

逢甲大學學生報告 ePaper

報告題名：

時間序列分析—桃園國際機場進出旅客人數

作者：張芸菁、顏嘉佺、柯雅潔、謝宇凡、呂嘉侑

系級：統計三甲

學號：D0738736、D0711505、D0780895、D0780906、D0738807

開課老師：陳婉淑老師

課程名稱：預測分析

開課系所：統計系

開課學年： 109 學年度 第 二 學期

目錄

第一章、緒論.....	5
第二章、研究方法與檢定.....	6
第一節、研究方法.....	6
第二節、檢定方法及預測準則.....	9
第三章、預測分析.....	11
第一節、資料敘述.....	11
第二節、分解法.....	12
第三節、Box-Jenkins ARIMA 模型	18
第四節、時間序列迴歸.....	23
第五節、指數平滑法.....	27
第四章、結論.....	30
附錄一、工作分配表.....	32
附錄二、參考資料.....	33



圖目錄

圖 1、殘差的自相關檢定範圍.....	9
圖 2、桃園國際機場進出旅客人數-原始資料	11
圖 3、分解法-原始時間序列圖	12
圖 4、分解法-趨勢循環圖	13
圖 5、分解法-季節性波動圖	13
圖 6、分解法-不規則變化圖	14
圖 7、分解法-去季節化觀察值圖	14
圖 8、分解法-樣本外預測圖	17
圖 9、ARIMA-時間序列、ACF、PACF 與 IACF	18
圖 10、ARIMA-一次差分後的 ACF、PACF 與 IACF	19
圖 11、ARIMA-一次季節差分後的 ACF、PACF 與 IACF	19
圖 12、ARIMA-一次差分與一次季節差分後的 ACF、PACF 與 IACF	20
圖 13、ARIMA-配適完後的 ACF、PACF、IACF 與 White Noise	21
圖 14、ARIMA-樣本外預測圖	22
圖 15、時間序列迴歸-白噪音與單根檢定	錯誤! 尚未定義書籤。
圖 16、時間序列迴歸-樣本外預測圖	26
圖 17、Holt-Winters Multiplicative-樣本外預測圖	28

表目錄

表 1、加法與乘法模型.....	6
表 2、分解法- Durbin-Watson 統計值(修改前).....	15
表 3、分解法- Durbin-Watson 統計值(修改後).....	15
表 4、分解法-自迴歸參數估計值	16
表 5、分解法-參數估計值	16
表 6、分解法-樣本外預測之 MAD、MSE、MPE、MAPE	17
表 7 (a)、ARIMA-參數估計表	20
表 7 (b)、ARIMA-參數估計表	20
表 8、ARIMA-樣本外預測之 MAD、MSE、MPE、MAPE	22
表 9、12 個月份虛擬變數.....	23
表 10、時間序列迴歸-參數估計值	24
表 11、時間序列回歸-樣本外預測之 MAD、MSE、MPE、MAPE	26
表 12、Holt-Winters Multiplicative 指數平滑法參數估計表	27
表 13、Holt-Winters Multiplicative 樣本外預測之 MAD、MSE、MPE、MAPE.....	29
表 14、樣本外預測之 MAD、MSE、MPE、MAPE 比較表.....	30





第一章、緒論

臺灣地理位置位於亞洲門戶，不論在經貿交流或者是觀光旅遊，臺灣都是相當重要的轉運站，因此桃園國際機場作為臺灣的門戶，為臺灣航空的交通樞紐，對於臺灣觀光產業以及貿易出口的發展有舉足輕重的地位。

桃園國際機場擁有聯外之高速公路及捷運系統，是臺灣最大的民航機場，根據 2019 年的統計，其客運吞吐量位居世界第 36 位，貨運處理量則居世界第 5 位。近年來，由於新南向政策的施行以及廉價航空的興起，臺灣桃園國際機場的旅客流量逐年攀升，在 2018 年當中更是創下 4653 萬人次的新紀錄，如此巨大的旅客流量，也成功讓桃園國際機場擠身進入全球前 50 大機場之列，此項成就也顯示了桃園國際機場作為東亞航空樞紐的重要地位。

此象徵著桃園國際機場的進出旅客人數多寡，勢必牽動著臺灣的經濟發展、文化交流，所以我們想要藉由桃園國際機場進出旅客人數的歷史資料，進行時間預測分析，以便了解未來桃園國際機場的旅客流量會如何發展。



第二章、研究方法與檢定

第一節、研究方法

1. 時間序列分解法(Time Series Decomposition)：

時間序列分解法主要由趨勢變動(Trend)、循環波動(Cycle)、季節性波動(Seasonal Variations)、不規則變化(Irregular Fluctuations)四個成分組成。

- **趨勢變動**：以長期來看，序列的發展狀態隨著時間成長或下降的趨勢。
- **循環波動**：在某段時間內，有週期性的連續變動，但波動的頻率與幅度較不嚴格。
- **季節性波動**：因季節變化所造成的規律波動。
- **不規則變化**：許多不同因素影響數值的變化。

時間序列分解法模型又分為加法模型與乘法模型。

- **加法模型**：觀察原始時間序列圖，當季節波動不會隨著時間序列的水平增長與遞減而大幅度變異，加法模型較適合分析該時間序列圖。
- **乘法模型**：當季節波動會隨著時間序列的水平增長而大幅度的變異時，乘法模型較適合分析該時間序列。

表 1、加法與乘法模型

加法模型	乘法模型
$y_t = Tr_t + Cl_t + Sn_t + Ir_t$	$y_t = Tr_t \times Cl_t \times Sn_t \times Ir_t$

2. Box-Jenkins ARIMA 模型：

Box-Jenkins 方法由 George Box 和 Gwilym Jenkins 在 1970 於教科書《Time Series Analysis: Forecasting and Control》提出。該方法之假設為，時間序列若是平穩的，可以使用 ARMA 模型來近似；時間序列若是不平穩的，則使用 ARIMA 模型來近似。

ARIMA 模型，全名為整合移動平均自迴歸模型(Autoregressive Integrated Moving Average model)，其中，ARIMA(p, d, q)可以拆成 AR 及 MA 兩個部分。

- AR(Autoregressive)：自迴歸，p 為自迴歸項數，表示對 p 階做自迴歸。
- MA(Moving Average)：移動平均，q 為移動平均項數，表示對 q 階做移動平均。
- d 表示使時間序列平穩所做的差分次數。

分析流程：

Step1： 觀察原始時間序列圖，判斷平均數、變異數是否平穩，若平均數不平穩需做一次差分；變異數不平穩需做變數轉換。

Step2： 觀察序列自相關圖 ACF 圖(Autocorrelation function)與偏自相關圖 PACF 圖(Partial Autocorrelation function)配適 AR 或 MA 模型。

Step3： 檢查參數估計是否顯著。

Step4： 診斷殘差，檢定白噪音及單根是否符合假設。

Step5： 用最適合模型的預測值來計算預測誤差。

3. 時間序列迴歸(Time Series Regression)：

時間序列迴歸之假設：

- 自變數與應變數有線性關係，可以由過去資料來推斷未來發展趨勢。
- 事件發展具隨機性，故預測值與線性模型的偏差可能與任何因素的發生有關。

時間序列迴歸是線性趨勢(Linear Trend)加上季節因子(Seasonal)所組成的基本模型：

$$y_t = Tr_t + Sn_t + \varepsilon_t ,$$

$$\varepsilon_t \sim iid N(0, \sigma^2)$$

其中，趨勢項配適線性迴歸 $Tr_t = \beta_0 + \beta_1 t$ 並設置季節性虛擬變數 $Sn_t = \{D_{1t} \dots D_{11t}\}$ 。

4. 指數平滑法(Exponential Smoothing Method)：

指數平滑法是一種時間序列分析的方法。其中分為：簡單平滑法(Simple Smoothing)、二次平滑法(Doble Smoothing)、季節性平滑法(Seasonal Smoothing)、線性平滑法(Linear Smoothing)、衰減趨勢平滑法(Damped-Trend Smoothing)、Holt-Winters-Additive、Holt-Winters-Multiplicative。由於指數平滑法為實務經驗累積所得出的分析法，因此統計檢定的理論不適用於評估參數估計是否顯著，僅以預測準則來檢驗預測結果的優劣。

- Holt-Winters-Additive：當季節性的變異程度在時間序列中大致恆定時使用，其等式如下：

$$\begin{aligned}L_t &= \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) , \\b_t &= \gamma(L_t - L_{t-1}) + (1 - \gamma)b_{t-1} , \\S_t &= \delta(Y_t - L_t) + (1 - \delta)S_{t-s} , \\F_{t+m} &= (L_t + b_t m) + S_{t-s+m}\end{aligned}$$

L_t ：The level, b_t ：The growth rate

S_t ：The seasonal factor of the time series in period t

- Holt-Winters-Multiplicative：當季節性的變異程度越來越大時使用，不需要做變數變換，其等式如下：

$$\begin{aligned}L_t &= \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}) , \\b_t &= \gamma(L_t - L_{t-1}) + (1 - \gamma)b_{t-1} , \\S_t &= \delta \frac{Y_t}{L_t} + (1 - \delta)S_{t-s} , \\F_{t+m} &= (L_t + b_t m)S_{t-s+m}\end{aligned}$$

