

# 逢甲大學學生報告 ePaper

## 國內航空載客現況與預測

Forecasting of Number of Passengers for Domestic Flights

作者：郭峙延、傅瀚民、楊庭語、黃義郡、林宜靜、趙之瑜、高御宸

系級：統計學系系統與精算碩士班

學號：M0422928、M0418625、M0405006、M0403154、M0419261、M0428908、M0492421

開課老師：陳婉淑 教授

課程名稱：統計預測方法

開課系所：統計學系

開課學年：104 學年度 第 2 學期

## 中文摘要

本報告主要研究每年國內航空載客量，期望藉由過去資料推測未來趨勢。資料來源為主計處總體統計資料庫，資料時間從 2002 年 1 月至 2016 年 3 月，共 171 筆資料，其中我們保留最後十二筆資料作為預測的對照依據。時間序列分析方法使用 ARIMA 分析法、時間序列迴歸法、分解法以及指數平滑法，配適模型並進行預測。本報告採用四種評估準則：MSE (Mean Squared Error)、MAE (Mean Absolute Error)、MAPE (Mean Absolute Percentage Error)，以及 MPE (Mean Percentage Error)，將十二筆預測結果與最後十二筆實際資料作比較，檢驗預估是否準確。結果發現時間序列迴歸法為最佳時間預測方法。



關鍵字：ARIMA 分析、時間序列回歸法、分解法、指數平滑法、樣本外預測、Durbin-Watson Test

## Abstract

The objective of this study is to forecast the monthly number of passengers of Taiwan's domestic flights. We predict 12-step-ahead numbers of passengers based on four forecasting models. Our data source is the macroeconomic database from National Statistics. The duration of data is from January 2002 to March 2016. There are 171 data during the period, and we hold out the last 12 observations for the out-of-sample forecasts. The forecasting methods include ARIMA (Autoregressive Integrated Moving Average) model, time series regression, decomposition methods, and exponential smoothing method. We use four evaluation criteria to examine the accuracy of the forecast, which are MSE (Mean Squared Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and MPE (Mean Percentage Error). It turns out that the time series regression is the best forecasting model based on the four criteria.



**Keyword :** ARIMA, Time Series Regression, Time Series Decomposition method, Exponential Smoothing method, out-of-sample forecasting, Durbin-Watson Test

## 目 次

第一章、	緒論.....	1
一、	研究動機.....	1
二、	資料說明.....	1
第二章、	ARIMA 分析 .....	8
第三章、	時間序列迴歸.....	14
第四章、	分解法.....	19
第五章、	指數平滑法.....	26
第六章、	結論.....	29
一、	時間序列預測法之比較.....	29
二、	時間序列預測之結論.....	29
參考文獻.....		31
附錄一、	國內航線原始資料.....	32

## 表目錄

表 1、原始資料自相關檢定.....	8
表 2、ARIMA 參數顯著檢定 .....	10
表 3、模型參數估計值.....	12
表 4、樣本外 12 筆預測觀測值及其 95%信賴區間上下界 .....	13
表 5、ANOVA 表.....	15
表 6、時間序列迴歸法—Durbin-Watson 統計值.....	16
表 7、時間序列迴歸法—自迴歸參數的估計值.....	16
表 8、時間序列迴歸法—參數估計值.....	17
表 9、參數估計值.....	17
表 10、標準 X11 分解法.....	19
表 11、分解法—調整後 R 平方.....	23
表 12、分解法—Durbin-Watson.....	24
表 13、分解法—自迴歸參數估計值.....	24
表 14、分解法—Durbin-Watson.....	24
表 15、分解法—參數估計值.....	25
表 16、指數平滑法參數估計.....	27
表 17、樣本外預測之 MAD、MSE、MPE、MAPE 比較表.....	29
表 18、國內航線原始資料.....	32

## 圖目錄

圖 1、國內航空客運每月載客人數-原始資料 .....	5
圖 2、台灣高速鐵路每月載客人數(資料來源:主計處總體統計資料庫) .....	6
圖 3、國內航空客運每月載客人數-盒型圖 .....	7
圖 4、國內航空客運每月載客人數-觀察值 .....	8
圖 5、資料之 ACF 與 PACF 圖 .....	9
圖 6、殘差診斷之 ACF 與 PACF 圖 .....	11
圖 7、ARIMA 白噪音檢定 .....	12
圖 8、國內航空客運每月載客人數-樣本外預測-ARIMA 分析法.....	13
圖 9、國內航空客運每月載客人數-觀察值 .....	14
圖 10、迴歸分析配適診斷圖.....	15
圖 11、Lag 2 White noise 檢定 .....	16
圖 12、時間序列迴歸樣本外預測趨勢圖.....	18
圖 13、國內航空客運每月載客人數-觀察值 .....	19
圖 14、季節因素之時間序列圖.....	21
圖 15、趨勢變動圖之時間序列圖.....	22
圖 16、去季節因子之時間序列圖.....	23
圖 17、分解法—white noise 檢定 .....	25
圖 18、國內航空客運每月載客人數-樣本外預測-分解法.....	25
圖 19、國內航空客運每月載客人數-樣本外預測-指數平滑法.....	28

# 第一章、緒論

談到國內航空，浮現於腦海中的關鍵字，似乎就是較老舊的飛機、復興航空空難，以及高速鐵路對於國內航空業的衝擊等等。由於多數民眾搭乘國內航空客運的機會較小，對於其營運與載客狀況因此不甚了解。我們希望可以藉由此次研究，能夠更瞭解目前國內航空載客量的變化與趨勢，以及其他交通工具對於國內航空載客量的影響。藉由研究載客量趨勢，希望可以預估未來的狀況，幫助國內航空公司評估所需班次數量，進而達到控制成本的目的。

本報告後續內容架構安排如下：第二章將介紹研究方法，我們在此章節會介紹分析的方法及研究流程。第三章將說明各分析預測之結果，並選出最佳模型。最後則為結論。

## 第二章、研究方法

在本章我們將會介紹本報告的研究流程，並說明我們所使用的 4 種預測模型及四種評估準則。

### 一、研究流程

首先決定本次報告的研究方向為國內航空載客量，資料時間從 2002 年 1 月至 2016 年 3 月，共 171 筆資料，使用此資料進行初步的 ARIMA 分析 (Autoregressive Integrated Moving Average Model, ARIMA)。時間序列分析方法使用 ARIMA 分析法、時間序列迴歸法、分解法以及指數平滑法，配適模型並進行預測。在預測前我們保留最後 12 筆資料，以便與後續的 12 筆預測資料做四種預測方法的評估，四種評估準則分別為：MSE(Mean Squared Error)、MAE(Mean Absolute Error)、MAPE(Mean Absolute Percentage Error)，以及 MPE(Mean Percentage Error)，檢驗預估是否準確。最後選定最佳模型，並給予我們的結論與建議。

### 二、研究方法

在這個小節，我們將介紹本報告會使用到的 4 種時間序列預測方法以及 4 種

預測評估準則，依次為 ARIMA 模型、時間序列迴歸法、分解法、指數平滑法及四種預測評估準則。

## 1. ARIMA 模型

ARIMA 模型，差分整合移動平均自迴歸模型，又稱整合移動平均自迴歸模型（移動也可稱作滑動），時間序列預測分析方法之一。ARIMA(p, d, q)中，AR 是"自迴歸"，p 為自迴歸項數；MA 為"移動平均"，q 為移動平均參數，d 為使之成為平穩序列所做的差分次數（階數）。「差分」一詞雖未出現在 ARIMA 的英文名稱中，卻是關鍵步驟。分析流程：

- 1). 根據資料的時間序列散佈圖等，判斷資料的平均數與變異數是否平穩，若變異數不平穩，可取對數或開根號；若平均數不平穩，則使用差分。
- 2). 由時間序列之相關函數 (Autocorrelation Function，以下簡稱 ACF)圖以及偏自相關函數 (Partial Autocorrelation Function)圖，觀察是否具有以下特徵：
  - I. 若超出兩倍標準差，則可能需要配適模型。
  - II. 資料若有季節性趨勢，則可能需要做季節差分。
  - III. 由趨勢，初步判斷要使用何種模型做預測。
- 3). 參數估計，檢驗其是否顯著，需不需要保留在模型中。
- 4). 診斷殘差是否為白噪音。
- 5). 使用配適的模型進行預測。

## 2. 時間序列迴歸法

時間序列迴歸法在使用前具有以下假設：

- 1). 事件發展呈現線性關係，因此可從過去數據找到相關法則來推測事件未來發展。
- 2). 事件發展具有隨機性，故所有事件的未來發展都可能受突發因素影響而使預測改變

時間序列中的殘差與各觀察值之間常具有直接相關性，如此相關性已違背觀察值為獨立的假設，因此時間序列迴歸分析是依照變數本身過去的數據資料所存在的變異型態來建立模型以便進行預測。而時間序列迴歸法是將時間序列以趨勢

(Trend,  $TR_t$ )、季節(Season,  $SN_t$ )以及殘差( $\varepsilon_t$ )來表示模型。

基本模型假設: $Y_t=TR_t+SN_t+\varepsilon_t$

### 3. 時間序列分解法

時間序列分解法(Time Series Decomposition methods)是受到趨勢變動(secular trend)、季節變動(seasonal variation)、週期變動(cyclical variation)和不規則變動(irregular fluctuation)這四個因素的影響，利用分解法確定了四種因素後，就能根據結果做進一步的預測。

- 1). **趨勢變動**：長期來看，時間數列隨著時間而增減變化而呈現上升或下降的緩慢趨勢，而造成一種有系統性的變動傾向，換句話說，時間數列呈現出緩慢而有規律的變化。例如近兩三年來，手機的銷售量逐漸增加；近十多年來，人口死亡率因醫藥日趨發達而漸漸降低。不過長期的變化也可能沒有明顯上升或下降的趨勢，呈現停滯狀態。
- 2). **季節變動**：時間數列在每一年中呈現出規律的週期性變動，即同一月份或同一季會有差不多的增減變化。季節變動存在的原因大部份是因為自然現象或社會現象所引起的，如四季的冷暖會影響冰品的銷售、寒暑假期間可能有較多的犯罪事件等。因此，以月或季為單位的資料其季節變動的週期為一年，以年為單位的資料則沒有季節變動之存在。
- 3). **循環變動**：一種有規律地伴隨著長期趨勢作週而復始上下變化的現象，常稱之為「循環變動」。循環變動與季節變動很類似，兩者都是有系統地作週期性的變化，只是循環變動的週期較長，約二至十五年不等（隨資料而不同）。通常循環變動的週期有時稍長，有時較短，上下變動的幅度也會稍微不一致，經濟方面的時間數列其循環變動較明顯，故循環變動常被稱為「景氣循環變動」。
- 4). **不規則變動**：時間數列資料隨機變動，具有不規則性變動，是不可預測的隨機變動。不規則變動是去除長期趨勢，季節變動，循環變動等三種成份後的不可預測的隨機變動，例如 921 大地震所引發股市的短暫下挫，聖嬰現象造成氣候短時間的異常，上游廠商供貨不足而影響成品銷售量，這些無法預期的因素造成時間數列短暫的異常變化。

