

時間序列分析期末報告

來台旅遊人數預測

報告大綱

1. 研究動機
2. 資料介紹
3. 方法與模型
4. 總結
5. 延伸探討
6. 資料來源

一、研究動機

經濟帶動觀光人潮，而觀光收益也會帶動經濟成長，兩者相輔相成，觀光活動絕對是衡量一個國家發展的重要指標之一，也是一個進行國民外交的機會。這篇報告的目的是預測未來一年的外國旅客人潮，藉此擬定對策，像是交通政策、經濟收益的估計。

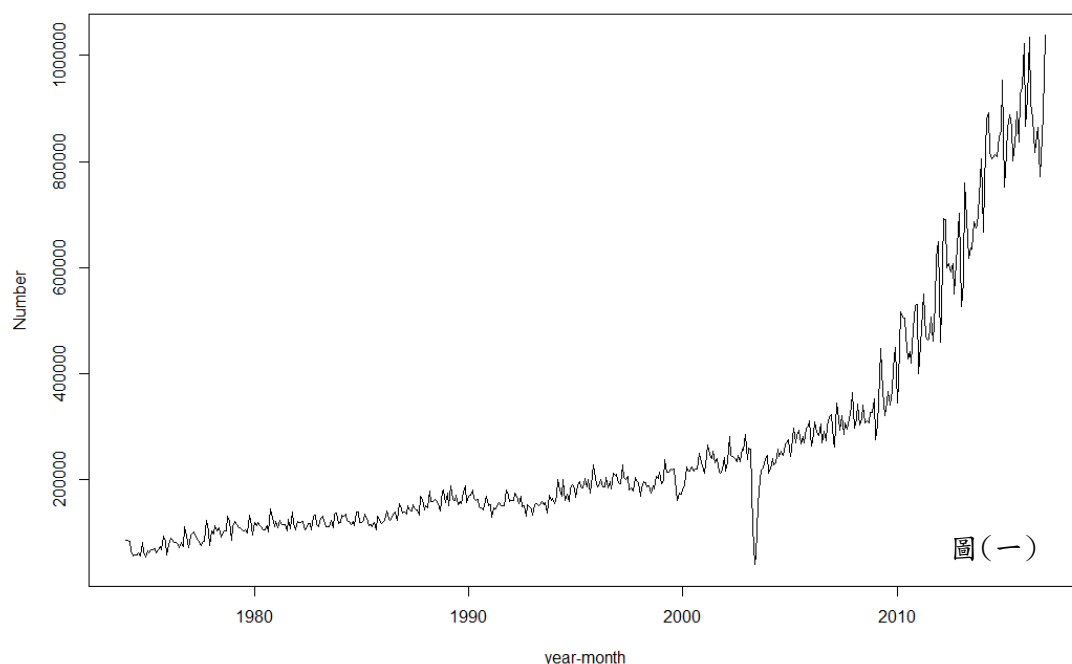
二、資料介紹

資料是由「AREMOS 經濟統計資料庫網路檢索系統」取得，紀錄 1974 年 1 月至 2017 年 12 月每月的來台觀光人數。為了做預測，我們將對 1974 年 1 月至 2016 年 12 月的資料用 ARIMA 建模，並預測 2017 年 12 個月的來台觀光人數，最後再與真實情況比較。