L_t ：The level, b_t ：The growth rate

S_t ：The seasonal factor of the time series in period t

第二節、檢定方法及預測準則

1. 殘差自相關檢定(Durbin-Watson)：

主要檢測迴歸分析中殘差項是否存在自相關，其公式如下：

$$D = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2},$$

殘差自相關檢定分別檢定自相關是否在顯著水準為正與負，其檢定如下：

$$H_0 : \rho_1 = 0,$$

$$H_a : \rho_1 > 0$$

若正的自相關檢定 D 小於 $d_{L,a}$ ，表示拒絕虛無假設；若正的自相關檢定 D 大於 $d_{U,a}$ 時，表示不拒絕虛無假設；若負的自相關檢定 D 介於 $d_{L,a}$ 與 $d_{U,a}$ ，則無法定論。

$$H_0 : \rho_1 = 0,$$

$$H_a : \rho_1 < 0$$

若負的自相關檢定 $(4 - D)$ 小於 $d_{L,a}$ ，表示拒絕虛無假設；若負的自相關檢定 $(4 - D)$ 大於 $d_{U,a}$ 時，表示不拒絕虛無假設；若負的自相關檢定 $(4 - D)$ 介於 $d_{L,a}$ 與 $d_{U,a}$ ，則無法定論。

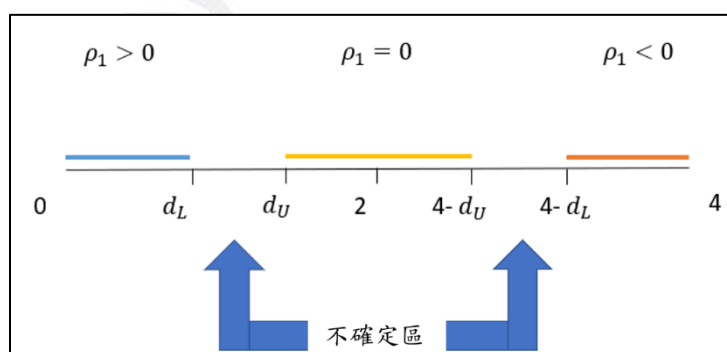


圖 1、殘差的自相關檢定範圍

由圖 1 可知，當檢定統計量結果接近 0，有負自相關；當檢定統計量結果接近 4，有正自相關；當檢定統計量結果接近 2，則為無自相關。

2. 四種預測準則：

為了解模型的預測值精確度，我們將使用 MAD(Mean Absolute Deviation)、MSE(Mean Square Error)、MPE(Mean Percentage Error)、MAPE(Mean Absolute Percentage Error)四種預測準則來衡量模型的優劣，藉由預測誤差的結果，了解預測模型所產生的模擬值與歷史資料之實際值的吻合程度。

- $MAD = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}$
- $MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}$
- $MPE = \frac{\sum_{t=1}^n PE_t}{n}$, $PE_t = \frac{y_t - \hat{y}_t}{y_t} \times 100$
- $MAPE = \frac{\sum_{t=1}^n APE_t}{n}$, $APE_t = \frac{|y_t - \hat{y}_t|}{y_t} \times 100$

3. 白噪音(White Noise)：

主要檢測 ARIMA 模型之序列及殘差是否合適，若無法通過白噪音檢定則表示該 ARIMA 模型之模式不適合，需再進行調整，直至通過白噪音檢定為止。

白噪音三種假設：

$$\begin{aligned} E(a_t) &= 0 , \\ var(a_t) &= \sigma^2 , \\ cov(a_i, a_j) &= 0 \quad \forall i \neq j \end{aligned}$$

4. 單根檢定(Unit root test)：

單根檢定的目的為檢驗時間序列是否平穩，若無法通過單根檢定，表示該時間序列不平穩，需再進行調整，直至通過單根檢定為止。

單根檢定之假設：

$$\begin{aligned} H_0 &: \text{Nonstationary or Unit Root ,} \\ H_a &: \text{Stationary} \end{aligned}$$

第三章、預測分析

第一節、資料敘述

我們的資料來源是採用 kaggle 數據建模和數據分析競賽平臺上的資料，該資料是關於桃園國際機場在 2008 年 9 月 1 日至 2019 年 8 月 31 日的每月進出機場旅客人數紀錄，單位為(萬)人次。

由於 2020 新冠肺炎的疫情爆發，桃園國際機場的進出旅客人數急遽下降，該突發事件並非常態性的規律事件，因此為了準確地預測，我們的資料只取 2008 年 9 月至 2019 年 8 月共 120 筆觀察值，建立模型及預測分析，並保留 2018 年 9 月至 2019 年 8 月共 12 筆觀察值，作為參考的原始值，以利計算模型的預測誤差，總計共 132 筆觀察值。

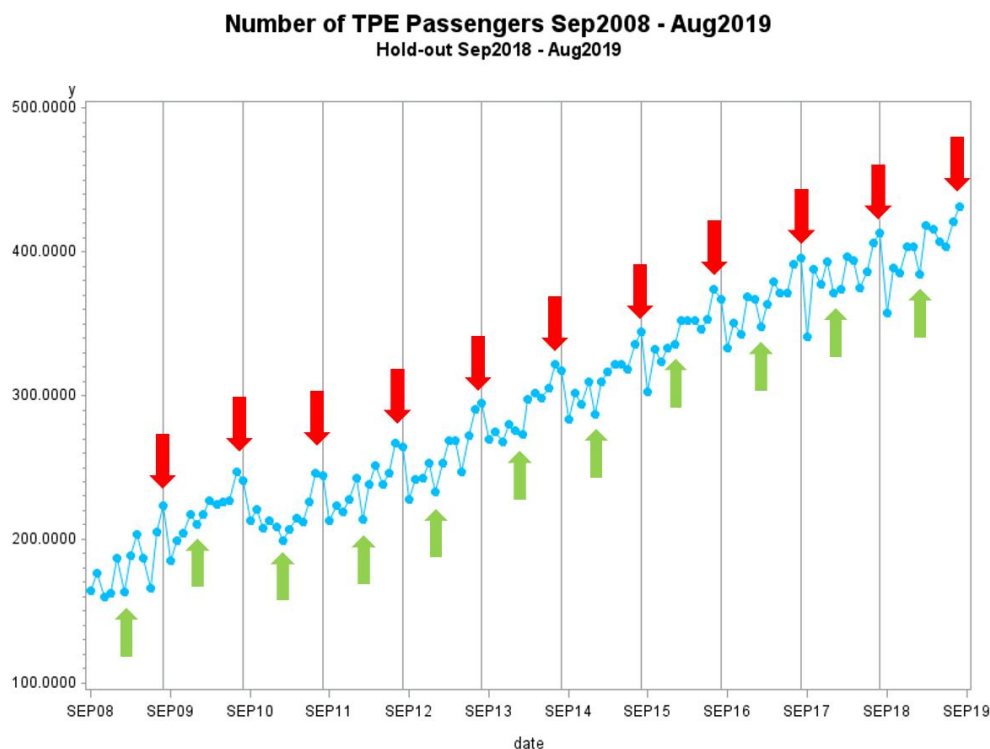


圖 2、桃園國際機場進出旅客人數-原始資料

為了方便觀察，在圖 2 每年的八月畫一條參考線輔助，可以發現七八月是出國的高峰期，相反的，一二月出國人數為低峰期。其原因可能為，七八月為大學以下學生之暑假，因此，多數家庭會規畫出國旅遊，造成機場進出人數暴增；然而，儘管一二月正值寒假期間，機場進出人數卻驟降，因此推論其原因為該期間正逢農曆過年，而為了驗證這項推論，我們調出每年農曆過年的日期，確實發現其月份與低峰月份吻合。

第二節、分解法

由於桃園國際機場進出旅客人數有明顯的季節波動，且會隨著時間序列的水平增長而大幅度的變異，因此分解法採用乘法模型：

$$y_t = Tr_t \times Cl_t \times Sn_t \times Ir_t$$

本研究是採用統計工具 SAS(Statistical Analysis Software)的 X11 方法來進行分解法，該方法源自於美國人口普查局 the US Census Bureau 和加拿大統計局 and Statistics Canada。其乘法模型並非典型的將資料分解為趨勢項、循環項、季節項和不規則項四個部分，而是將趨勢項與循環項合併為趨勢循環項，將資料分解為趨勢循環項、季節項和不規則項三個部分，其模型為：