常見的方法有兩種，加法模型(additive model)和乘法模型(multiplicative model)，加法模型適用於時間序列趨勢較平穩的資料，乘法模型適用於時間序列趨勢較不平穩、變異較大的資料上。

## 4. 指數平滑法

指數平滑法(Exponential Smoothing Method)是生產預測中常用的一種方法。簡單的全期平均法是對時間數列的過去數據一個不漏地全部加以同等利用；移動平均法則不考慮較遠期的數據，併在加權移動平均法中給予近期資料更大的權重；而指數平滑法則兼容了全期平均和移動平均所長，不捨棄過去的數據，但是僅給予逐漸減弱的影響程度。

## 5. 四種預測評估準則

$$MAD = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}$$

$$MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}$$

$$MPE = \frac{\sum_{t=1}^n PE_t}{n} \quad PE_t = \frac{y_t - \hat{y}_t}{y_t} \times 100$$

$$MAPE = \frac{\sum_{t=1}^n APE_t}{n} \quad APE_t = \frac{|y_t - \hat{y}_t|}{y_t} \times 100$$



### 第三章、 原始資料分析

在本章節中，我們會先介紹我們的原始資料：國內航空客運的每月載客人數。本報告資料來源為主計處總體統計資料庫，資料時間從 2002 年 1 月至 2016 年 3 月，共 171 筆資料，原始資料內容於附錄一。我們保留最後十二筆資料作為預測的對照依據。

如圖 1 所示，2003 年 4 月至 2003 年 6 月的每月載客人數明顯降低，其中五

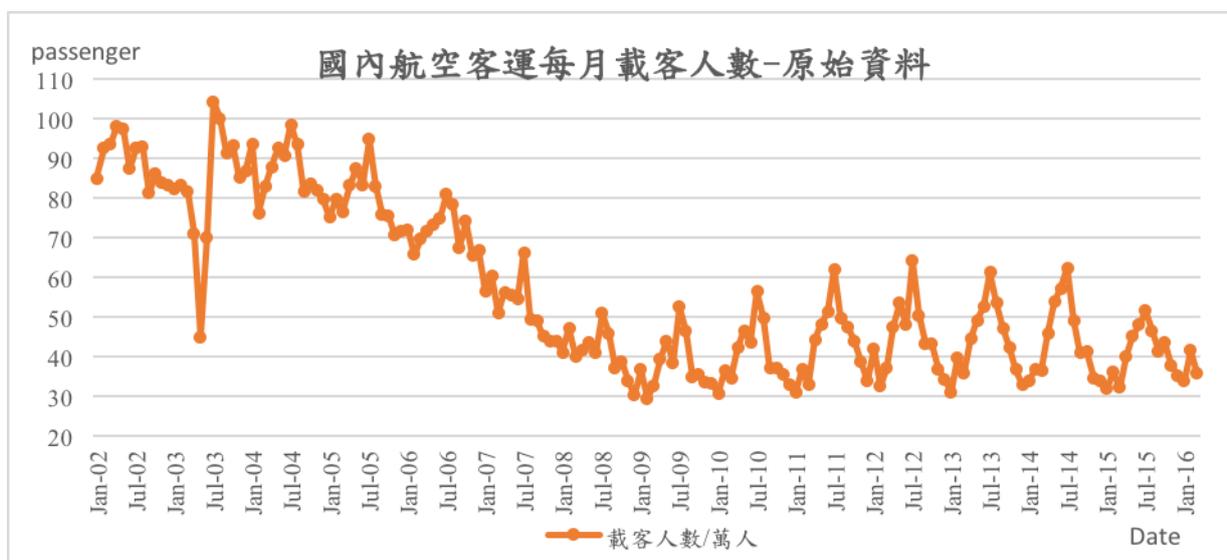


圖 1、國內航空客運每月載客人數-原始資料

月為將近 3 月份的一半，有顯著的降低，推測原因為台灣於 2003 年 3 月開始出現 SARS 疫情，4 月中前疫情仍在控制之下，但至四月底疫情爆發，五月中世界衛生組織（WHO）正式將台灣列為 SARS 感染區，直到 6 月中才正式從感染區除名。從 2004 年五月開始，整體的人數趨勢開始緩慢下降，直至 2009 年 1 月，開始有較穩定的波動。部分原因推測為 2007 年 1 月台灣高速鐵路通車，由於高鐵與國內空運兩者性質相似，因此通車後，影響到國內空運的載客量。

如圖 2 所示，高鐵每月載客人數從 2007 年 1 月至 2008 年 8 月有顯著上升，國內航空客運載客人數於同時段顯著下降；2009 年後高鐵載客人數緩慢上升，空運載客人數漸漸呈現平穩波動趨勢。由於前期資料受疾病、高鐵因素影響，且後期資料呈現相對平穩狀態，因此進行資料分析及預測時，只採用最後一百筆資料。

## 國內航空載客現況與預測

由於倒數第一百筆資料為 2007 年 12 月，為方便分析整年資料，因此將此資料扣除，由 2008 年 1 月開始作資料分析，且載客人數皆為萬人以上，因此全部資料以萬人為單位做計算。

載客人數與各月天數呈正相關，因此在做月資料估計時，須排除天數的不同，將各月份資料除以各月天數再乘以 30，使得各月份天數基準相同。得到預測十二筆資料後，比較與原始資料的誤差。

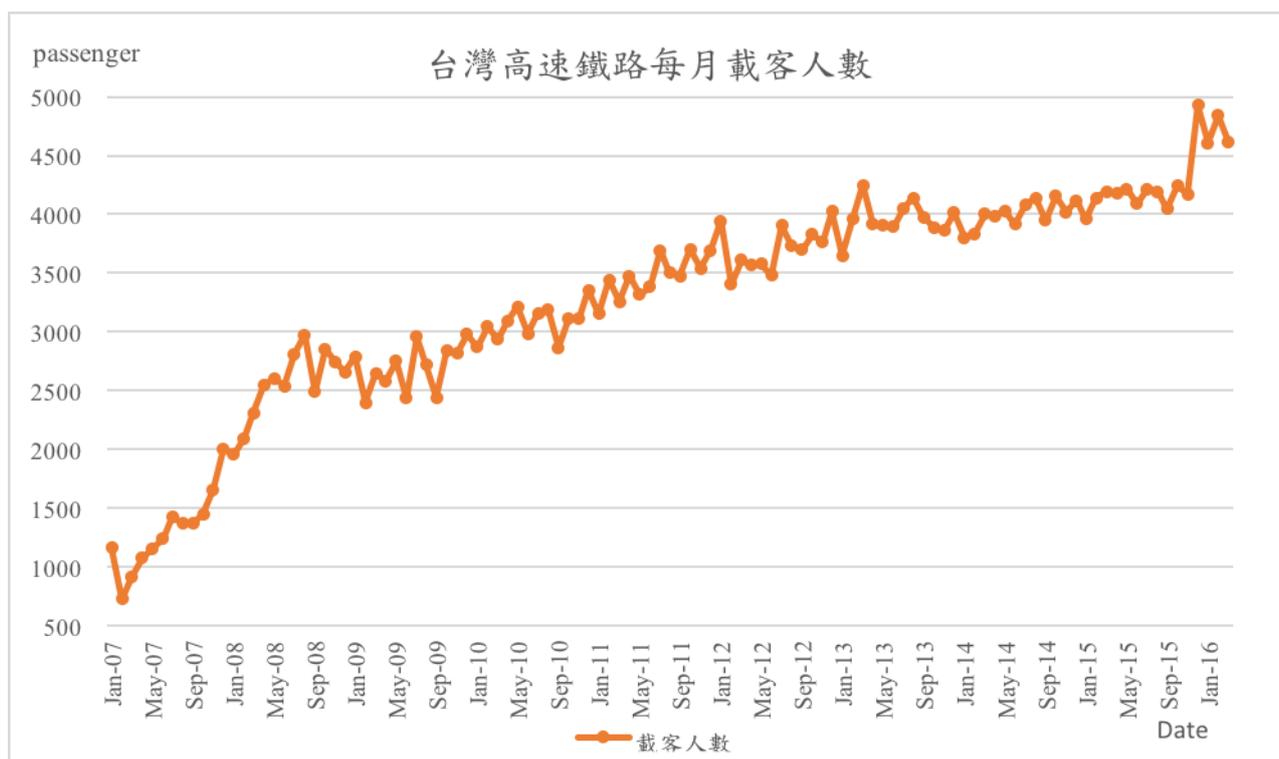


圖 2、台灣高速鐵路每月載客人數（資料來源：主計處總體統計資料庫）

為更清楚闡述資料特性，將歷年各月份資料歸類於一起，繪製盒形圖，顯示各月份的最大值、最小值、中位數、下四分位數及上四分位數，如圖 3。統計資料共有 87 個月份，如圖中可見，資料具有季節性，其中以七月份為最高。