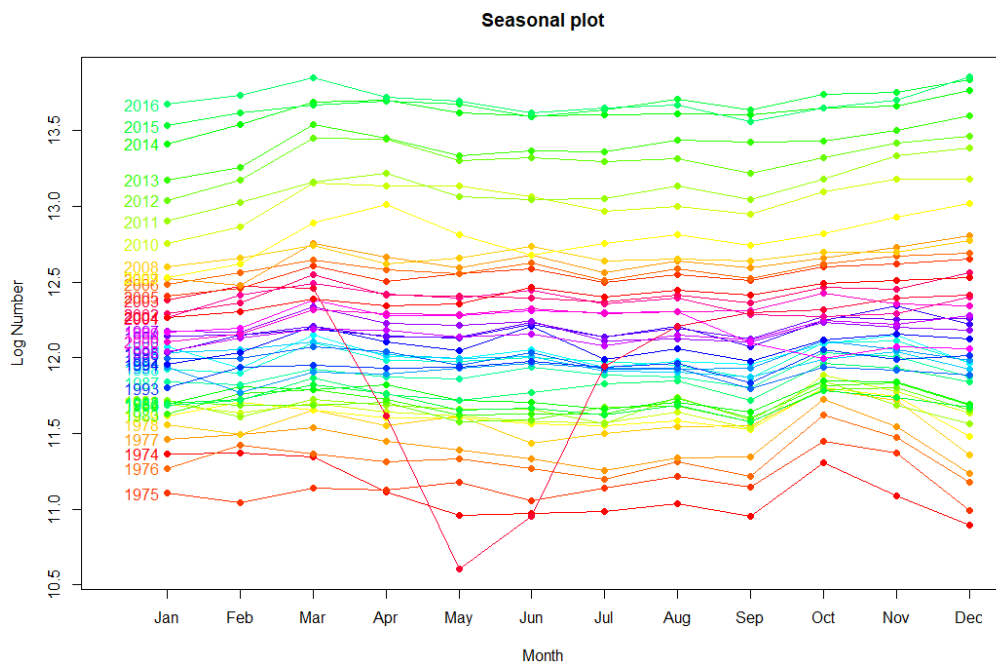
三、方法與模型

方法一：ARIMA

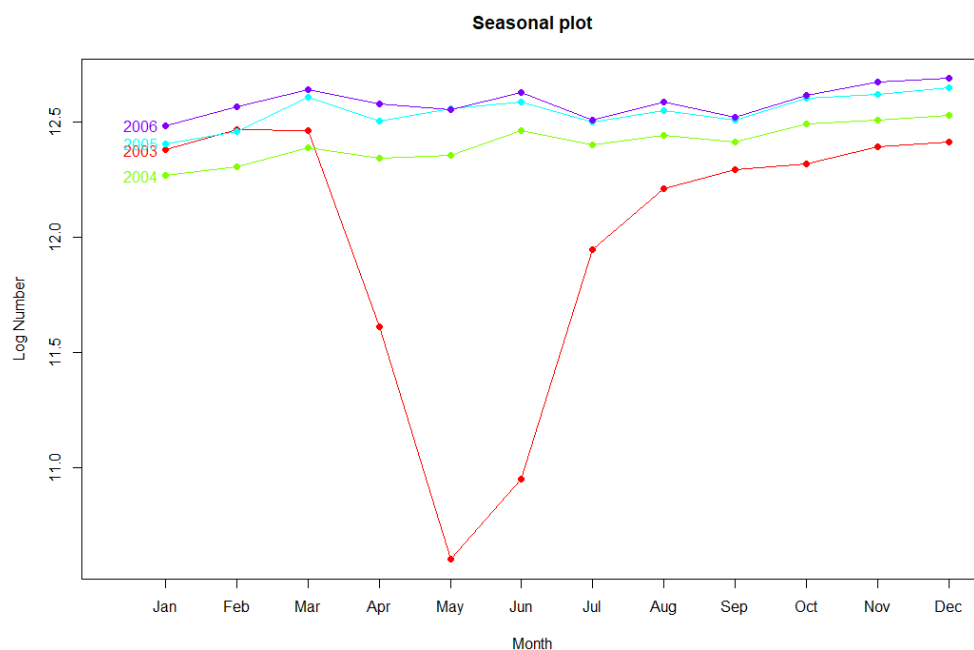
(1) 資料觀察與處理



圖(一)是 1974~2016 歷年逐月的觀光人數圖，可以發現整體有指數上升的趨勢，有必要做差分；變異也有變大的情況，因此會對資料取對數。



圖(二)



圖(三)

從圖(二)可以發現觀光人潮有季節(12 個月)趨勢，夏天可能因為颱風季和高溫導致觀光人數下滑，圖(三)可以看出因為 SARS 疫情導致 2003 年的觀光人數創新低，因此將 2003 年當作特殊情況刪除。觀察資料後，初步的處理為：將資料取對數以及將 2003 年刪除。

(2) 非季節與季節的 AR 和 MA 判別

從圖(一)和圖(二)來看，非季節和季節有差分必要，因此分別做一皆差分並以 Dickey-Fuller Test 檢定差分過後是否為穩態過程。

非季節做 1 次差分

```
Augmented Dickey-Fuller Test

data: diff(dat)
Dickey-Fuller = -11.155, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

季節做 1 次差分