$$y_t = Trc_t \times Sn_t \times Ir_t$$

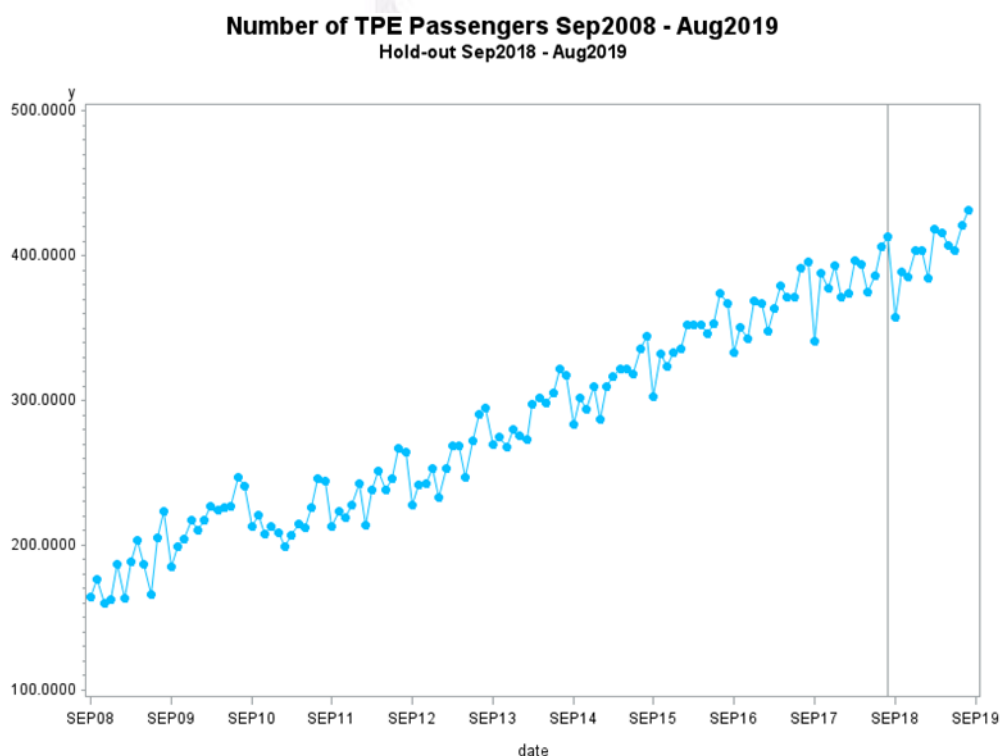


圖 3、分解法-原始時間序列圖

觀察圖 3，發現每一年有明顯的週期循環，其平均數不穩定，變異數穩定。

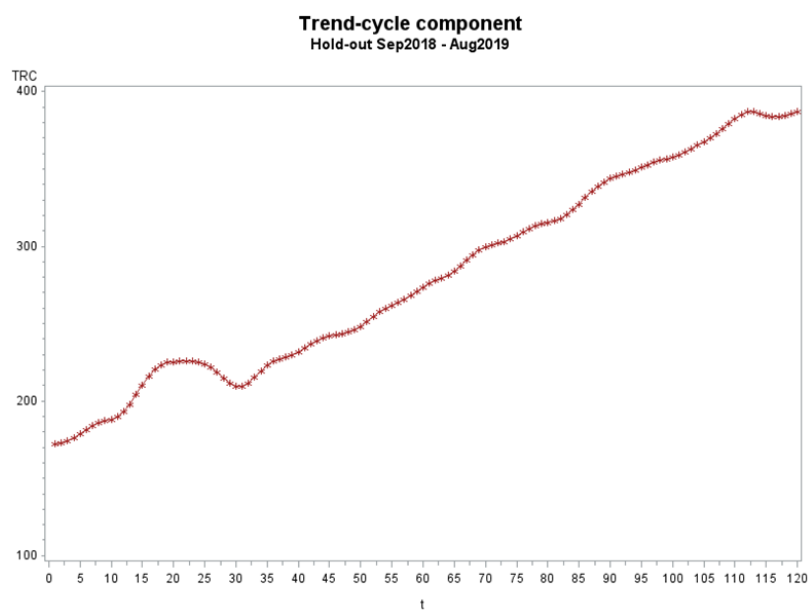


圖 4、分解法-趨勢循環圖

觀察圖 4，發現觀察值確實有線性關係。

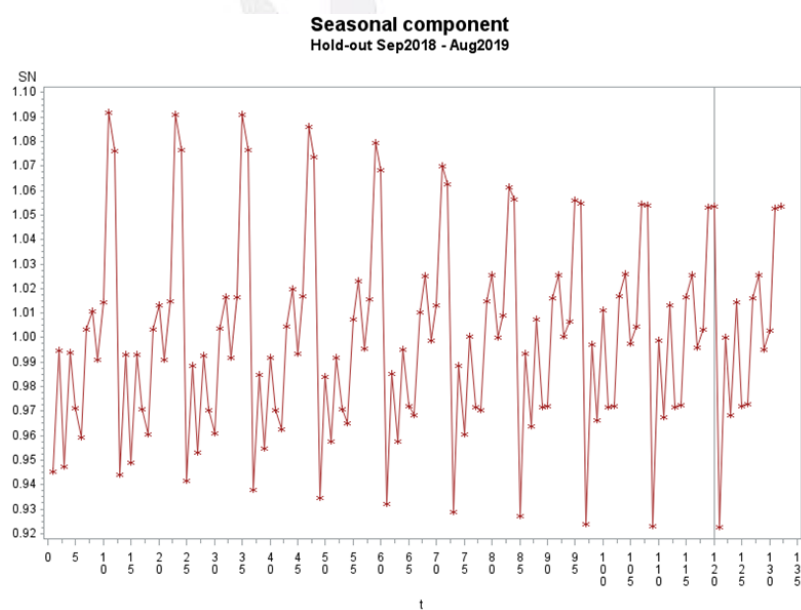


圖 5、分解法-季節性波動圖

觀察圖 5，發現 12 筆觀察值就一個循環，代表此資料有季節性。

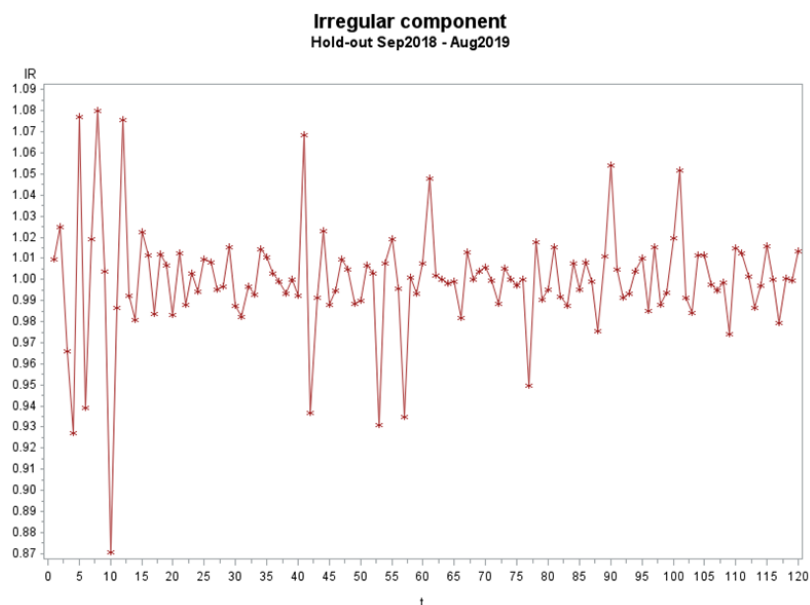


圖 6、分解法-不規則變化圖

觀察圖 6，觀察值可能因任何因素影響，呈現不規則的狀態。

在原始觀察值 y_t 分解為圖 4、圖 5 和圖 6 項三個部分後，針對趨勢循環項去季節化 (deseasonalized)，成為去季節化觀察值 dy_t 。由於季節性這種重複的循環，可能會模糊我們在預測時希望建模的信號，所以需在去除季節性之後，方可對去季節化觀察值 dy_t 配適迴歸模型。

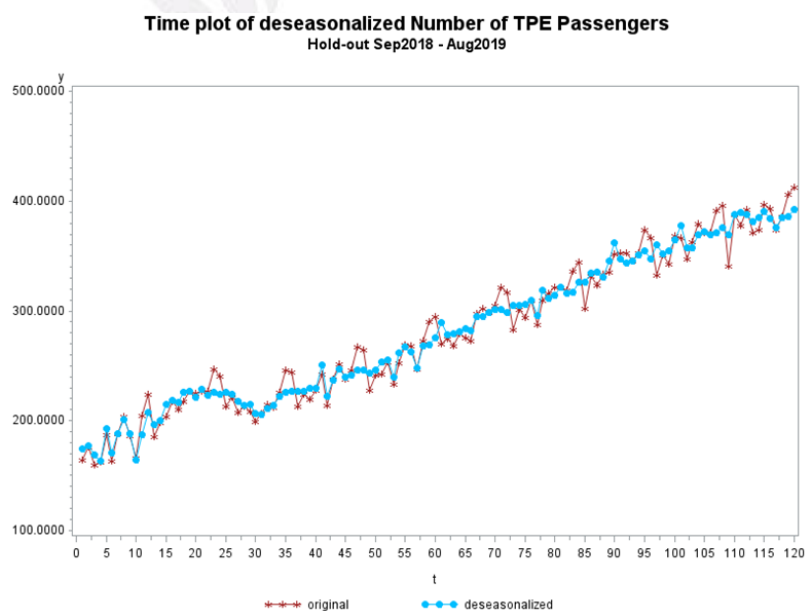


圖 7、分解法-去季節化觀察值圖

觀察圖 7，發現去季節化觀察值 dy_t 與時間有線性關係，故對其配適迴歸模型，並且檢定模型的殘差之間是否有自相關。

Durbin-Watson Test

正自相關假設

$$H_0: \rho = 0$$

$$H_a: \rho > 0$$

負自相關假設

$$H_0: \rho = 0$$

$$H_a: \rho < 0$$

修改前

表 2、分解法- Durbin-Watson 統計值(修改前)

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	0.7780	<.0001	1.0000

正自相關檢定的 p-value < 0.0001 小於 5%顯著水準，故拒絕 H_0 ，表示殘差之間有正自相關，需配適一階殘差，檢定是否有自相關。

修改後

表 3、分解法- Durbin-Watson 統計值(修改後)