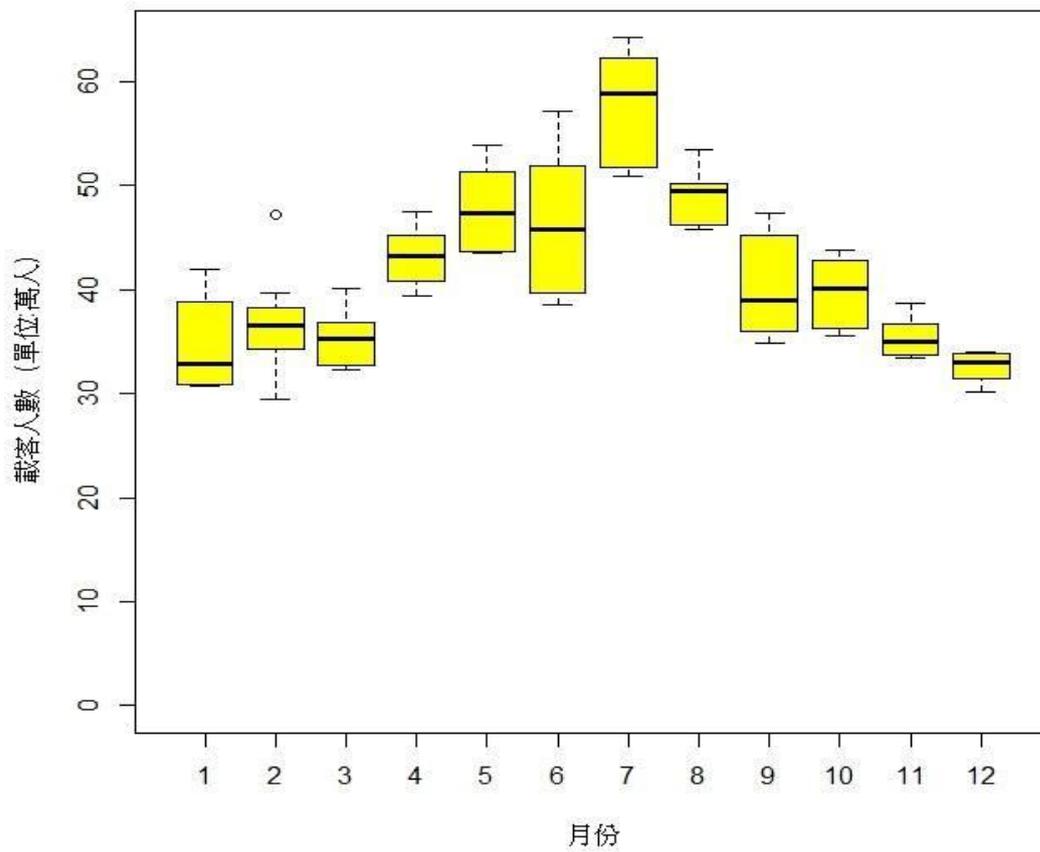


圖 3、國內航空客運每月載客人數-盒型圖

## 第四章、ARIMA 分析

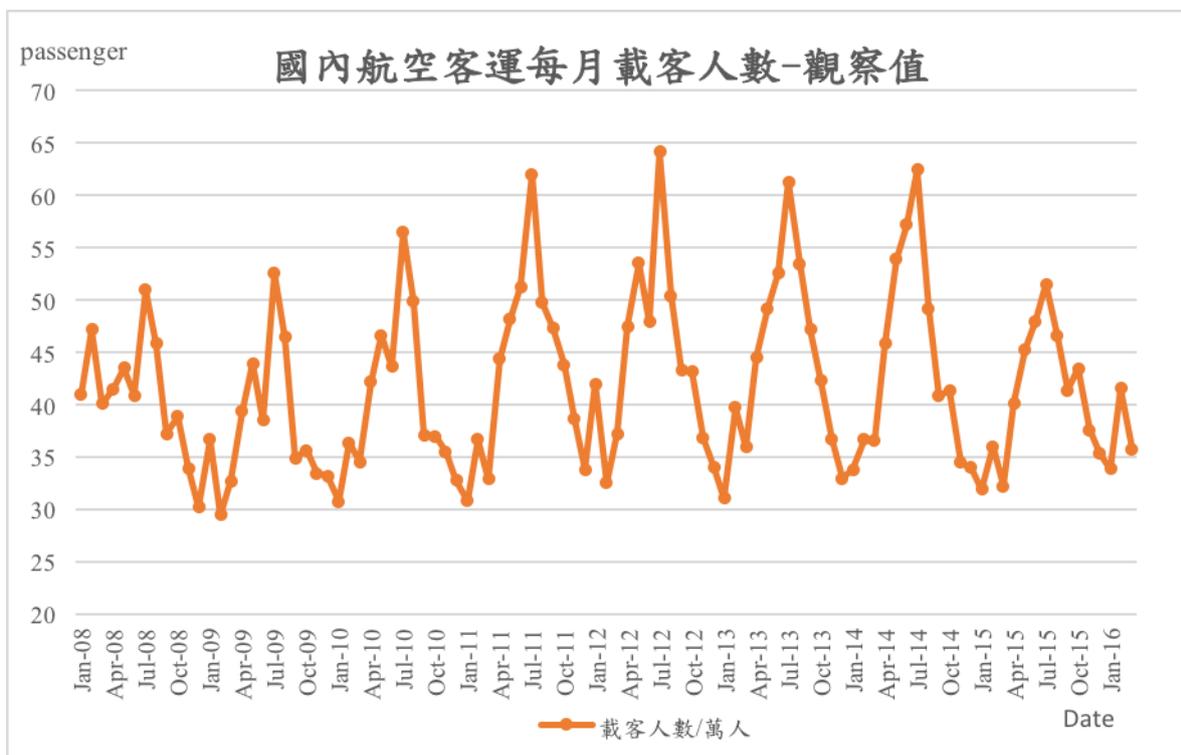


圖 4、國內航空客運每月載客人數-觀察值

以下對最後的九十九筆資料進行 ARIMA 分析。

首先判斷資料的變異數以及平均數是否平穩。如圖 4 所示，最後 99 筆資料平均數與變異數皆相對平穩，因此對於此資料不需要做差分以及取對數。

在配適模型前，需先檢定資料是否有自相關。如表 1 所示，由於 P-value 皆小於 0.0001，因此拒絕虛無假設，至少有一個  $\rho_i \neq 0$ ，有自相關，因此需配適模型。

表 1、原始資料自相關檢定

白噪音的自相關檢查									
至滯後	卡方	DF	Pr>ChiSq	自相關					
6	122.34	6	<.0001	0.667	0.416	0.125	-0.294	-0.535	-0.560
12	262.03	12	<.0001	-0.537	-0.349	0.063	0.315	0.568	0.740

18	366.11	18	<.0001	0.525	0.291	0.052	-0.310	-0.473	-0.524
24	477.24	24	<.0001	-0.491	-0.322	0.012	0.222	0.428	0.596

接著觀察資料之 ACF 圖及 PACF 圖，如圖 5 所示。

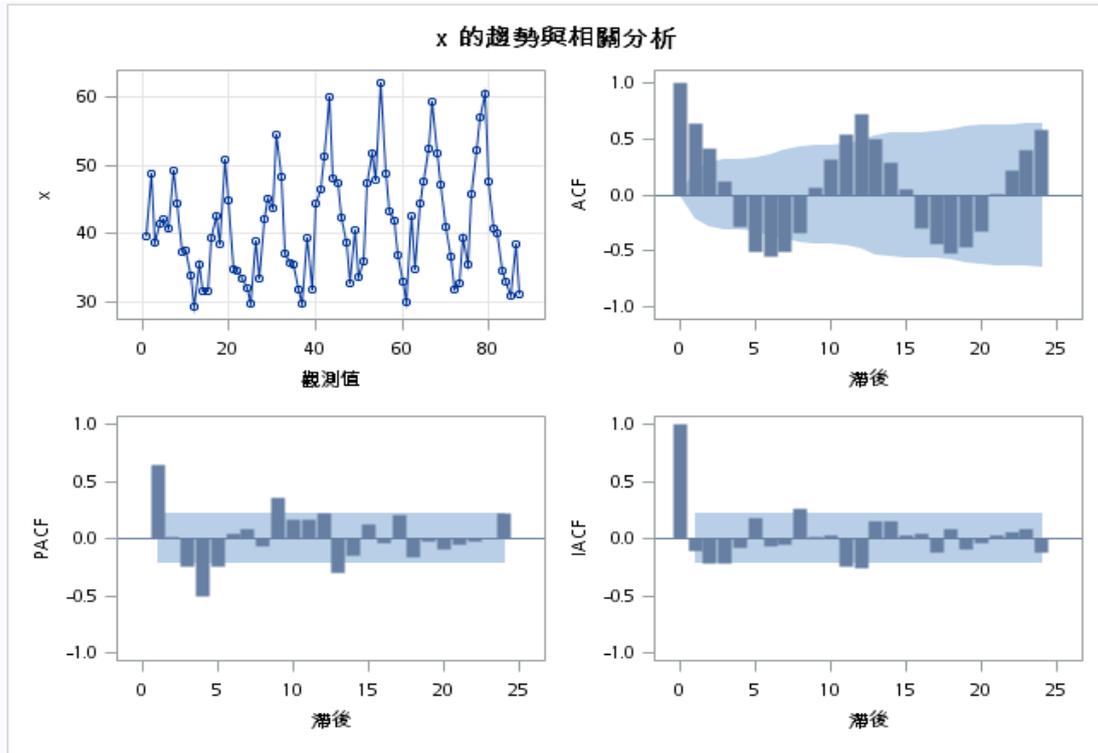


圖 5、資料之 ACF 與 PACF 圖

由於 ACF 圖中，lag1, lag2 die down，PACF 圖 cuts off，且圖形呈現季節性變化，因此預估配 AR 模型並做季節差分。

由 ACF 與 PACF 圖初步判斷模型後，經過嘗試，最後選擇配適 ARIMA(2,0,0)(2,1,0)<sub>s</sub> 模型。

配適模型後，首先檢查模型中各參數是否需要保留在模型中。若  $p\text{-value} < \alpha = 0.05$ ，則拒絕虛無假設，參數必須保留在模型內。檢查結果如表 2，參數項之 P-value 皆小於 0.05，因此參數需保留在模型中。最後進行殘差的診斷，檢查 ACF 與 PACF 是否落入兩倍標準差之內，結果如圖 6。雖然 PACF 於 lag5 及 lag11 略微超過，但由於超過不多，模型配適應屬合適。



表 2、ARIMA 參數顯著檢定

Model Parameter	Estimate	Std. Error	T	Prob>   T
Intercept	0.39622	0.6536	0.6062	0.5463
Autoregressive, Lag 1	0.37697	0.1108	3.4019	0.0011
Autoregressive, Lag 2	0.33958	0.109	3.1147	0.0027
Seasonal Autoregressive, Lag 12	-0.76862	0.1149	-6.6912	<0.0001
Seasonal Autoregressive, Lag 24	-0.43139	0.1126	-3.8305	0.0003
Model Variance (sigma Squared)	11.86			

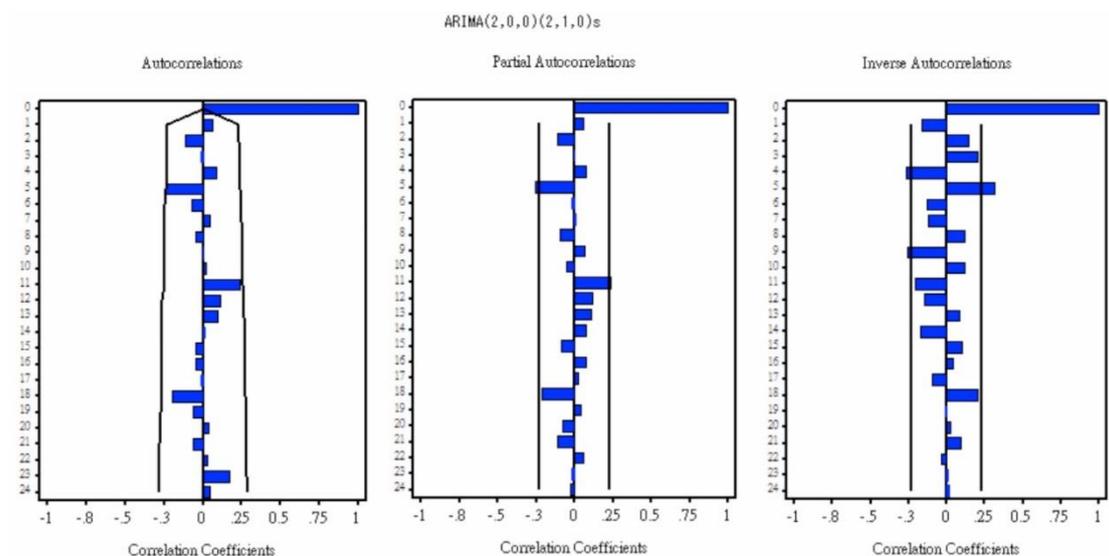


圖 6、殘差診斷之 ACF 與 PACF 圖

檢定殘差項是否為白噪音，診斷結果如圖 7 所示。除了 lag5 之 P-Value 略小於 0.05，其他皆大於 0.05，因此不拒絕虛無假設，殘差項為白噪音，殘差項大致上無自我相關。