```
Augmented Dickey-Fuller Test

data: diff(dat, lag = 12, differences = 1)
Dickey-Fuller = -5.7482, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

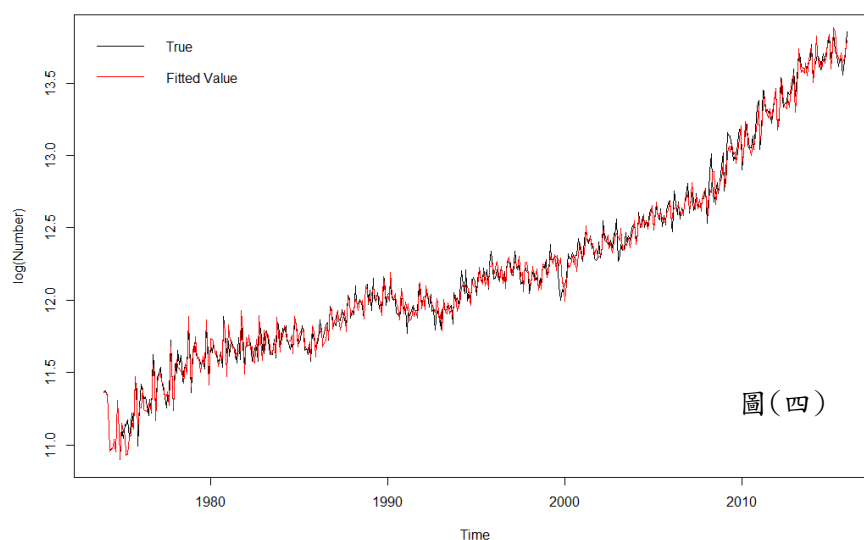
由上述兩個檢定結果，我們可以確認分別做 1 次差分即可。

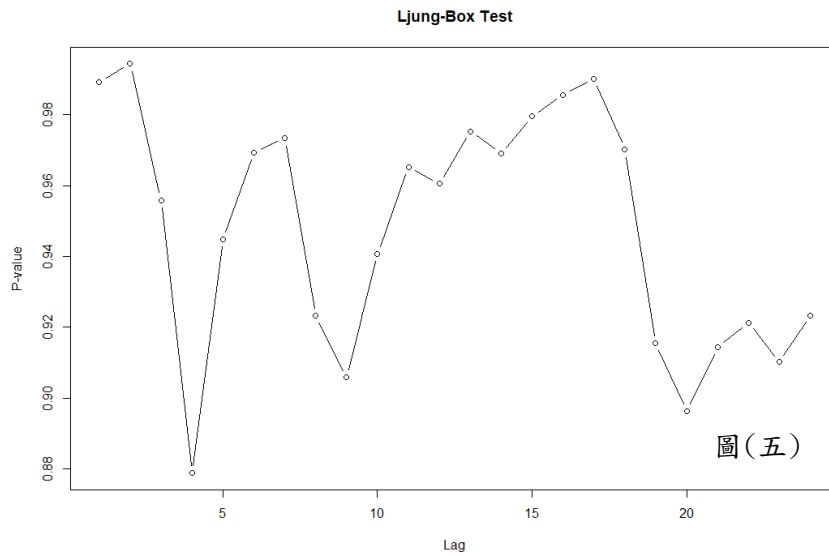
假設模型為： $ARIMA(p, 1, q)(P, 1, Q)_{12}$

至於 AR 和 MA 的判斷，我們使用 R 中的 *auto.arima*，得出好幾種不同組合下的 AIC，最後選擇的最適模型為：

$ARIMA(2,1,3)(1,1,2)_{12}$

(3) 配適檢定