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	2.2200	0.8682	0.1318

正自相關與負自相關的 p-value 皆大於 5%顯著水準，故不拒絕 H_0 ，表示一階殘差之間沒有正自相關與負自相關。

參數估計值

表 4、分解法-自迴歸參數估計值

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.609946	0.073261	-8.33

表 5、分解法-參數估計值

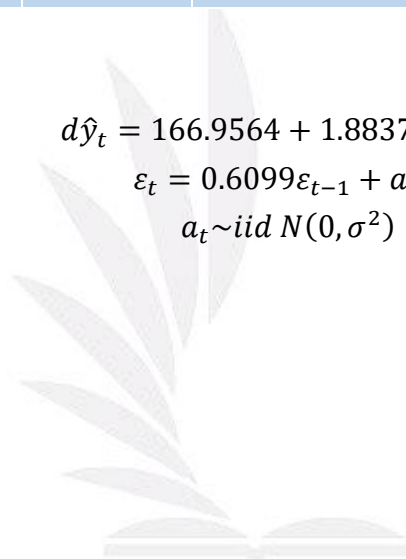
Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	166.9564	4.0494	41.23	<.0001
t	1	1.8837	0.0577	32.64	<.0001

Model :

$$d\hat{y}_t = 166.9564 + 1.8837t + \varepsilon_t ,$$

$$\varepsilon_t = 0.6099\varepsilon_{t-1} + a_t ,$$

$$a_t \sim iid N(0, \sigma^2)$$



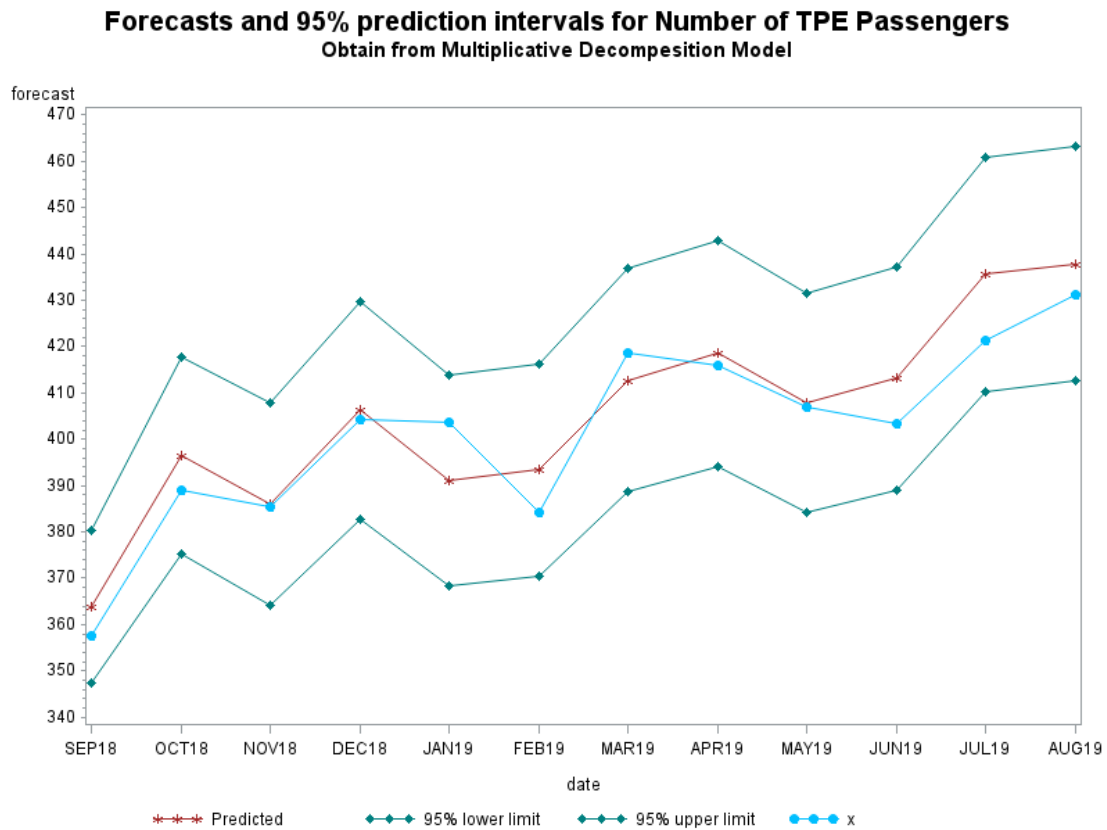


圖 8、分解法-樣本外預測圖

為了解模型的預測值模式與保留年原始值模式的相符程度，因此繪製圖 8 樣本外預測圖來比較。由圖 8 可知，預測值在 2019 年 1 月有明顯低於實際值的情形，以及在 2019 年 7 月有明顯高過實際值的情形。其原因推測為在 2019 年 6 月下旬時發生了空服員罷工事件，此一事件可能導致桃園機場 7 月份的進出人數降低，造成預測值有高估的情況，而其他月份的預測值皆與實際值相近，且 12 筆實際值和預測值皆位於 95% 的預測區間內，表示分解法配適的模型對於預測桃園機場進出旅客人數具有相當的精確度。

為了解分解法模型之預測值與實際值的偏離的程度，需藉由預測誤差來判斷模型的優劣。表 6 為四種預測準則 MAD、MSE、MPE、MAPE 所計算的預測誤差之平均數報表。

表 6、分解法-樣本外預測之 MAD、MSE、MPE、MAPE

Variable	N	Mean
MAD	12	6.506739
MSE	12	60.48708
MPE	12	-0.870499
MAPE	12	1.617469

第三節、Box-Jenkins ARIMA 模型

根據 Box-Jenkins ARIMA 假設，時間序列若是平穩的，可以使用 ARMA 模型來近似；時間序列若是不平穩的，則使用 ARIMA 模型來近似。因此，需觀察序列的自相關圖(ACF 圖)與偏自相關圖(PACF 圖)配適模型。