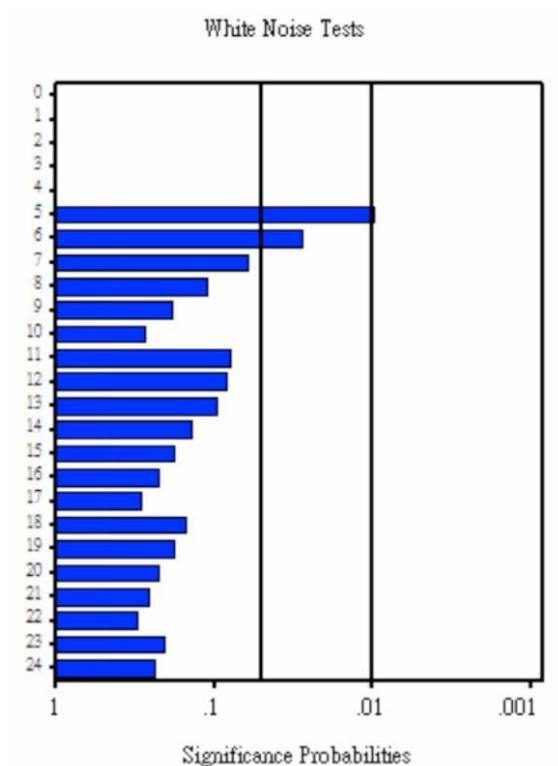


圖 7、ARIMA 白噪音檢定

由參數及白噪音檢定，此模型應屬配適合適。由表 3，我們可寫出此模型的公式：

$$(1-B^{12})(1 - 0.37697B - 0.33958B^2)(1 + 0.76862B^{12} + 0.43139B^{24})Z_t = 0.39622 + a_t$$

表 3、模型參數估計值

Model Parameter	Estimate
Intercept	0.39622
Autoregressive, Lag 1	0.37697
Autoregressive, Lag 2	0.33958
Seasonal Autoregressive, Lag 12	-0.76862
Seasonal Autoregressive, Lag 24	-0.43139
Model Variance (sigma Squared)	11.859995

寫出公式後，便可使用所配適之模型預估未來值。進行樣本外的十二筆資料預測，其值以及 95%信賴區間上下界，如表 4：

表 4、樣本外 12 筆預測觀測值及其 95%信賴區間上下界

變數的預測 x			
觀測值	預測	95%信賴區間	
88	44.94	38.1935	51.6931
89	48.81	41.5973	56.0242
90	50.86	42.9489	58.7735
91	60.17	51.9885	68.3557
92	49.27	40.8710	57.6699
93	43.97	35.4494	52.4875
94	41.27	32.6684	49.8688
95	36.47	27.8201	45.1203
96	32.97	24.2902	41.6553
97	31.61	22.9044	40.3106
98	41.09	32.3786	49.8112
99	34.78	26.0560	43.5056

最後，將十二筆預測值與實際值做比較，觀察實際值是否落於預測之信賴區間內。結果如圖 8。

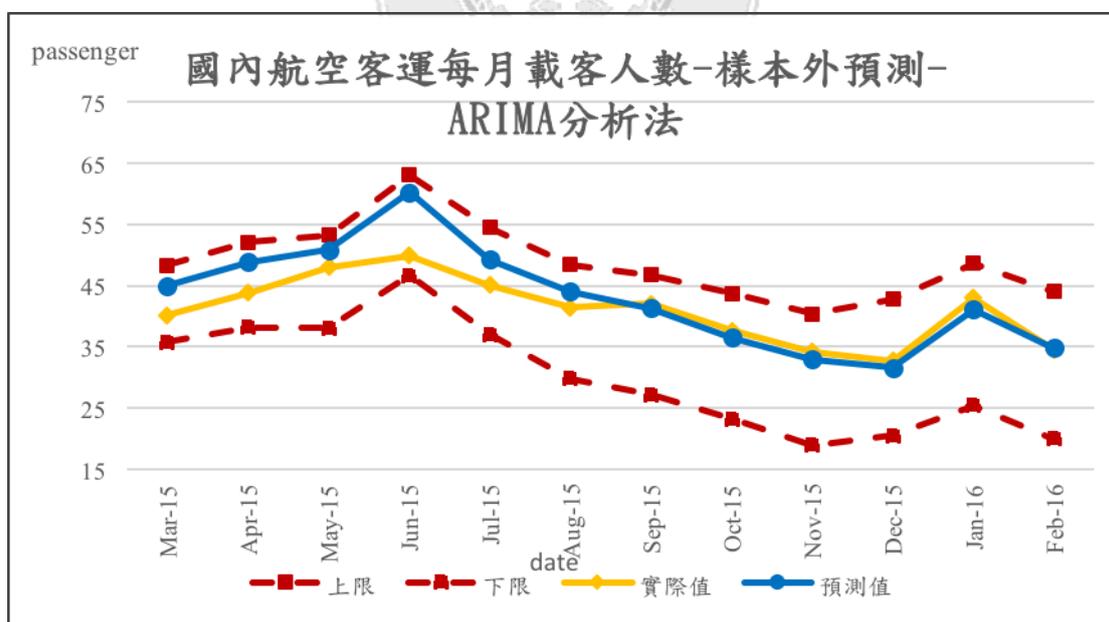


圖 8、國內航空客運每月載客人數-樣本外預測-ARIMA 分析法

## 第五章、時間序列迴歸

本章節使用時間序列迴歸的方式，對國內航空客運每月載客人數進行預測，圖 9 為國內航空每月載客人數之原始時間序列圖，資料時間為 2008 年 1 月至 2015 年 3 月，其中保留 12 筆資料做預測，其預測時間為 2015 年 4 月至 2016 年 3 月。

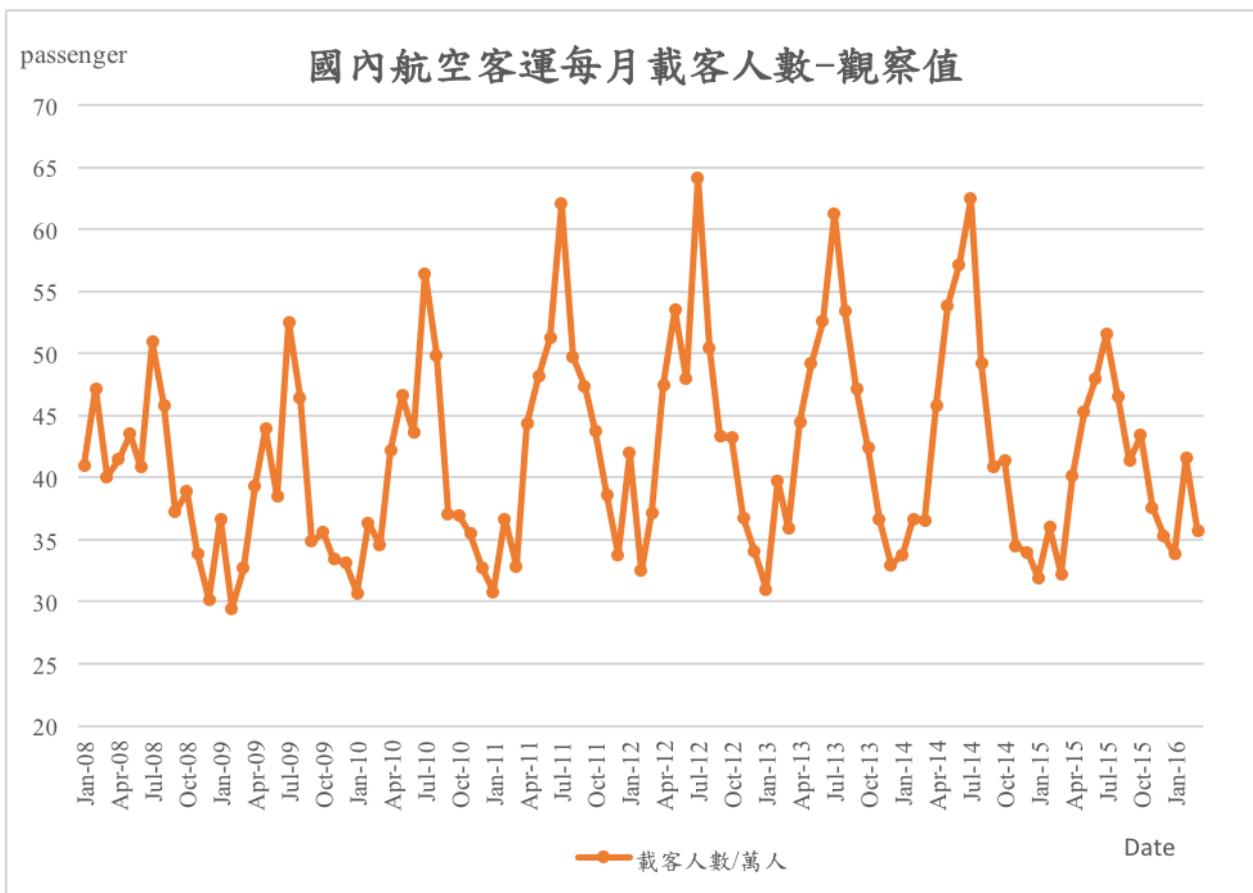


圖 9、國內航空客運每月載客人數-觀察值

由圖 9 初步判斷，平均數與變異數皆平穩，不需將原始資料進行變異數的轉換，所以直接使用 SAS 軟體進行時間序列迴歸分析，結果如表 5，調整 R 平方為 0.7800，由圖 10 可判斷殘差經過排序之後接近 45 度線，且直方圖近似常態分配，此迴歸模型是合適的。

表 5、ANOVA 表

觀測值數目：87					
變異數分析					
來源	自由度	平方和	平均值平方	F 值	Pr > F
模型	12	4444.58	370.38	26.41	<.0001
誤差	74	1037.92	14.03		
已校正的總計	86	5482.50			
根 MSE	應變平均值	變異係數	R 平方	調整 R 平方	
3.75	41.18	9.10	0.81	0.78	