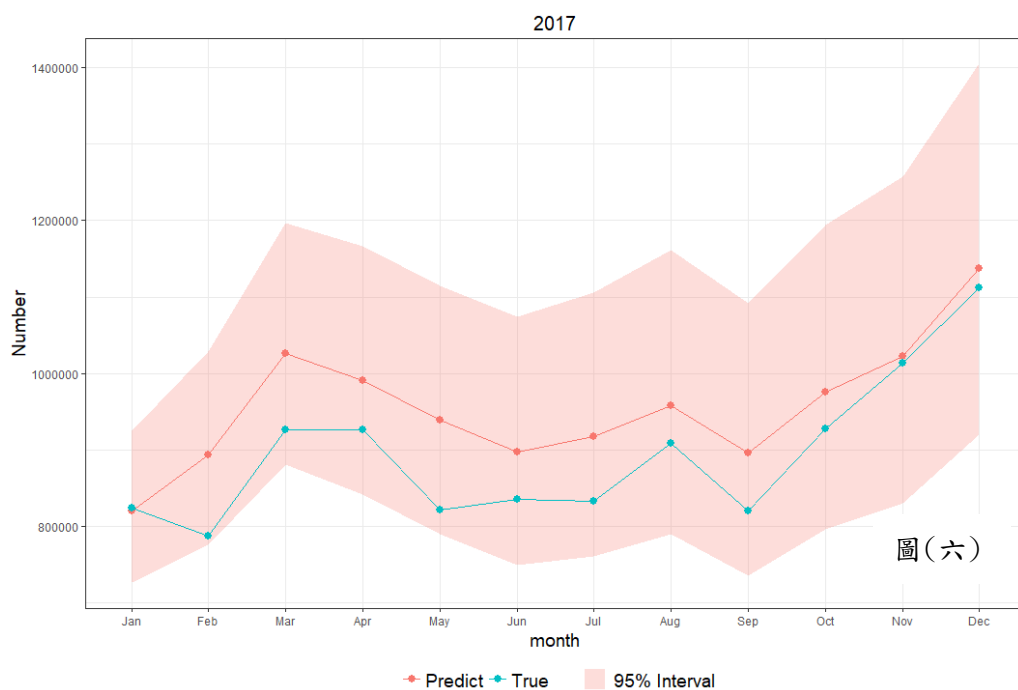


從圖(四)和圖(五)顯示 $ARIMA(2,1,3)(1,1,2)_{12}$ 是合適的模型。

係數為：

	ar1	ar2	ma1	ma2	ma3	sar1	sma1	sma2
	1.0559	-0.1079	-1.4629	0.4416	0.0284	-0.3545	-0.3203	-0.1729
s.e.	0.4400	0.4055	0.4387	0.5936	0.1658	0.3551	0.3570	0.2355

(4) 預測



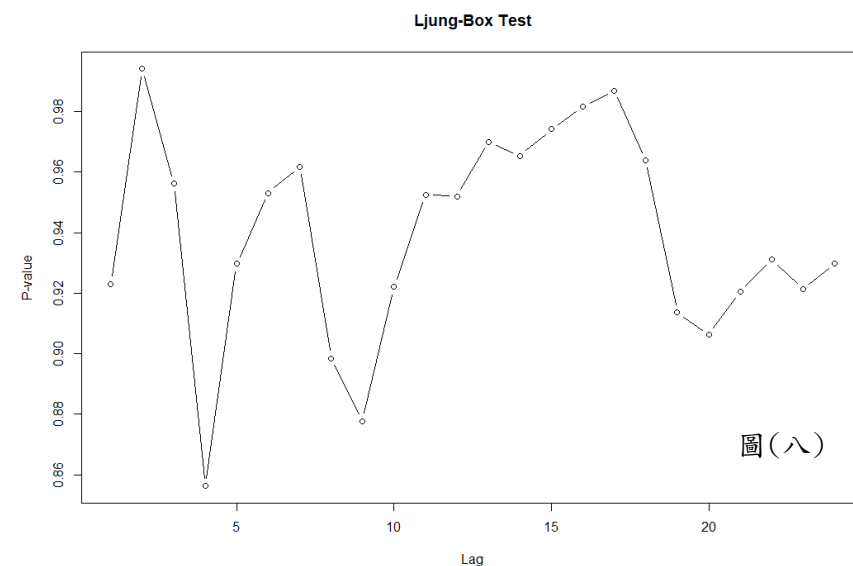
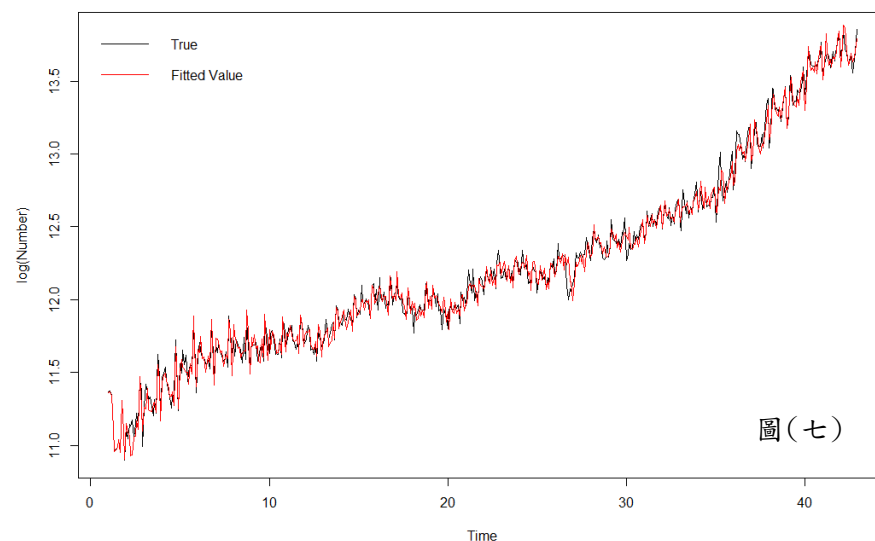
圖(六)為 2017 年來台觀光人數的預測結果，真實值皆落在信賴區間內，趨勢蠻符合真實情況，MAPE 為 7.183217；MPE 為 -7.091451，整體預測情況還算不錯。