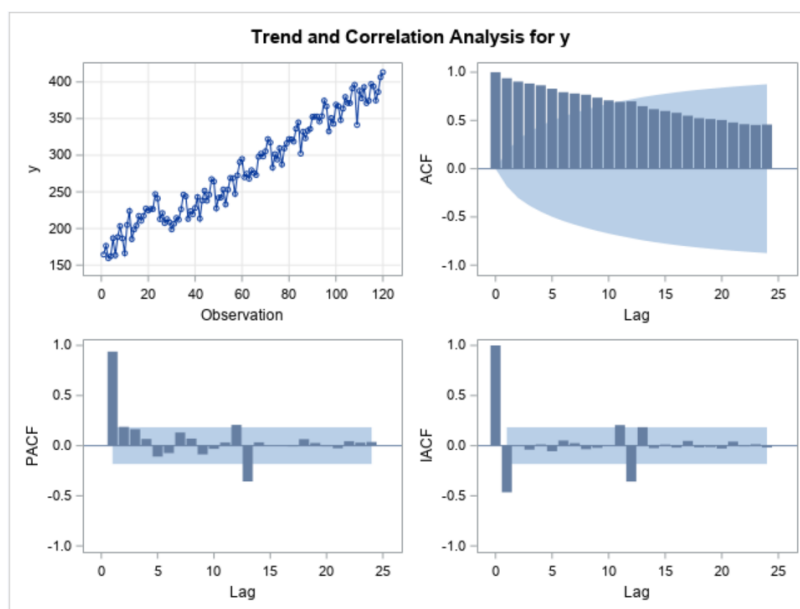


圖 9、ARIMA-時間序列、ACF、PACF 與 IACF

觀察圖 9 的時間序列，其平均數不穩定，變異數穩定，且 ACF 圖 Dies down slowly，故需做一次差分確認序列是否達到平穩。

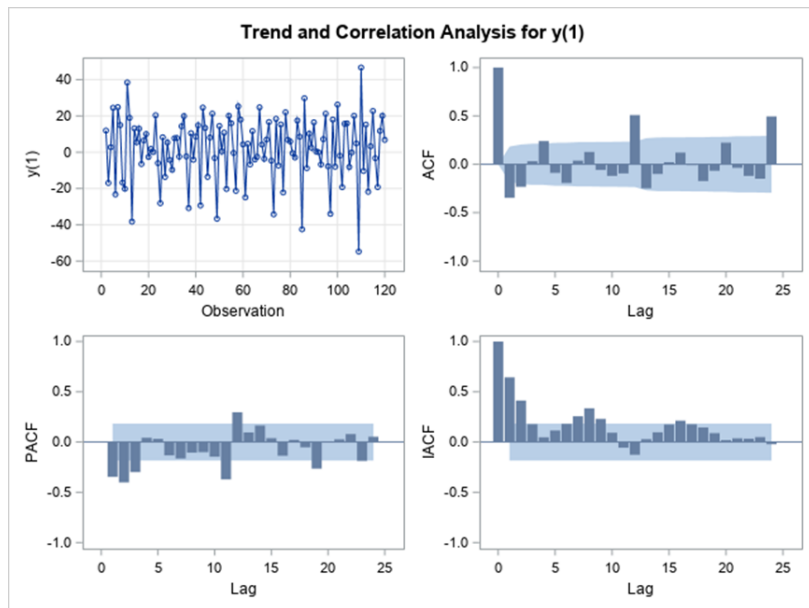


圖 10、ARIMA-一次差分後的 ACF、PACF 與 IACF

經過一次差分後，發現圖 10 的 ACF 圖的 Lag12 與 Lag24 突出，表示序列有季節性，需做一次季節性差分，確認序列是否達到平穩。

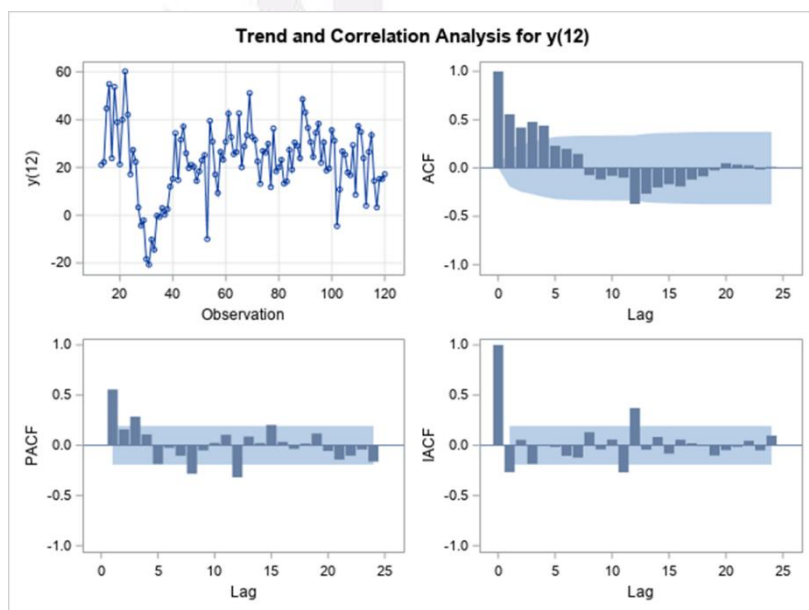


圖 11、ARIMA-一次季節差分後的 ACF、PACF 與 IACF

經過一次季節性差分後，發現圖 11 的 ACF 圖 Dies down slowly 尚未平穩，需改做一次差分及一次季節性差分，確認序列是否達到平穩。

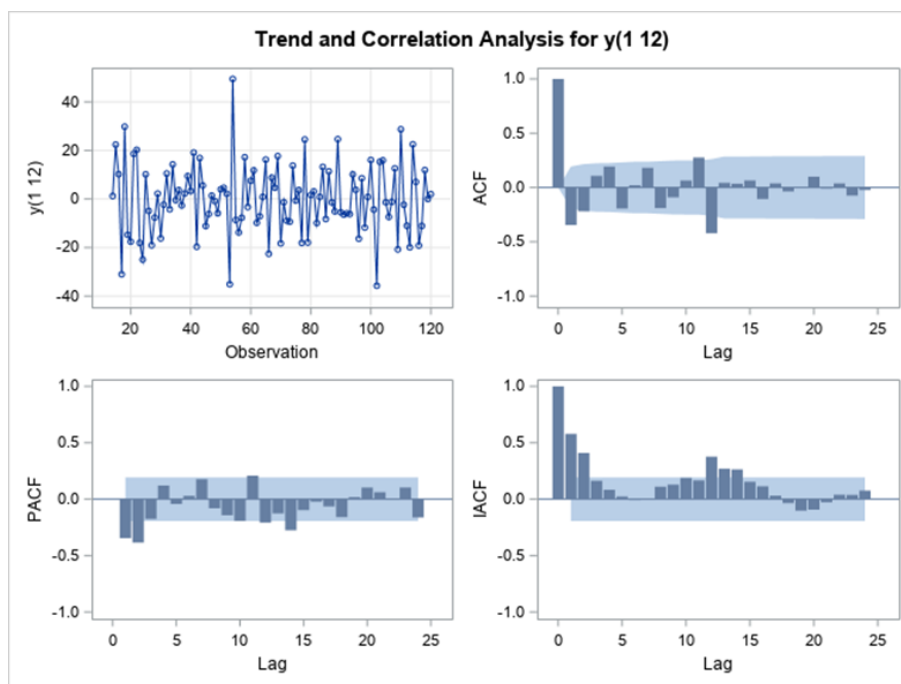


圖 12、ARIMA-一次差分與一次季節差分後的 ACF、PACF 與 IACF

經過一次差分與一次季節性差分後，圖 12 的平均數及變異數都變得穩定，且 ACF 圖呈現 Cuts off，且在 Lag1 與 Lag12 突出，而 PACF 圖呈現 Dies down，這種型態的資料適合配適 MA 模型。

表 7 (a)、ARIMA-參數估計表

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.60937	0.07292	8.36	<.0001	1
MA2,1	0.82865	0.14107	5.87	<.0001	12

表 7 (b)、ARIMA-參數估計表

Variance Estimate	96.12936
Std Error Estimate	9.804558
Number of Residuals	107

表 7(a)為配適 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 模型之參數估計，模型如下：

Model : $ARIMA(0,1,1) \times (0,1,1)_{12}$

$$(1 - B)(1 - B^{12})z_t = (1 - 0.6094B)(1 - 0.8287B^{12})a_t ,$$

$$a_t \sim iid N(0, \sigma^2) ,$$

$$\hat{\sigma}^2 = 96.1294$$

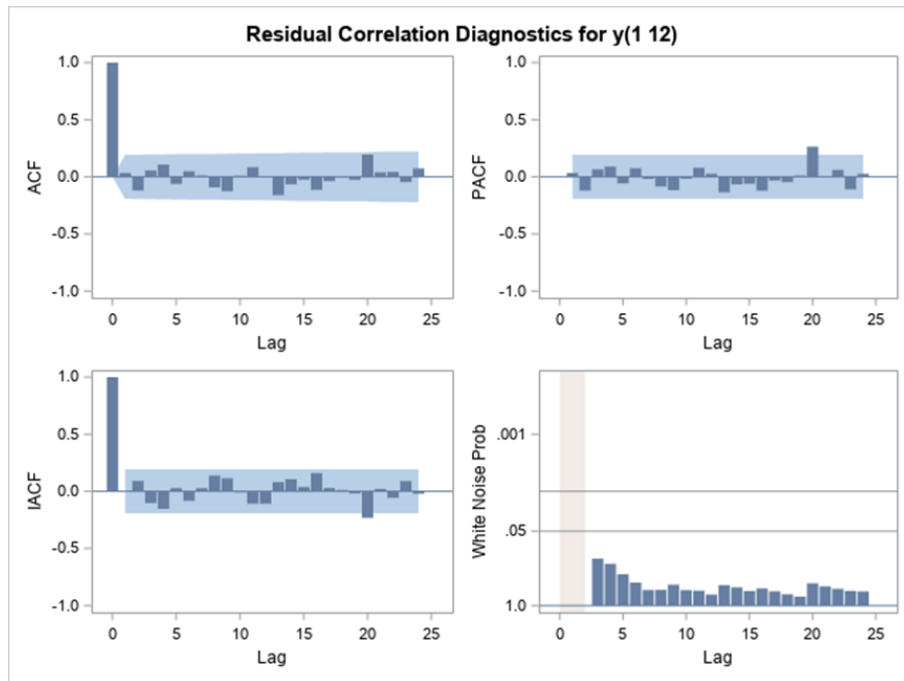


圖 13、ARIMA-配適完後的 ACF、PACF、IACF 與 White Noise

經過配適 $ARIMA(0,1,1) \times (0,1,1)_{12}$ ，觀察圖 13 的 ACF 圖及 PACF 圖發現序列的自相關及偏自相關幾乎都落在兩倍標準差之內，僅有 Lag20 稍有突出，需做白噪音檢定 (White Noise Test) 進一步確認模型的殘差之間無自相關。

