.483 | 1857.404 | -0.0015747017339 | 4.904682246580955 | (0, 0, 2) |
| 資料集B | 1377.146 | 1398.669 | -0.0011266093749 | 5.690410352653589 | (1, 0, 4) |

資料來源 : 本研究整理。

針對 SARIMA之比較，評估方法與ARIMA相同，比較當中針對不同週期來比較訓練成效，將會從兩資料集訓練後的結果個別挑選出前5名以此來做數值評估的比較，表(2)為資料集A，表(3)為資料集B之模型訓練比較，將會透過在不同週期下所得出的結果，以AIC、BIC當作主要評斷標準，且以R平方、RMSE為輔。根據研究結果可得出當週期參數挑選為44或48週期時，在這兩個資料集的訓練下都會有不錯的效果，且44與48週期的結果可以44週期在AIC與BIC的表現與48週期並不會有太大的差距，且在R平方與RMSE的表現上，44週期也呈現出較好的表現。

表 2. SARIMA之資料集A結果比較

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **資料集A** | AIC | BIC | R平方 | RMSE | 參數(p, d, q)(P, D, Q) |
| 週期 : 44 | 1626.542 | 1653.527 | 0.3066786238745429 | 4.080717199879909 | (2, 0, 1)x(2, 1, []) |
| 週期 : 48 | 1634.827 | 1661.732 | 0.01619726619479811 | 4.886060297715177 | (0, 0, 2)x(2, 1, [1]) |
| 週期 : 32 | 1676.354 | 1703.576 | 0.006793701838250565 | 4.884149352556783 | (0, 0, 2)x(2, 1, [1]) |
| 週期 : 42 | 1714.125 | 1741.151 | - 0.019712850790046 | 4.948893974923373 | (0, 0, 2)x(2, 1, [1]) |
| 週期 : 47 | 1717.119 | 1744.044 | -0.2684956411685837 | 5.519677362240856 | (3, 0, 0)x(2, 1, 0) |

資料來源 : 本研究整理。

表 3. ARIMA之資料集B結果比較

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **資料集B** | AIC | BIC | R平方 | RMSE | 參數(p, d, q)(P, D, Q) |
| 週期 : 48 | 1192.333 | 1212.667 | 0.10292558706214 | 5.386582529303062 | (0, 0, 2)x(1, 1, [1]) |
| 週期 : 44 | 1203.208 | 1227.058 | 0.37473254091437 | 4.497094326595549 | (0, 0, 2)x(2, 1, [1]) |
| 週期 : 47 | 1227.911 | 1251.666 | -0.22265923374315 | 6.288568550392904 | (2, 0, 1)x(2, 1, []) |
| 週期 : 46 | 1234.659 | 1255.048 | 0.12719378128166947 | 5.313222454327524 | (0, 0, 2)x(2, 1, []) |
| 週期 : 12 | 1256.449 | 1281.238 | 0.189274025031811 | 5.120780108962516 | (1, 0, 0)x(2, 1, [1, 2]) |

資料來源 : 本研究整理。

針對ARIMA所做出小結論，由以上結果可以得出三個結論，第一，當選擇單變量自我回歸模型時，可以發現透過季節性ARIMA時，在不同的資料集下AIC與BIC均會表現出較好的結果。第二，資料集的選擇上，在ARIMA與SARIMA之訓練成效比較下，可以發現，當使用資料集B以訓練模型時，AIC與BIC均會呈現出較理想的結果。第三，參數設定上，當今天加入季節性之參數時，週期為44時，對於模型效果上會有最好的結果。

針對第二類多變量迴歸的SVR，本研究將以python中的sklearn套件見建構模型，並透過調整正則化參數C來調整SVR模型的擬合程度，並使用均方誤差(MSE)來衡量模型的預測效果。本研究將以構成工業生產指數的礦業及土石採取業、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、用水供應業指標作為特徵選取，進行總指數以及礦業及土石採取業礦業及土石採取業的預測。根據研究結果，在單純SVR模式中，C的數值為25，MSE為17.1694，模型的預測效果極為不佳，而在PCA-SVR模式中，經過PCA篩選出的主成分中，電力及燃氣供應業的貢獻度最大，高達70%，而礦業及土石採取業以用水供應業，