方法二：ARIMA+事件分析

2008 年 7 月 4 號，陸客來台旅遊首發團抵達，開啟兩岸觀光發展，而執政黨的政治傾向會影響陸客來台的情況。自開放陸客來台觀光以來，台灣經過一次政黨輪替，我們將政黨當作「事件」並給定兩外生變數，變數一為國民黨，變數二為民進黨，2008 年 7 月至 2016 年 12 月為該執政黨則紀錄為 1，其餘為 0。

模型為 $ARIMA(2,1,3)(1,1,2)_{12}$ ，加上「事件」變數。

(1) 配適檢定



從圖(七)和圖(八)顯示ARIMA(2,1,3)(1,1,2)₁₂ 加上「事件」變數是合適的模型。

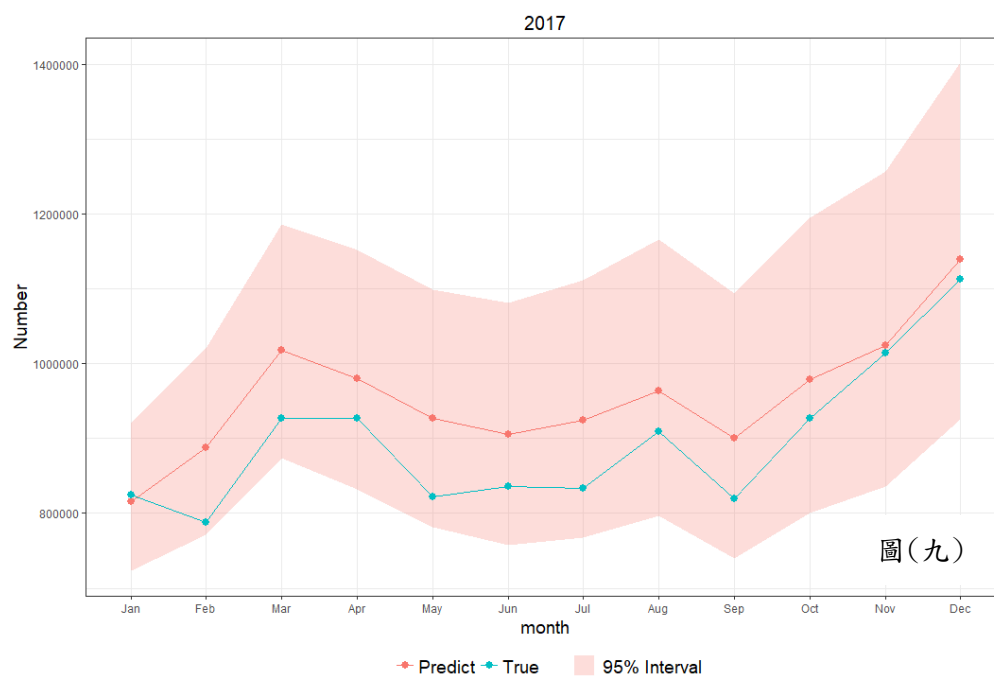
係數為：

	ar1	ar2	ma1	ma2	ma3	sar1	sma1	sma2
	0.9972	-0.0642	-1.4109	0.3814	0.0396	-0.5090	-0.1602	-0.2811
s.e.	0.4284	0.3910	0.4274	0.5754	0.1600	0.3578	0.3735	0.2624