根據圖 13 的白噪音檢定圖，檢定統計量的 p-value 均大於 5% 顯著水準，故不拒絕 H_0 ，表示殘差之間無自相關，因此該模型為合適的。

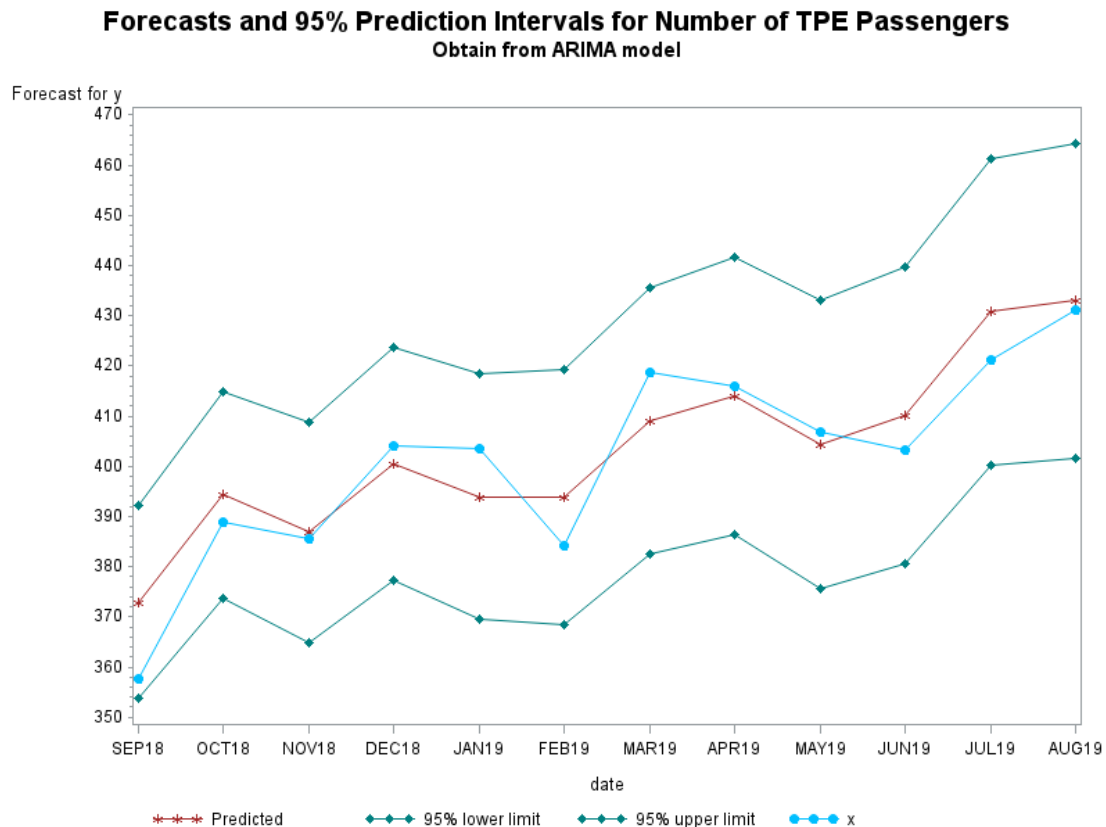


圖 14、ARIMA-樣本外預測圖

為了解模型的預測值模式與保留年原始值模式的相符程度，因此繪製圖 14 樣本外預測圖來比較。由圖 14 可知，預測值在 2018 年 9 月及 2019 年 2 月有明顯高於實際值的情形，以及在 2019 年 1 月及 2019 年 3 月有明顯低於實際值的情形，其餘月份的預測值皆和實際值相近，且 12 筆實際值和預測值皆位於 95% 的預測區間內，表示 ARIMA 配適的模型對於預測桃園機場進出旅客人數具有相當的精確度。

為了解 ARIMA 模型之預測值與實際值的偏離的程度，需藉由預測誤差來判斷模型的優劣。表 8 為四種預測準則 MAD、MSE、MPE、MAPE 所計算的預測誤差之平均數報表。

表 8、ARIMA-樣本外預測之 MAD、MSE、MPE、MAPE

Variable	N	Mean
MAD	12	6.389589
MSE	12	58.62307
MPE	12	-0.521547
MAPE	12	1.620391

第四節、時間序列迴歸

時間序列迴歸是線性趨勢(Linear Trend)加上季節因子(Seasonal)所組成的基本模型：

$$y_t = Tr_t + Sn_t + \varepsilon_t \quad ,$$

$$\varepsilon_t \sim iid N(0, \sigma^2)$$

其中，趨勢項配適線性迴歸 $Tr_t = \beta_0 + \beta_1 t$ 並設置季節性虛擬變數 $Sn_t = \{D_{1t} \dots D_{11t}\}$ ，模型如下：

$$y_t = \beta_0 + \beta_1 t + \beta_{01} D_{1t} + \beta_{02} D_{2t} + \beta_{03} D_{3t} \dots + \beta_{11} D_{11t} + \varepsilon_t \quad ,$$

$$\varepsilon_t \sim iid N(0, \sigma^2)$$

D1 至 D11：

表 9、12 個月份虛擬變數

月份	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
一月	1	0	0	0	0	0	0	0	0	0	0
二月	0	1	0	0	0	0	0	0	0	0	0
三月	0	0	1	0	0	0	0	0	0	0	0
四月	0	0	0	1	0	0	0	0	0	0	0
五月	0	0	0	0	1	0	0	0	0	0	0
六月	0	0	0	0	0	1	0	0	0	0	0
七月	0	0	0	0	0	0	1	0	0	0	0
八月	0	0	0	0	0	0	0	1	0	0	0
九月	0	0	0	0	0	0	0	0	1	0	0
十月	0	0	0	0	0	0	0	0	0	1	0
十一月	0	0	0	0	0	0	0	0	0	0	1
十二月	0	0	0	0	0	0	0	0	0	0	0

表 10、時間序列迴歸-參數估計值

PARM	Value	Std.Error	T	Prob > T
Intercept	167.9073	6.7979	24.6998	<.0001
Lag 1	0.4013	0.0940	4.2690	<.0001
Lag 2	0.0994	0.1017	0.9775	0.3306
Lag 3	0.2671	0.0942	2.8373	0.0055
Liner Trend	1.8641	0.0889	20.9758	<.0001
d1	-5.7529	3.3840	-1.7000	0.0921
d2	-9.1195	3.7304	-2.4447	0.0162
d3	4.4349	3.5512	1.2489	0.2145
d4	8.1979	3.8811	2.1123	0.0371
d5	-2.6486	3.9982	-0.6625	0.5091
d6	0.7128	3.9437	0.1807	0.8569
d7	20.1333	4.0058	5.0260	<.0001
d8	20.2300	3.9066	5.1784	<.0001
d9	-17.0021	3.5467	-4.7938	<.0001
d10	-1.3153	3.7272	-0.3529	0.7249
d11	-10.0655	3.3824	-2.9758	0.0036
Model Variance	81.5629	.	.	.

觀察表 10 的 12 個月虛擬變數的參數估計值，發現 d8(八月)的參數估計值 $\hat{\beta}_{08} = 20.2328$ ，為 12 個月中最大的係數，表示八月時桃園機場的旅客進出人數最多；d7(七月)的參數估計值 $\hat{\beta}_{07} = 20.1333$ ，為 12 個月中第二大的係數，表示七月時依然是旅客進出機場的高峰期。相反的，d9(九月)的參數估計值 $\hat{\beta}_{09} = -17.0021$ ，為 12 個月中最小的係數，表示九月旅客進出機場的人數最少。

Model :

$$\hat{y}_t = 167.9073 + 1.8641t - 5.7529D_{1t} - 9.1195D_{2t} + 4.4349D_{3t} + 8.1979D_{4t} - 2.6486D_{5t} + 0.7128D_{6t} + 20.1333D_{7t} + 20.2300D_{8t} - 17.0021D_{9t} - 1.3153D_{10t} - 10.0655D_{11t} + \varepsilon_t ,$$

$$\varepsilon_t = 0.4013\varepsilon_{t-1} + 0.0994\varepsilon_{t-2} + 0.2671\varepsilon_{t-3} + a_t ,$$

$$a_t \sim iid N(0, \sigma^2) ,$$

$$\hat{\sigma}^2 = 96.1294$$

為了確保模型的殘差並無自相關及模型序列的平穩，需進一步做白噪音檢定(White Noise Test)及單根檢定(Unit Root Test)確認。

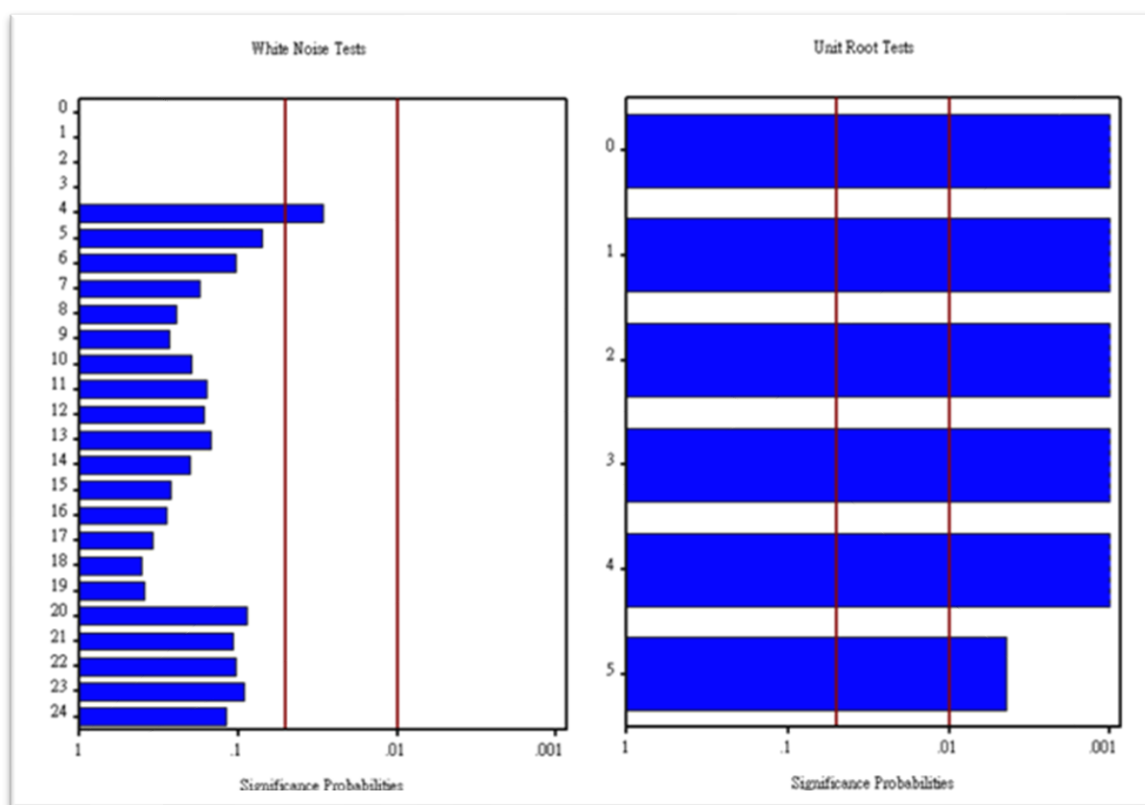


圖 15、時間序列迴歸-白噪音與單根檢定

根據圖 15 的白噪音檢定(White Noise Test)，檢定統計量的 p-value 均 $> 1\%$ 顯著水準，故不拒絕 H_0 ，表示模型的殘差無自相關；根據圖 15 的單根檢定(Unit Root Test)，檢定統計量的 p-value 均 $< 1\%$ 顯著水準，故拒絕 H_0 ，表示模型的序列是平穩的，故該模型是合適的。