殘差總和	0
平方殘差和	1037.91849
預測殘差 SS (PRESS)	1441.10872

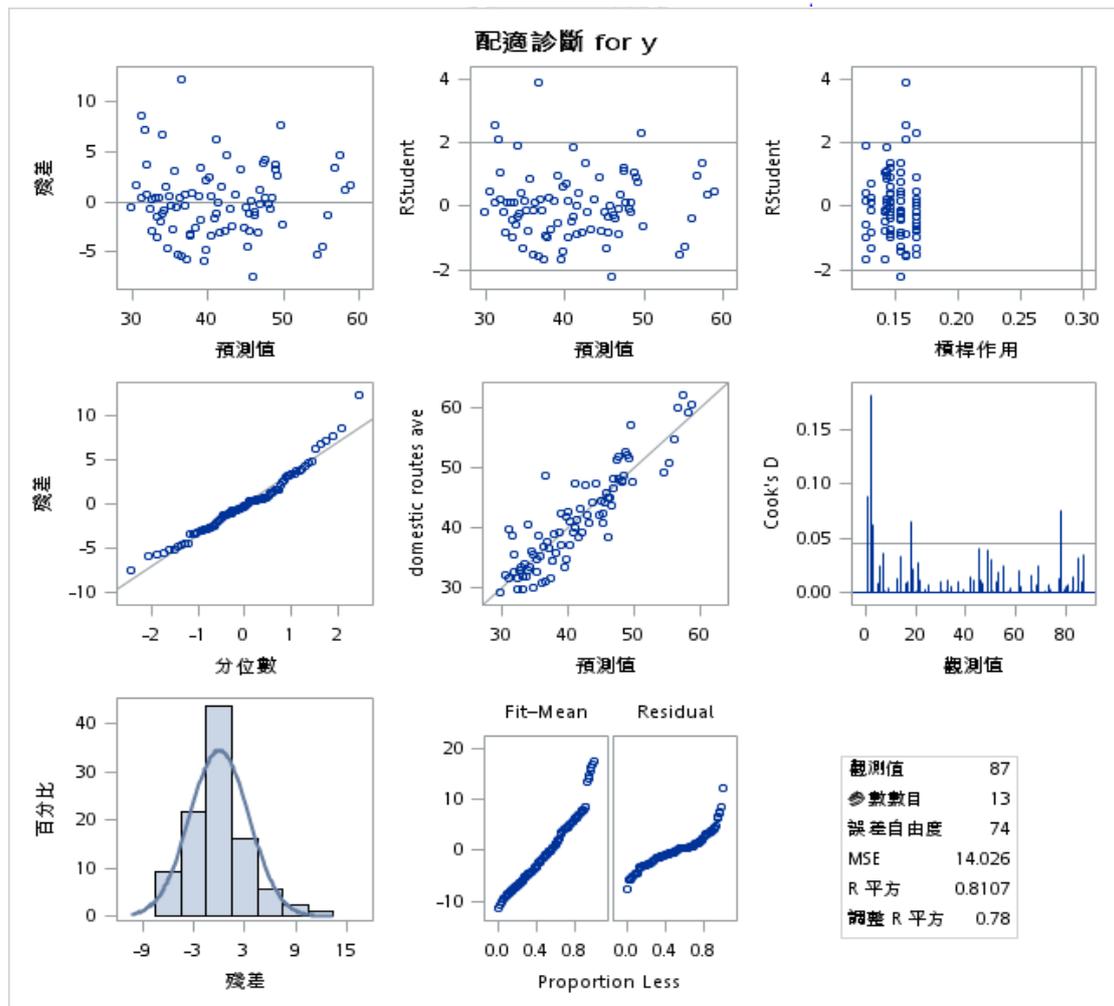


圖 10、迴歸分析配適診斷圖

以上小結，由迴歸與殘差分析，判斷模型配適合適，首先檢查 ACF 是否 Dies down，並且在兩倍標準差之內。確定模型為 AR 模型 lag2，如圖 11。

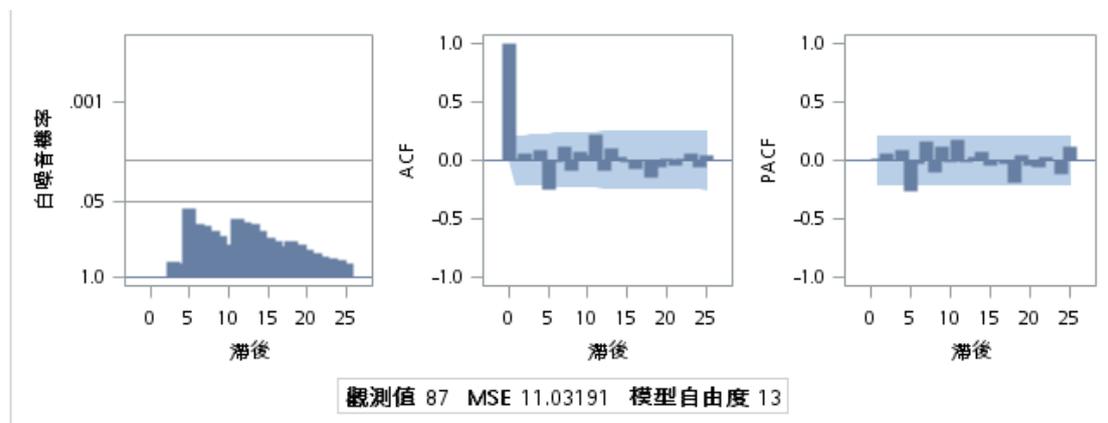


圖 11、Lag 2 White noise 檢定

確定模型 AR(2)，接下來做 Durbin-Watson 自相關的檢定。首先使用  $Pr < DW$  來檢定殘差項是具有正自我相關，若  $P$ -value 值大於 0.05，則拒絕虛無假設，表示殘差項無正自我相關。接著使用  $Pr > DW$  來檢定殘差項是否有負自我相關，若  $P$ -value 值大於 0.05，表示殘差項無負自我相關。由表 6 可看出，0.2649 大於 0.05，因此無正自我相關，0.7351 大於 0.05，因此無負自我相關，由此可知誤差項無自相關性，模型配適合適。

表 6、時間序列迴歸法—Durbin-Watson 統計值

Durbin-Watson 統計值			
順序	DW	Pr<DW	Pr>DW
1	1.8579	0.2649	0.7351

表 7、時間序列迴歸法—自迴歸參數的估計值

自迴歸參數的估計值			
滯後	係數	標準誤差	t 值
1	-0.35853	0.116523	-3.08
2	-0.14973	0.116523	-1.28

表 8、時間序列迴歸法－參數估計值

參數估計值					
變數	自由度	估計值	標準誤差	t 值	近似 pr>   t
INTERCEPT	1	30.4364	1.8609	16.36	<0.0001
t	1	0.044	0.0276	1.59	0.1156
d1	1	1.2712	1.4774	0.86	0.3924
d2	1	6.6977	1.6091	4.16	<0.0001
d3	1	1.7084	1.7389	0.98	0.3292
d4	1	11.1064	1.8502	6	<0.0001
d5	1	14.3878	1.8868	7.63	<0.0001
d6	1	15.0305	1.8978	7.92	<0.0001
d7	1	24.3061	1.8875	12.88	<0.0001
d8	1	15.3432	1.8527	8.28	<0.0001
d9	1	8.81	1.7903	4.92	<0.0001
d10	1	6.7673	1.6482	4.11	0.0001
d11	1	3.4936	1.5036	2.32	0.023

由上述 DW 檢定可知此時間序列迴歸已無自相關，最後檢測殘差是否符合 White noise 檢定，若殘差不符合則必須放棄此組合，改用其他組合，p-value 大於 0.05 顯示符合 white noise，如圖 11。並依表 7、表 8 的自迴歸參數估計以及參數估計寫出國內航空客運每月載客人數時間序列迴歸式：

$$y_t = 30.4364 + 0.044t + 1.2712d_1 + 6.6977d_2 + 1.7084d_3 + 11.1064d_4 \\ + 14.3878d_5 + 15.0305d_6 + 24.3061d_7 + 15.3432d_8 + 8.81d_9 \\ + 6.7673d_{10} + 3.4936d_{11} + 0.358527\varepsilon_{t-1} + 0.149725\varepsilon_{t-2} + a_t$$

$$d_1 \begin{cases} 1, \text{ when Jan} \\ 0, \text{ otherwise} \end{cases}, d_2 \begin{cases} 1, \text{ when Jan} \\ 0, \text{ otherwise} \end{cases} \dots \dots, d_{11} \begin{cases} 1, \text{ when Jan} \\ 0, \text{ otherwise} \end{cases}$$

其中，誤差項為  $\varepsilon_t = 0.358527\varepsilon_{t-1} + 0.149725\varepsilon_{t-2} + a_t$  符合 white noise 無自我相關，模型配適度佳。

## 國內航空載客現況與預測

總結，圖 12 為利用時間序列迴歸做樣本外預測趨勢圖，在前六筆預測都有高估的情況，除了在第四筆 2015 年 7 月有明顯高估的情形，而其他預測都非常接近，表示迴歸時間序列適合用於分析此資料。

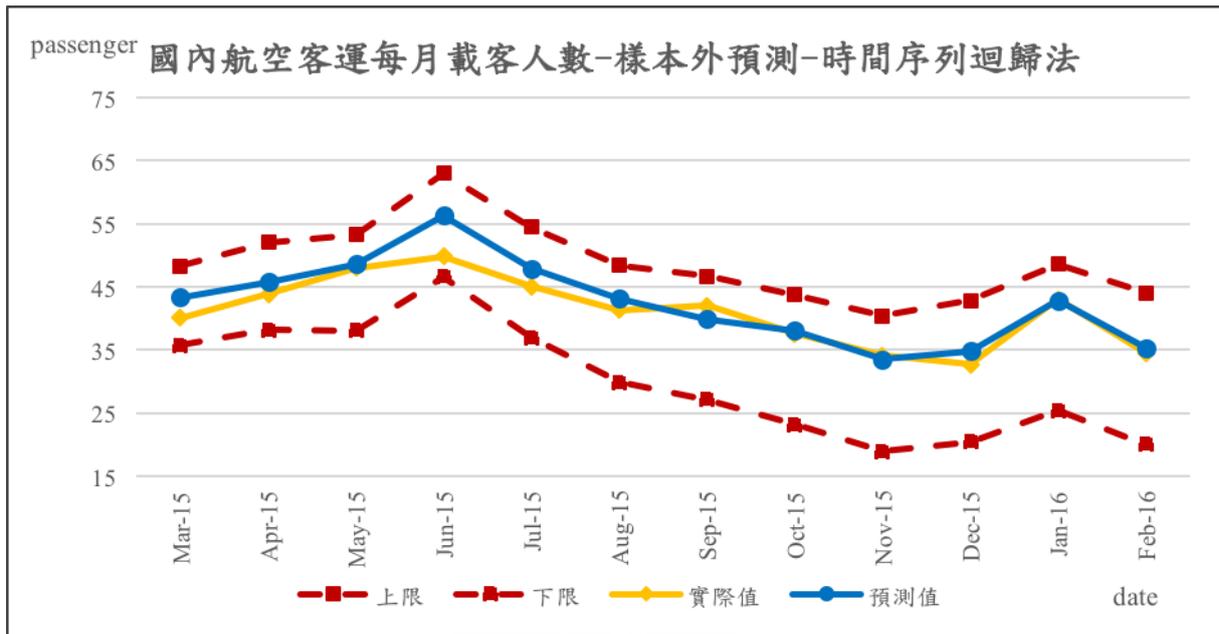


圖 12、時間序列迴歸樣本外預測趨勢圖

## 第六章、時間序列分解法

根據我們使用的資料，本章節使用加法模型對國內航空客運每月載客人數進行預測，圖 13 為國內航空每月載客人數之原始時間序列圖，資料時間為 2008 年 1 月至 2015 年 3 月，其中保留 12 筆資料做預測，其預測時間為 2015 年 4 月至 2016 年 3 月。