	var1	var2
	0.0528	-0.0183
s.e.	0.0508	0.0785

可以從係數看出國民黨執政時對觀光人數的影響為正；而民進黨執政時對人數的影響為負。

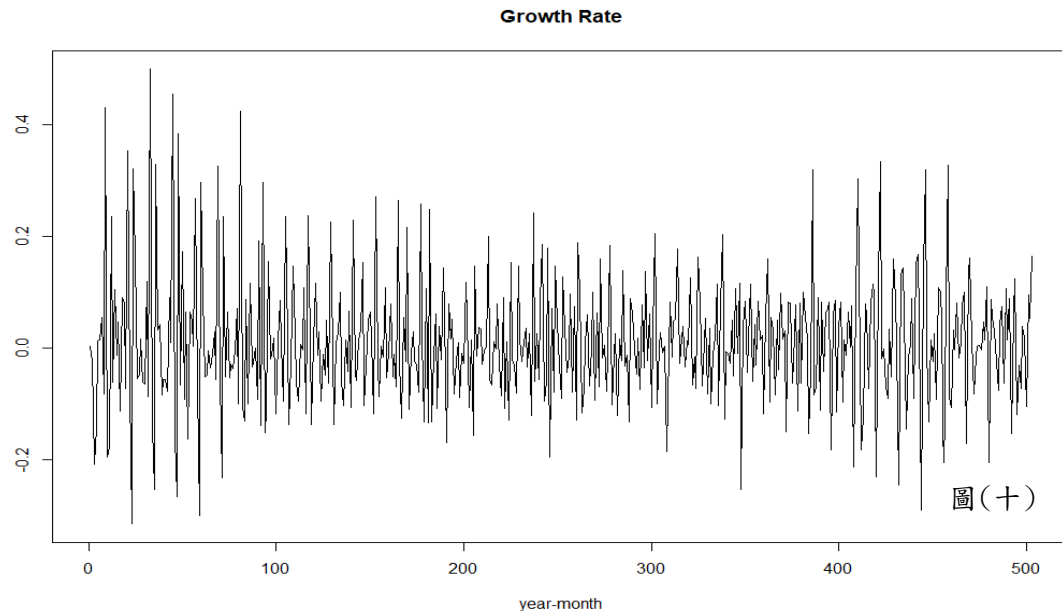
(2) 預測



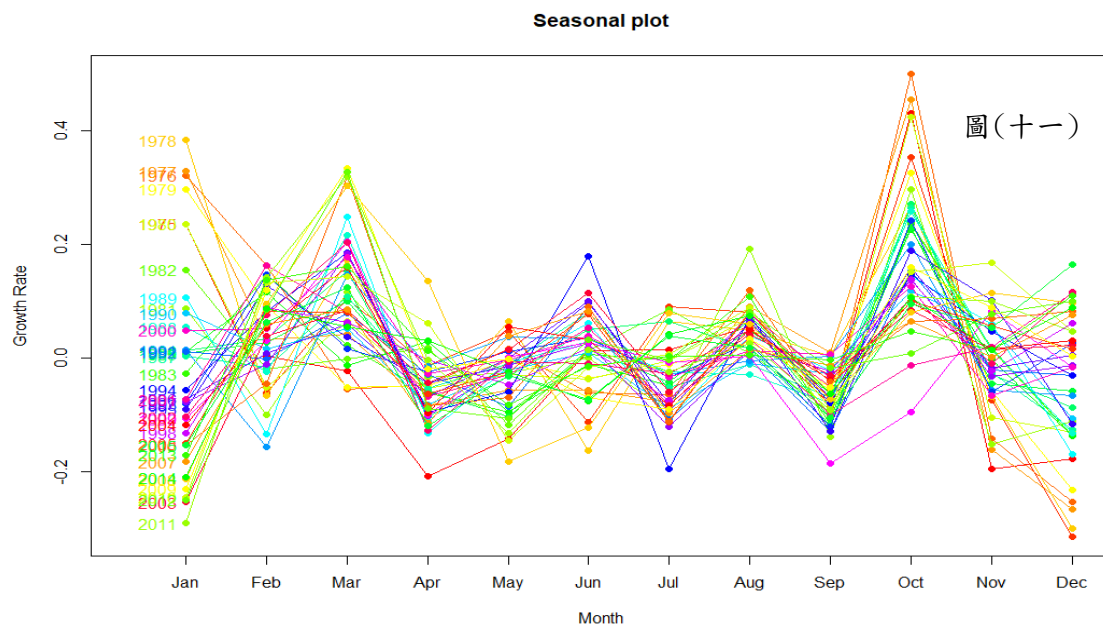
圖(九)為 2017 年來台觀光人數的預測結果，真實值皆落在信賴區間內，趨勢也符合真實情況，MAPE 為 7.14725；MPE 為 -6.975647，整體預測表現比只用 ARIMA model 來得好。

方法三：月成長率的 ARIMA 分析

除了預測分析人數之外，我們也可以對月成長率做預測，再利用前一個月已知的來台觀光人數來估計未來一年的人數。在此也是將 **2003** 年當作離群值刪除。



從圖(十)可以初步判斷歷年逐月的成長率為穩態過程，因此不必做差分。



由圖(十一)可以看出有季節趨勢，而看起來需要一次季節差分。

(1) 非季節與季節的 AR 和 MA 判別

從圖(十)和圖(十一)來看，季節有差分必要，因此做一皆差分並以 Dickey-Fuller Test 檢定差分過後是否為穩態過程。

非季節做 0 次差分

```
Augmented Dickey-Fuller Test

data: growth.rate(tour)
Dickey-Fuller = -10.672, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

季節做 1 次差分

```
Augmented Dickey-Fuller Test

data: diff(growth.rate(tour), lag = 12, differences = 1)
Dickey-Fuller = -9.3674, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```