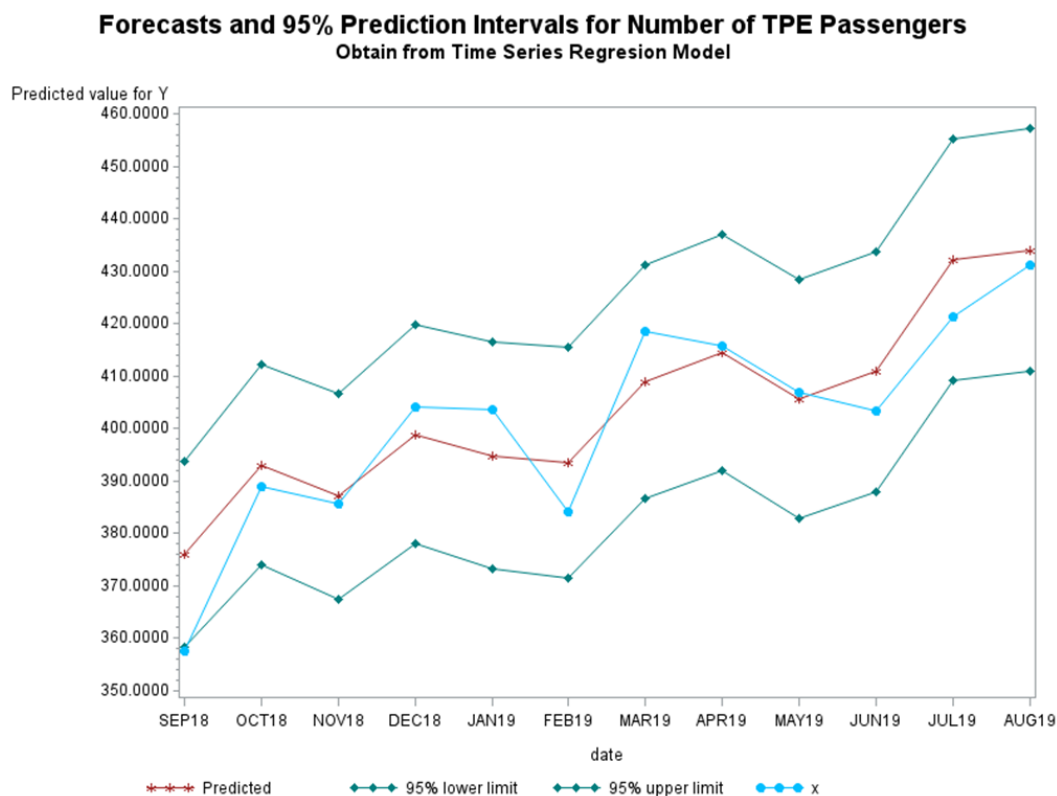


圖 16、時間序列迴歸-樣本外預測圖

為了解模型的預測值模式與保留年原始值模式的相符程度，因此繪製圖 16 樣本外預測圖來比較。由圖 16 可知，預測值在 2019 年 2 月有明顯低於實際值的情形，以及在 2019 年 1 月及 2019 年 3 月有明顯高於實際值的情形，其餘月份的預測值皆和實際值相近。值得注意的是，2018 年 9 月的實際值幾乎落在 95% 預測區間的外緣，其原因為強烈颱風山竹在 2018 年 9 月上旬形成，外圍環流造成出現驚人雨勢，導致航班暫停，所以九月的出國人數不如預期。除了該月以外，其餘 11 筆實際值和預測值皆位於 95% 的預測區間內，故時間序列迴歸模型對於預測桃園機場進出旅客人數具有相當的精確度。

為了解時間序列迴歸模型之預測值與實際值的偏離的程度，需藉由預測誤差來判斷模型的優劣。表 11 為四種預測準則 MAD、MSE、MPE、MAPE 所計算的預測誤差之平均數報表。

表 11、時間序列迴歸-樣本外預測之 MAD、MSE、MPE、MAPE

Variable	N	Mean
MAD	12	6.7412422
MSE	12	69.229574
MPE	12	-0.6459356
MAPE	12	1.7123276

第五節、指數平滑法

指數平滑法是另一種時間序列分析常見的方法，根據圖 5 分解法-季節性波動圖，得知序列的季節性的變異程度隨著時間改變而越來越大，因此較宜採用 Holt-Winters-Multiplicative 的指數平滑法，其方法之等式如下：

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}) ,$$

$$b_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)b_{t-1} ,$$

$$S_t = \delta \frac{Y_t}{L_t} + (1 - \delta)S_{t-s} ,$$

$$F_{t+m} = (L_t + b_t m)S_{t-s+m}$$

L_t : The level, b_t : The growth rate

S_t : The seasonal factor of the time series in period t

表 12、Holt-Winters Multiplicative 指數平滑法參數估計表

Obs	PARAM	VALUE	STDERR	T	P
1	LEVEL Smoothing Weight	0.35547	0.0481	7.3971	<.0001
2	TREND Smoothing Weight	0.00100	0.0217	0.0461	0.9633
3	SEASONAL Smoothing Weight	0.31309	0.0586	5.3403	<.0001
4	Residual Variance (sigma squared)	96.14759	.	.	.
5	Smoothed Level	388.86848	.	.	.
6	Smoothed Trend	1.88516	.	.	.
7	Smoothed Seasonal Factor 1	0.96985	.	.	.
8	Smoothed Seasonal Factor 2	0.96282	.	.	.
9	Smoothed Seasonal Factor 3	1.00513	.	.	.
10	Smoothed Seasonal Factor 4	1.01816	.	.	.
11	Smoothed Seasonal Factor 5	0.98441	.	.	.
12	Smoothed Seasonal Factor 6	0.99729	.	.	.
13	Smoothed Seasonal Factor 7	1.05834	.	.	.
14	Smoothed Seasonal Factor 8	1.06485	.	.	.
15	Smoothed Seasonal Factor 9	0.93700	.	.	.
16	Smoothed Seasonal Factor 10	1.00348	.	.	.
17	Smoothed Seasonal Factor 11	0.96674	.	.	.
18	Smoothed Seasonal Factor 12	0.99958	.	.	.

根據表 12、Holt-Winters Multiplicative 指數平滑法參數估計表，7 月及 8 月的季節因子參數估計值為 1.0583 及 1.0649，是一年當中最高的兩個數值，表示此兩月份的桃園機場旅客人數最高；十月的季節因子參數估計值為 1.0035，是一年當中最底的數值，表示此月份的桃園機場旅客人數最低；一月及二月由於受到農曆春節的影響，因此季節因子參數估計值也較低。其等式如下：

$$\text{Level: } L_t = 0.3555 \frac{Y_t}{S_{t-s}} + (1 - 0.3555)(L_{t-1} + b_{t-1}) ,$$

$$\text{Trend: } b_t = 0.001(L_t - L_{t-1}) + (1 - 0.001)b_{t-1} ,$$

$$\text{Seasonal: } S_t = 0.3131 \frac{Y_t}{L_t} + (1 - 0.3131)S_{t-s} ,$$

$$\text{Forecast: } F_{t+m} = (L_t + b_t m)S_{t-s+m}$$

Forecasts and 95% Prediction Intervals for Number of TPE Passengers
Obtain from Exponential Smoothing