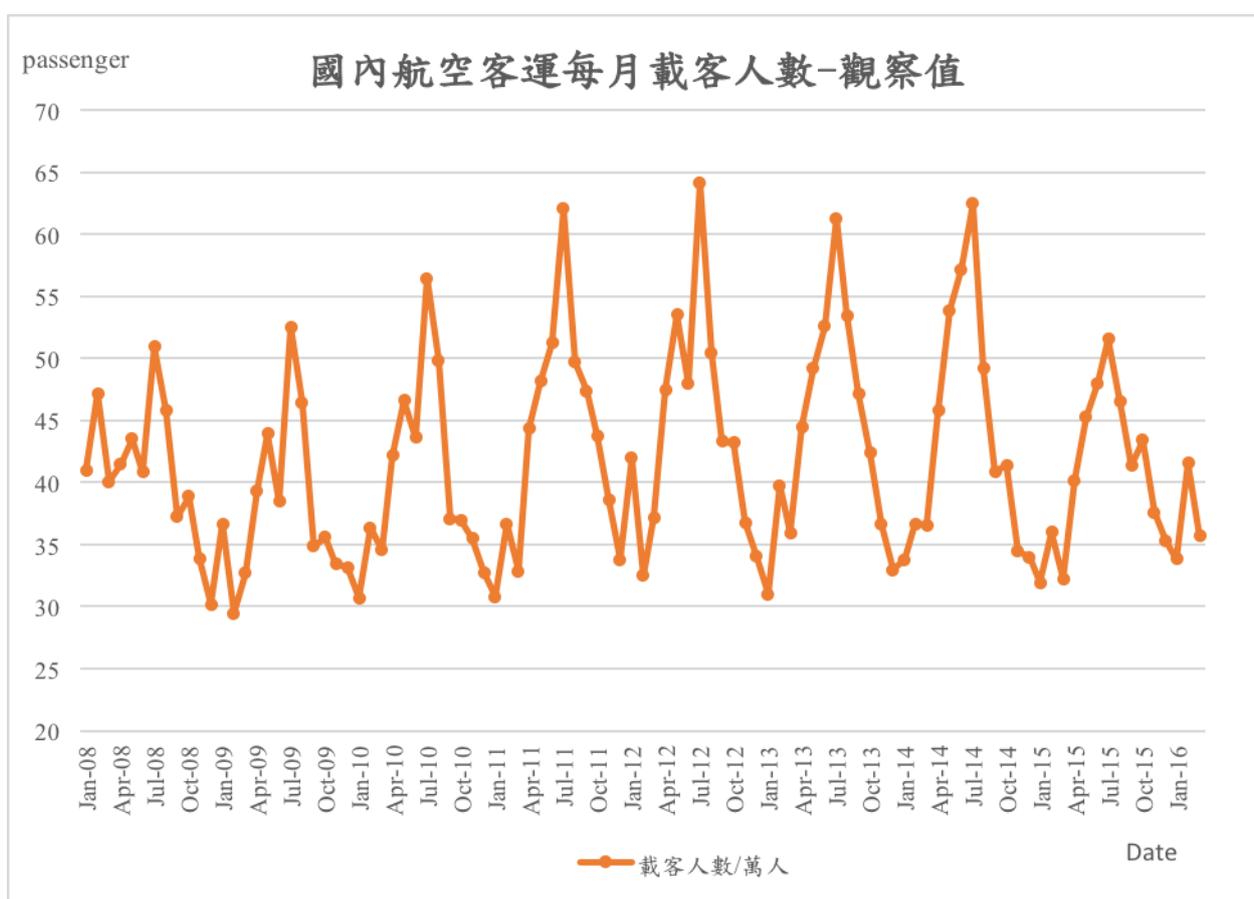


圖 13、國內航空客運每月載客人數-觀察值

表 10、標準 X11 分解法

Time plot of minthly number of people  
(in hundreds of cases)

Obs	number	date	t	Y	SN	DY	TRC	IR
-----	--------	------	---	---	----	----	-----	----

Time plot of minthly number of people  
(in hundreds of cases)

Obs	number	date	t	Y	SN	DY	TRC	IR
1	39.6565	Jan-08	1	39.6565	-8.035	47.6915	43.7206	3.97085
2	48.8136	Feb-08	2	48.8136	-0.5336	49.3472	42.703	6.64425
3	38.7798	Mar-08	3	38.7798	-5.678	44.4578	41.6976	2.76014
4	41.4682	Apr-08	4	41.4682	1.8564	39.6118	40.7262	-1.11433
5	42.0830	May-08	5	42.083	4.3162	37.7668	39.8326	-2.06581
6	40.8423	Jun-08	6	40.8423	3.27	37.5723	39.0365	-1.46413
7	49.2851	Jul-08	7	49.2851	13.9178	35.3673	38.408	-3.04072
8	44.3472	Aug-08	8	44.3472	7.3824	36.9648	37.9646	-0.99987
9	37.2063	Sep-08	9	37.2063	-1.2713	38.4776	37.6982	0.77937
10	37.5957	Oct-08	10	37.5957	-2.2071	39.8028	37.5979	2.20486
11	33.8854	Nov-08	11	33.8854	-4.3938	38.2792	37.6174	0.66174
12	29.1884	Dec-08	12	29.1884	-8.3037	37.4921	37.6455	-0.15337
13	35.4826	Jan-09	13	35.4826	-8.3663	43.8489	37.6315	6.21738
14	31.5631	Feb-09	14	31.5631	-0.7328	32.2959	37.5096	-5.21377
15	31.6391	Mar-09	15	31.6391	-6.0672	37.7063	37.3021	0.40419
16	39.3178	Apr-09	16	39.3178	1.9576	37.3602	37.0478	0.31247
17	42.4896	May-09	17	42.4896	4.466	38.0236	36.8393	1.18425
18	38.4773	Jun-09	18	38.4773	3.5893	34.888	36.7093	-1.82132
19	50.8274	Jul-09	19	50.8274	14.4061	36.4213	36.7417	-0.3204
20	44.9222	Aug-09	20	44.9222	7.2367	37.6855	36.9272	0.75833

由表 10 標準 X11 分解法輸出結果，我們採用加法模型，其中表頭符號意義如下：

Y(the observed time series)：所觀察到的時間序列

SN(the seasonal component)：季節因子

DY(deseasonalized observations)：去掉季節因子

TRC(trend-cycle component)：趨勢變動因子

IR(the irregular component)：不規則變動因子

由此表可以觀察到兩個重點：第一點，從 SN 可以觀察到，第一筆為 2008 年 1 月(-8.035)和第 13 筆為 2009 年 1 月(-8.3663)的數值接近但不相等，由此可知，此資料可能具有季節循環，但波動幅度不同。第二點，我們可以藉由 SN 的

正負值去判別旅遊的淡旺季，結果顯示，以 2008 年為例，旺季為七月(13.9178)，因為 SN 值為正且數值較高，則 12 月(-8.3037)、1 月(-8.035)為淡季，因為 SN 值為多為負值且小。

也可從可從圖 14 季節因素之時間序列圖去佐證，可以觀察到圖形波動大致相同，但波動高度些許不同，代表每年同季節影響載客人數的因子可能不同。

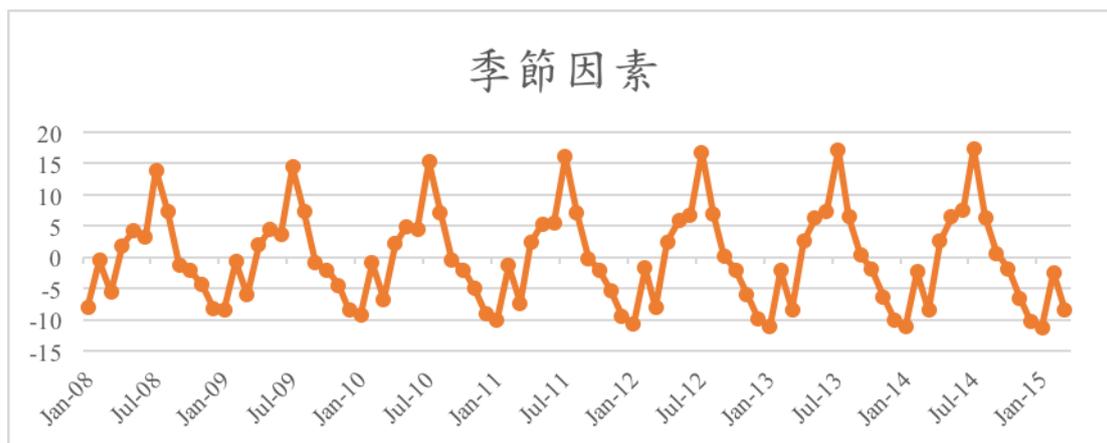


圖 14、季節因素之時間序列圖

從圖 15 趨勢變動圖，可以觀察到趨勢循環線非常複雜，且不是線性圖形，導致增加未來預測上的困難。又因為時間序列迴歸變異數平穩，y 不需做轉換，因此我們採用加法模型來做預測，也就是：

$$y_t = TR_t + SN_t + CL_t + IR_t$$

$y_t$ ：在時間區間t所觀察到的時間序列

$TR_t$ ：在時間區間t的趨勢變動因子

$SN_t$ ：在時間區間t的季節因子

$CL_t$ ：在時間區間t的週期變動

$IR_t$ ：在時間區間t的不規則變動因子

而做預測時則使用：

$$DY = y_t - SN_t = TRC + CL_t + IR_t$$

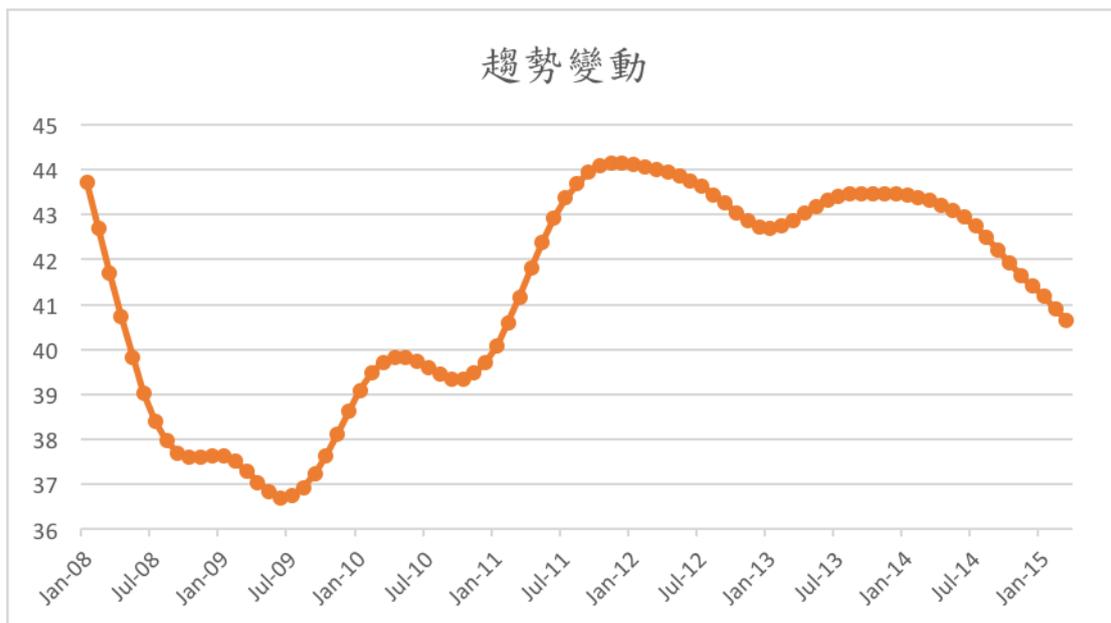


圖 15、趨勢變動圖之時間序列圖

將季節因子去掉之後，可以得到圖 16，橘色線為原始資料，藍色線為去掉季節因子( $DY = y_t - SN_t = TRC + CL_t + IR_t$ )，接著就可以開始配適模型。