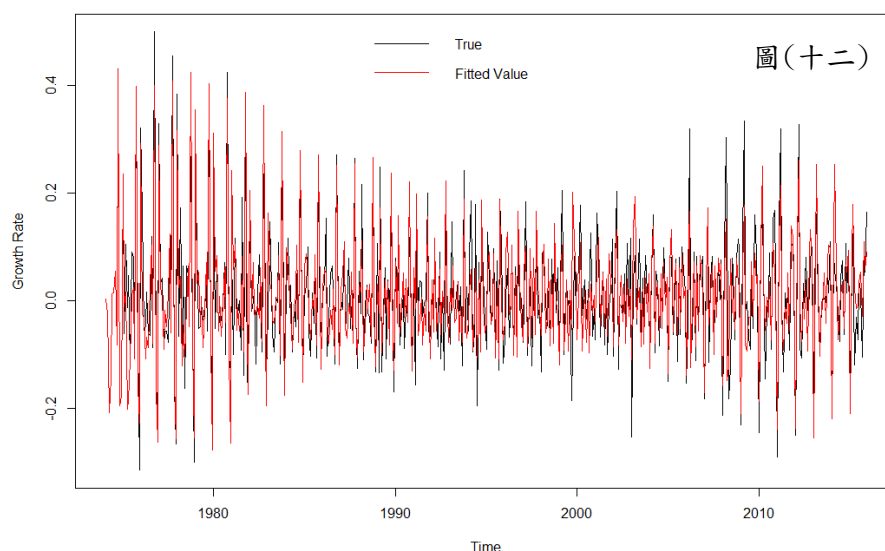
由上述兩個檢定結果，我們可以確認季節做 1 次差分即可。

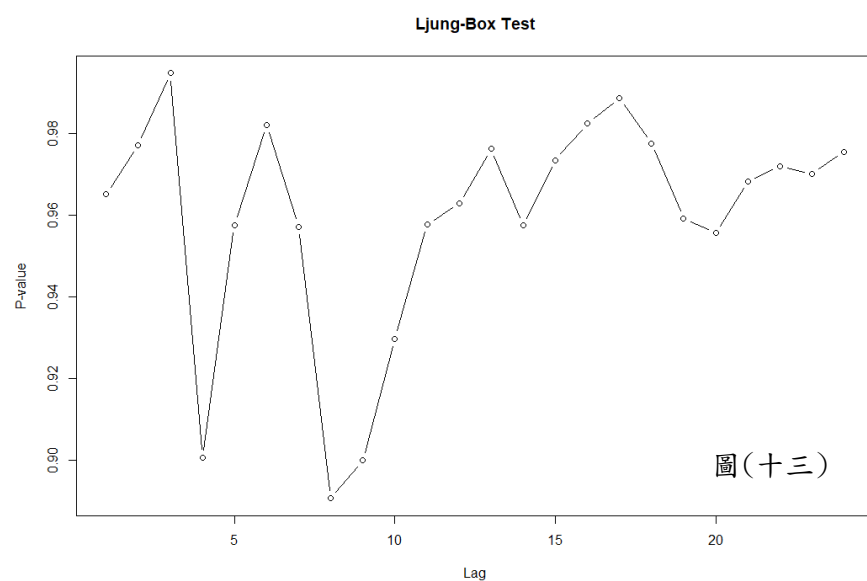
假設模型為： $ARIMA(p, 0, q)(P, 1, Q)_{12}$

至於 AR 和 MA 的判斷，我們使用 R 中的 *auto.arima*，得出好幾種不同組合下的 AIC，最後選擇的最適模型為：

$ARIMA(1,0,2)(1,1,2)_{12}$