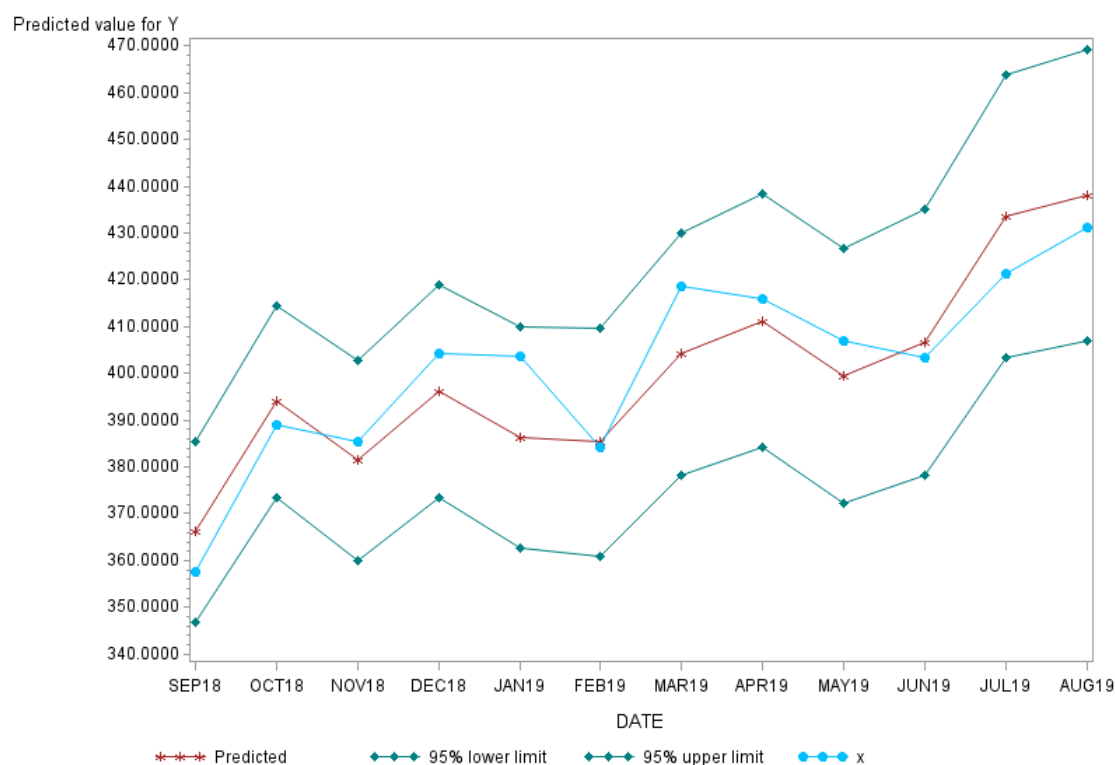


圖 17、Holt-Winters Multiplicative-樣本外預測圖

為了解指數平滑法的預測值模式與保留年原始值模式的相符程度，因此繪製圖 17 Holt-Winters Multiplicative-樣本外預測圖來比較。由圖 17 可知，預測值在 2019 年 1 月及 2019 年 3 月有明顯低於實際值的情形，而預測值在 2019 年 7 月和 2019 年 8 月則是有略高於實際值的情形，除了前述所提及的月份外其餘月份的預測值皆和實際值相近，這 12 筆實際值和預測值皆位於 95% 的預測區間內，表示指數平滑法對於預測桃園機場進出旅客人數具有相當的精確度。

為了解指數平滑法之預測值與實際值的偏離的程度，需藉由預測誤差來判斷模型的優劣。表 13 為四種預測準則 MAD、MSE、MPE、MAPE 所計算的預測誤差之平均數報表。

表 13、Holt-Winters Multiplicative 樣本外預測之 MAD、MSE、MPE、MAPE

Variable	N	Mean
MAD	12	7.7287
MSE	12	80.8058
MPE	12	0.3667
MAPE	12	1.9119



第四章、結論

一、時間序列分析法比較

為了解桃園機場旅客進出人數的模式，本研究使用四項時間序列分析方法進行分析，依序為分解法(X11)、Box-Jenkins ARIMA、時間序列迴歸和指數平滑法，而最終判斷模型預測值的主臬則為 MAD、MSE、MPE 以及 MAPE 這四項預測準則，藉由預測準則判斷模型的優劣。

表 14、樣本外預測之 MAD、MSE、MPE、MAPE 比較表

	X11	ARIMA	Time Series Regression	Holt-Winters Multiplicative	準則	The Best Model
MAD	6.5067	6.3896	6.7412	7.7287	越小越好	ARIMA
MSE	60.4871	58.6231	69.2296	80.8058	越小越好	ARIMA
MPE	-0.8705%	-0.5215%	-0.6459%	0.3667%	越接近 0 越好	Holt-Winters Multiplicative
MAPE	1.6175%	1.6204%	1.7123%	1.9119	越小越好	X11

表 14 為四種分析法的預測誤差比較報表。由表 14 可知，MAD 的判斷準則為該數值越小越好，因此 ARIMA 是四種分析法之中最好的模型；MSE 的判斷準則為該數值越小越好，因此 ARIMA 亦是四種分析法之中最好的模型；MPE 的判斷準則為越接近 0 越好，因此指數平滑法 Holt-Winters Multiplicative 是四種分析法之中最好的模型；MAPE 的判斷準則為該數值越小越好，因此 X11 分解法是四種分析法之中最好的模型為。

綜合四項預測準則的結果，Box-Jenkins ARIMA 為預測桃園機場旅客進出人數的最佳方法，但由於四種分析法的預測誤差數值都很小也很相近，因此四種分析法都具有相當的參考價值。

二、預測值與實際值偏離原因

根據四種時間序列分析法之樣本外預測圖，預測值在 2018 年 9 月與 2019 年 7 月都有高於實際值的情形。其原因為，從歷年颱風與社會事件來分析，在 2018 年 9 月上旬山竹颱風形成，出現驚人雨勢，導致航班暫停，所以 9 月份桃園機場的進出人數不如預期。而在 2019 年 6 月下旬，發生了空服員罷工事件，此一事件可能導致桃園機場 7 月份的進出人數降低，造成預測值有高估的情況。因此，氣候不佳與空服員罷工這兩事件為可能造成桃園機場進出人數下降的原因。

三、預測建議與回饋

依據全球飛安統計，超過七成的飛航事故與人為因素有關。我國飛安統計也顯示，事故主因近七成與飛行員有關，若包含其他如維修人員、管制員及組織管理因素，則約有八成涉及人為因素。近年來，機師、空服員發動罷工，其罷工原因都與「超時工作」有關，代表航空業人力資源嚴重不足。飛行安全一直是國人搭乘飛機最重視的問題，適當的人力安排，才能避免機師疲勞駕駛，以減少飛航事故的發生。因此，我們建議在旅客進出人數的高峰期，應增加及調派人力至機場地勤，協助旅客登機及出入境手續，亦增加空服人員輪班次數，以縮短飛行時數。本文預測未來桃園機場旅客進出人數為高準確度，因此航空公司可透過本文預測來擬定因應對策。

除此之外，桃園國際機場作為台灣的門戶，為臺灣航空的交通樞紐，對於台灣觀光產業以及貿易出口的發展有舉足輕重的地位，因此桃園國際機場的航廈數量及機場跑道數量將決定可容納旅客人數及航班的吞吐量，進而影響台灣的觀光業及貿易經濟的發展。根據預測，桃園國際機場的進出旅客人數有逐年攀升趨勢，因此，我們建議桃園國際機場修建第一航廈與第二航廈，加蓋第三航廈以及增加機場跑道，以服務未來增加的旅客流量。



附錄一、工作分配表

組員	工作分配	工作及討論時數	貢獻百分比
張芸菁	Word、ppt、SAS 程式撰寫、 口頭報告、規劃及統整	>24 小時	40%
柯雅潔	Word、SAS 程式撰寫、口頭報 告	24 小時	20%
呂嘉侑	Word、口頭報告	24 小時	15%
謝宇凡	Word、口頭報告	20 小時	15%
顏嘉玄	ppt、口頭報告	16 小時	10%



附錄二、參考資料

1. Box-Jenkins Method :

Brownlee, J. (n.d.). A Gentle Introduction to the Box-Jenkins Method for Time Series Forecasting. Time Series. <https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/>

2. 〈單位根檢驗〉:

MBA 智庫百科. (2016, December 21). Retrieved from <https://wiki.mbalib.com/zh-tw/%E5%8D%95%E4%BD%8D%E6%A0%B9%E6%A3%80%E9%AA%8C>

3. 〈時間序列分解總結〉:

(2020, October 24). Retrieved from http://www.360doc.com/content/20/1024/18/99071_942194322.shtml

4. 〈杜賓-瓦森統計量〉:

維基百科. Retrieved from <https://zh.wikipedia.org/wiki/%E6%9D%9C%E5%AE%BE-%E7%93%A6%E7%89%B9%E6%A3%AE%E7%BB%9F%E8%AE%A1%E9%87%8F>

5. Forecasting, Time Series, and

Bowerman, B. L., O'connell, R. T., & Koehler, A. B. (2005). Forecasting, Time Series, and Regression (4th ed.). America: Curt Hinrichs.

6. 〈長榮協商破裂 空服員下午 4 時起罷工 首日影響 22 航班〉中央通訊社:

張雄風. (2019, June 20). Retrieved from <https://www.cna.com.tw/news/firstnews/201906205006.aspx>

7. 〈山竹颱風侵台機率不高，但還是會對台灣造成這些影響〉 yahoo 新聞：

何渝婷. (2018, September 13). Retrieved from

<https://tw.news.yahoo.com/%E5%B1%B1%E7%AB%B9%E9%A2%B1%E9%A2%A8%E4%B%E5%E5%8F%B0%E6%A9%9F%E7%8E%87%E4%B8%8D%E9%AB%98-%E4%BD%86%E9%82%84%E6%98%AF%E6%9C%83%E5%B0%8D%E5%8F%B0%E7%81%A3%E9%80%A0%E6%88%90%E9%80%99%E4%BA%9B%E5%BD%B1%E9%9F%BF-073500803.html>

8. 〈人為因素與飛航安全〉：

鄭永安. (2014). 人為因素與 飛航安全. 科學發展, (495), 1-5.

<http://nccur.lib.nccu.edu.tw/bitstream/140.119/34793/7>

9. 〈白噪音〉，

http://www.fin.kuas.edu.tw/uploads/bulletin_file/file/5ef05c456325c20e79000079/_E8_AC_9B_E7_BE_A92_E6_99_82_E9_96_93_E5_BA_8F_E5_88_97_E6_A8_A1_E5_9E_8B_28CH7.pdf

10. 趙宥寧，2019.01.06，〈長桃園機場旅運再創新高！運量人次躍全球前 50 大 躋身機場界「大聯盟」〉風傳媒

<https://www.cna.com.tw/news/firstnews/201906205006.aspx>