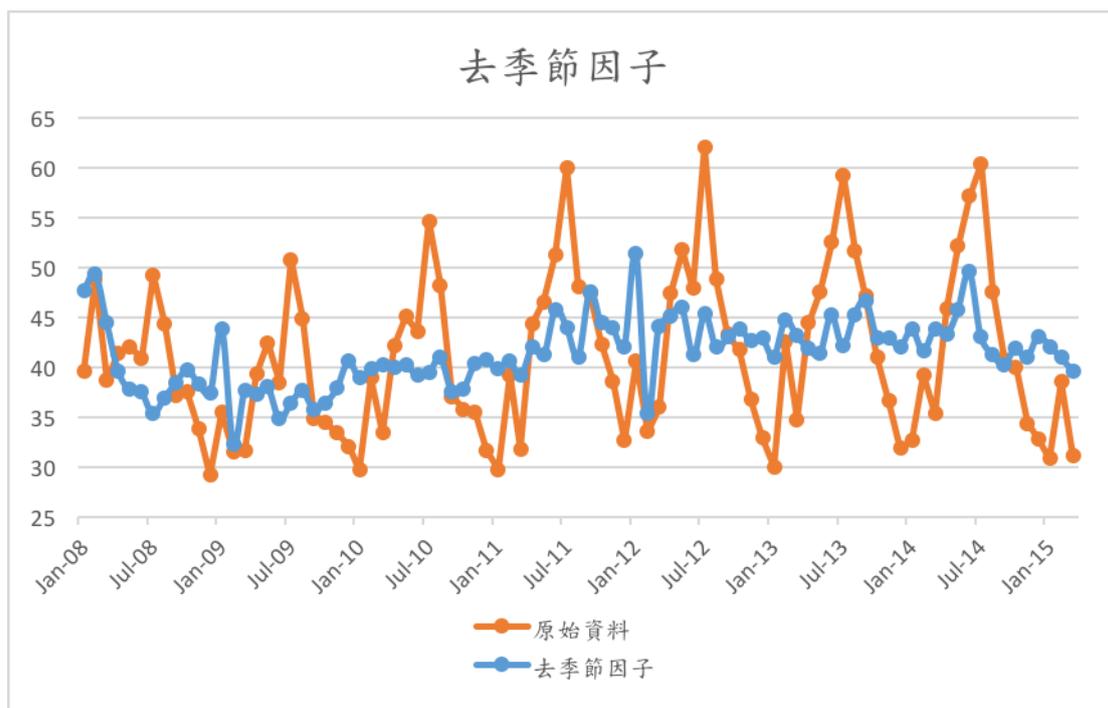


圖 16、去季節因子之時間序列圖

利用變異數檢定，可參考表 11，可以寫出線性模型為：

$$TRC + CL_t = 38.7137 + 0.06143t$$

表 11、分解法—調整後 R 平方

根 MSE	3.15156	R 平方	0.1969
應變平均值	41.41645	調整 R 平方	0.1875
變異係數	7.60945		

參數估計值

變數	自由度	參數估計值	標準誤差	t 值	Pr >  t
Intercept	1	<b>38.7137</b>	0.6816	56.8	<.0001
t	1	<b>0.06143</b>	0.0135	4.57	<.0001

接著對殘差項做 Durbin-Watson test，可以得到表 12 結果，Durbin-Watson 值 1.263 介於 0-2 之間，為正相關，誤差項也和自己有正的自我相關，因此配適

AR(1)模型。

表 12、分解法—Durbin-Watson

Durbin-Watson D	<b>1.263</b>
觀測值數目	87
一階自相關	0.31

下一步做自迴歸參數估計值，SAS 結果如表 13，又因 SAS 系統緣故，必須將得出的結果乘上負號，進而得到

$$\varepsilon_t = 0.309951\varepsilon_{t-1} + a_t$$

表 13、分解法—自迴歸參數估計值

初步 MSE		8.7718	
自迴歸參數的估計值			
滯後	係數	標準誤差	t 值
1	<b>-0.309951</b>	0.103736	-2.99

完成誤差項配適 AR(1)模型之後，必須再做一次檢定，檢測模型是否合適，檢定結果為表 14，得到 Durbin-Watson 值 2.0049，介於 2-4 之間，為負相關，表示模型適當，而此 DW 值接近 2，再由  $Pr < DW$  是用於檢定正自相關的 P-value，由表可知 0.465 大於 0.05，因此拒絕正相關，而  $Pr > DW$  是用於檢定負自相關的 P-value，0.535 大於 0.05，因此拒絕負相關，由此可知道無相關性。而當誤差項配適 AR(1)模型時，趨勢項也受到了影響，得到新的估計，

$$TRC + CL_t = 38.9436 + 0.0567t$$
，如表 15。

表 14、分解法—Durbin-Watson

AUTOREG 程序			
Yule-Walker 估計值			
SSE	753.01949	DFE	84
MSE	8.96452	根 MSE	2.99408
SBC	448.15595	AIC	440.75823
MAE	2.0764707	AICC	441.04739
MAPE	5.0000529	HQC	443.73706

Durbin-Watson	2.0049	轉換的迴歸 R 平方	0.1032
		總 R 平方	0.2837
Durbin-Watson 統計值			
順序	DW	Pr < DW	Pr > DW
1	<b>2.0049</b>	0.465	0.535

表 15、分解法－參數估計值

參數估計值					
變數	自由度	估計值	標準誤差	t 值	近似 Pr >  t
Intercept	1	<b>38.9436</b>	0.9265	42.04	<.0001
t	1	<b>0.0567</b>	0.0182	3.11	0.0026

下一步觀察是否符合 White noise 檢定，如圖 17 結果顯示，符合 White noise 且 ACF 沒有超過兩倍誤差，表示此模型適合。圖 18 為利用分解法做樣本外預測趨勢圖。可以得到最終迴歸

$$D_y = y_t - SN_t = 38.9436 + 0.0567t + 0.309951\varepsilon_{t-1} + a_t$$

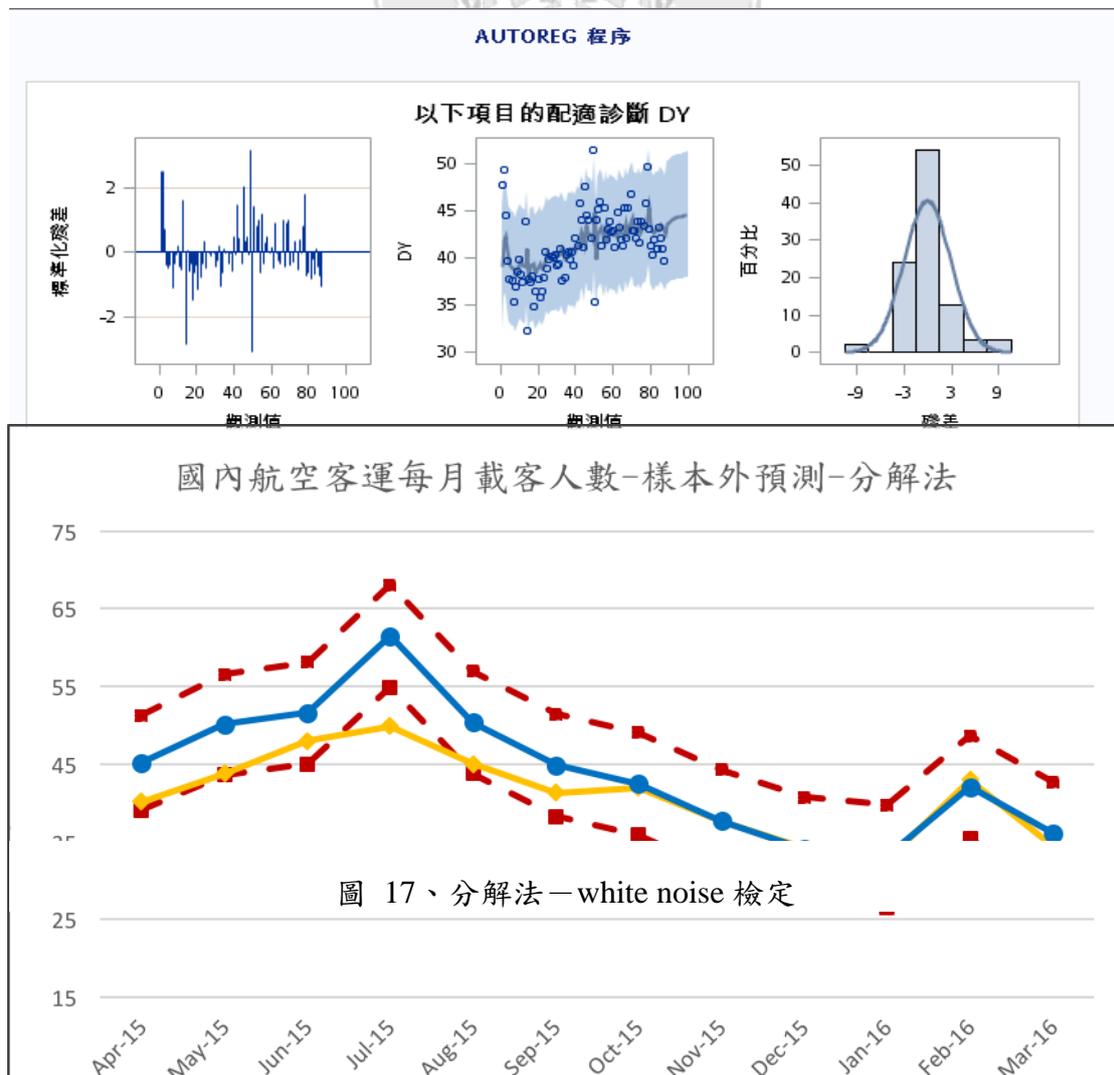


圖 18、國內航空客運每月載客人數-樣本外預測-分解法

## 第七章、指數平滑法

本章採用指數平滑法進行時間序列預測分析。由於變異數平穩，我們採用指數平滑法的加法模型進行預測分析。

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$$

其中

$F_t$  = 第  $t$  期的預測值

$F_{t-1}$  = 前一期的預測值(亦即第  $t-1$  期)

$\alpha$  = 平滑常數(代表預測誤差的百分比)

$A_{t-1}$  = 前一期的實際值

模型：

$$\text{Level: } L_t = 0.48338(y_t - S_{t-s}) + (1 - 0.48338)(L_{t-1} + b_{t-1}),$$

$$\text{Trend: } b_t = 0.00344(L_t - L_{t-1}) + (1 - 0.00344)b_{t-1},$$

$$\text{Seasonal: } S_t = 0.001(y_t - L_t) + (1 - 0.001)S_{t-s},$$

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



表 16、指數平滑法參數估計

<b>Parameter Estimates For Number Of People(Winter method --additive)</b>					
Model Parameter	Estimate	Std.	Error	T	Prob> T
LEVEL Smooth Weight	0.48338		0.065	7.4381	<0.0001
TREND Smooth Weight	0.00344		0.007	0.4926	0.6236
SEASONAL Smooth Weight	0.001		0.0654	0.0153	0.9878
Residual Variance (sigma squared)	10.23624				
Smooth Level	39.5657				
Smooth Trend	-0.021				
Smooth Seasonal Factor 1	-7.72146				
Smooth Seasonal Factor 2	-2.30937				
Smooth Seasonal Factor 3	-7.31305				
Smooth Seasonal Factor 4	2.44354				
Smooth Seasonal Factor 5	5.61531				
Smooth Seasonal Factor 6	6.14072				
Smooth Seasonal Factor 7	15.33479				
Smooth Seasonal Factor 8	6.29204				
Smooth Seasonal Factor 9	-0.33177				
Smooth Seasonal Factor 10	-2.49507				
Smooth Seasonal Factor 11	-5.95056				
Smooth Seasonal Factor 12	-9.71132				

由表 16 可看出，1 月至 3 月為負值，4 月至 8 月為正值，最後 9 月到 12 月為負值，就我們所知由於 1 月接近過年大家都在準備過年所需的東西、大部分的公司都會比較忙，而 2 月為台灣的過年期間但全家出遊大多應以汽車為主要交通工具，另一個原因是有離島在本島工作的居民要回家過年，可能是造成國內航空搭乘人數下降但是降不多的原因，4 月至 8 月逐漸增加到 7 月為最大值，可知這段時期為旅遊旺季，並且可能與暑假期間有關，而 9 月為開學季開始減少出遊人數。

## 國內航空載客現況與預測

由圖 19 可觀察出，利用指數平滑法做樣本外預測，在前六筆資料點相當接近，但是在在此之後的六筆資料點則有較明顯的低估，但也相差不遠。

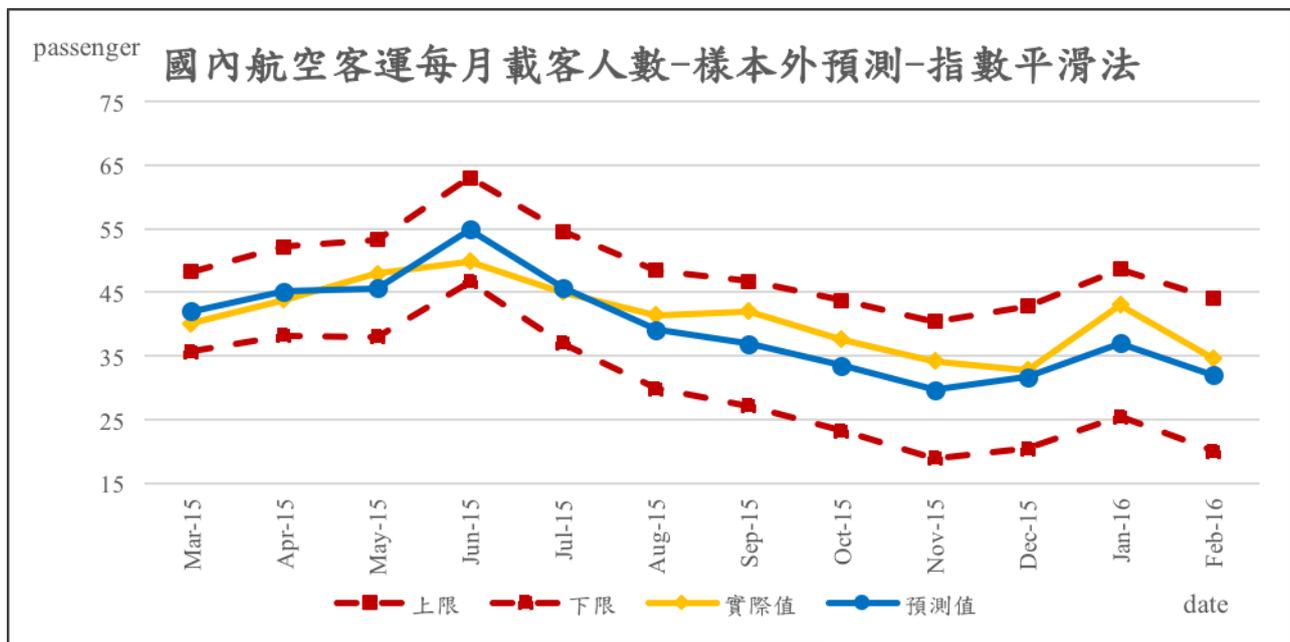


圖 19、國內航空客運每月載客人數-樣本外預測-指數平滑法



## 第八章、 結論

### 一、時間序列預測法之比較

本文對國內航空載客人數之樣本外預測，係透過 ARIMA、時間序列迴歸法、分解法、指數平滑法運算，並計算 MAD(mean absolute deviation)、MSE(mean squared error)、MPE(Mean Percentage Error)、MAPE(Mean Absolute Percentage Error)統計量，比較四種方法何者配適度較高，用已確認較適合之模型。由表 17 中可以看出個統計量最小的值：指數平滑法的 MAD 統計量為 0.13，時間序列迴歸法的 MSE 統計量為 4.83，分解法的 MPE 統計量為-6.93，時間序列迴歸法的 MAPE 統計量為 4.04。得出時間序列迴歸法有 MSE 和 MAPE 統計量表現較優，顯示時間序列迴歸法配適度最佳。

表 17、樣本外預測之 MAD、MSE、MPE、MAPE 比較表

預測方法 \ 統計量	MAD	MSE	MPE	MAPE
ARIMA	0.40	17.17	-4.21	6.88
時間序列迴歸法	0.27	4.83	-1.66	4.04
分解法	0.42	21.53	-6.93	7.31
指數平滑法	0.13	11.13	4.02	7.09

### 二、時間序列預測之結論

由本文使用四種方法之樣本外預測趨勢圖，在第四筆 2015 年 7 月時都有明顯落差，預測皆高估了實際值。為探討其明顯高估的原因，從歷年七月國內航空客運每月載客人數來分析，各年皆約為 50 萬近 60 萬的載客人次，而預測年(2015)的 7 月實際值約為 49 萬人。國內航線分為「西部航線」、「花東航線」及「離島航線」，載客人數以「離島航線」為主，而預測年的前一年(2014) 7 月因復興航

空澎湖空難事件，降低國民隔年前往離島旅遊的意願；另一方面在「花東航線」部分的影響，由於 2014 年 7 月東部鐵路電氣化延伸至台東，普悠瑪號可以從台北直至台東，乘車時間從 6 小時大幅縮減至 3.5 小時，票價更只有機票的三分之一，降低不少民眾搭乘飛機的意願，北東航線平均載客率從通車前的 8 成多一路跌，2015 年初僅剩 5 成多，因此這兩點為可能造成載客人數下降的原因，而其他月份之預測與實際值都相差不遠，表示時間序列迴歸法適合用於分析此資料。

本文預測未來載客人數為高準確度，因此航空公司可以透過本文預測，了解每個時間可能載客量，規劃增加或減少航班次數，準確評估成本及預算，避免在人潮較少之時段發生航班排程過多之情形，以減少損失。同時為提高載客量，可透過預測觀察熱門時段，以增加航班排程，在非熱門時段，不同族群提供專案優惠，選擇更適合的行銷模式，以增加獲利能力。站在消費者角度在熱門時段有足夠的航班可以搭乘，在非熱門時段又可享優惠，以上係本文預測可為航空公司與社會大眾帶來雙贏之局面。



## 參考文獻

Bowerman, B., O'Connell, R., and Koehler, A. (2005) Forecasting, Time Series, and Regression, 4th edition, Duxbury Press.

中華民國統計資訊網，總體統計資料庫，  
<http://statdb.dgbas.gov.tw/pxweb/dialog/statfile9L.asp>。

中國時報，2015，〈普悠瑪 太魯閣 打趴東部航線〉，中國時報網頁，  
<http://www.chinatimes.com/newspapers/20151207001204-260114>，2015/12/07。

自由時報，2014，〈《復興空難》澎湖空難〉，自由時報網頁，  
<http://news.ltn.com.tw/news/life/breakingnews/1062977>，2014/07/24。



## 附錄一、國內航線原始資料

表 18、國內航線原始資料

國內航線	載客人次	國內航線	載客人次	國內航線	載客人次	國內航線	載客人次
200201	849396	200401	937139	200601	720247	200801	409784
200202	926161	200402	761932	200602	658010	200802	471865
200203	935628	200403	828894	200603	696292	200803	400725
200204	981967	200404	879103	200604	717012	200804	414682
200205	974171	200405	926267	200605	731502	200805	434858
200206	874224	200406	908745	200606	750166	200806	408423
200207	928082	200407	986315	200607	809506	200807	509279
200208	931109	200408	934927	200608	783659	200808	458254
200209	812102	200409	817833	200609	673508	200809	372063
200210	861896	200410	836094	200610	742337	200810	388489
200211	840655	200411	819268	200611	656126	200811	338854
200212	832891	200412	799080	200612	667974	200812	301613
200301	823237	200501	750868	200701	566134	200901	366654
200302	832518	200502	796013	200702	604601	200902	294589
200303	817204	200503	764978	200703	510396	200903	326937
200304	710826	200504	831537	200704	562128	200904	393178
200305	449550	200505	875571	200705	555506	200905	439059
200306	700954	200506	834289	200706	544699	200906	384773
200307	1044096	200507	949323	200707	662891	200907	525216
200308	1001774	200508	830508	200708	493884	200908	464196
200309	914494	200509	759962	200709	491276	200909	348501
200310	934005	200510	755024	200710	450234	200910	355828
200311	851388	200511	706933	200711	439775	200911	334085
200312	869364	200512	716442	200712	439416	200912	331500

國內航線	載客人次	國內航線	載客人次	國內航線	載客人次	國內航線	載客人次
201001	307011	201201	419535	201401	337619	201601	338149
201002	363467	201202	324736	201402	366439	201602	415613

國內航空載客現況與預測

國內航線	載客人次	國內航線	載客人次	國內航線	載客人次	國內航線	載客人次
201003	345343	201203	371704	201403	365529	201603	357050
201004	421754	201204	474905	201404	458423		
201005	465892	201205	535091	201405	538688		
201006	436336	201206	479152	201406	571525		
201007	564416	201207	641719	201407	624420		
201008	498188	201208	504220	201408	491903		
201009	370731	201209	432767	201409	408271		
201010	369748	201210	431853	201410	413837		
201011	354761	201211	367677	201411	344263		
201012	327270	201212	340391	201412	339776		
201101	307814	201301	310099	201501	319041		
201102	366357	201302	397270	201502	359765		
201103	328646	201303	359413	201503	322175		
201104	443440	201304	444700	201504	401384		
201105	481325	201305	491490	201505	452577		
201106	512256	201306	525579	201506	479740		
201107	620303	201307	612590	201507	515167		
201108	496990	201308	534365	201508	465321		
201109	473684	201309	471324	201509	413534		
201110	437755	201310	423505	201510	434201		
201111	386242	201311	366312	201511	375871		
201112	337529	201312	329276	201512	352845		