(2) 配適檢定





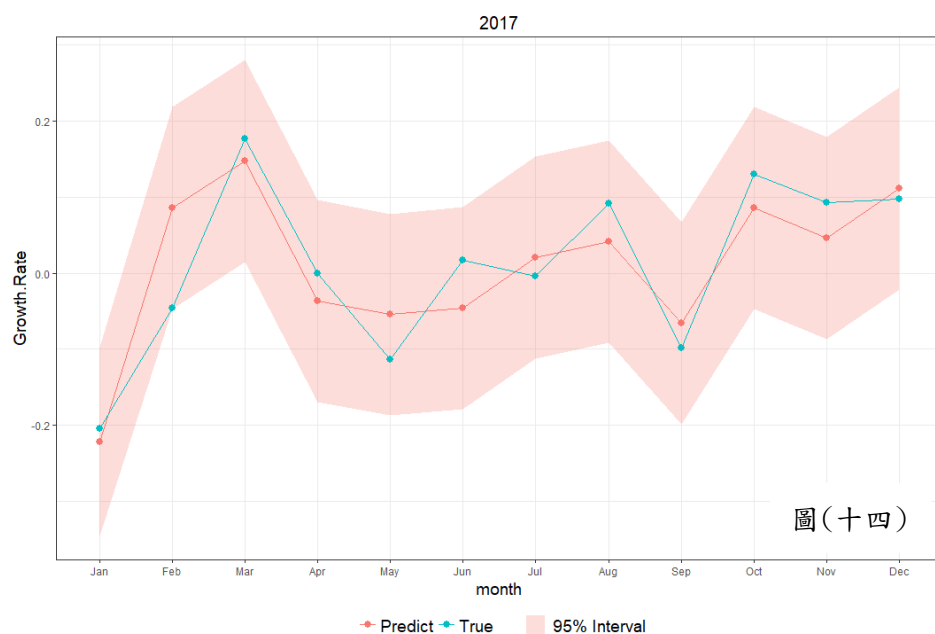
從圖(十二)和圖(十三)顯示 $ARIMA(1,0,2)(1,1,2)_{12}$ 是合適的模型。

係數為：

	ar1	ma1	ma2	sar1	sma1	sma2
	0.0256	-0.6298	0.0284	-0.2961	-0.3738	-0.1395
s.e.	0.5680	0.5700	0.2392	0.3885	0.3886	0.2529

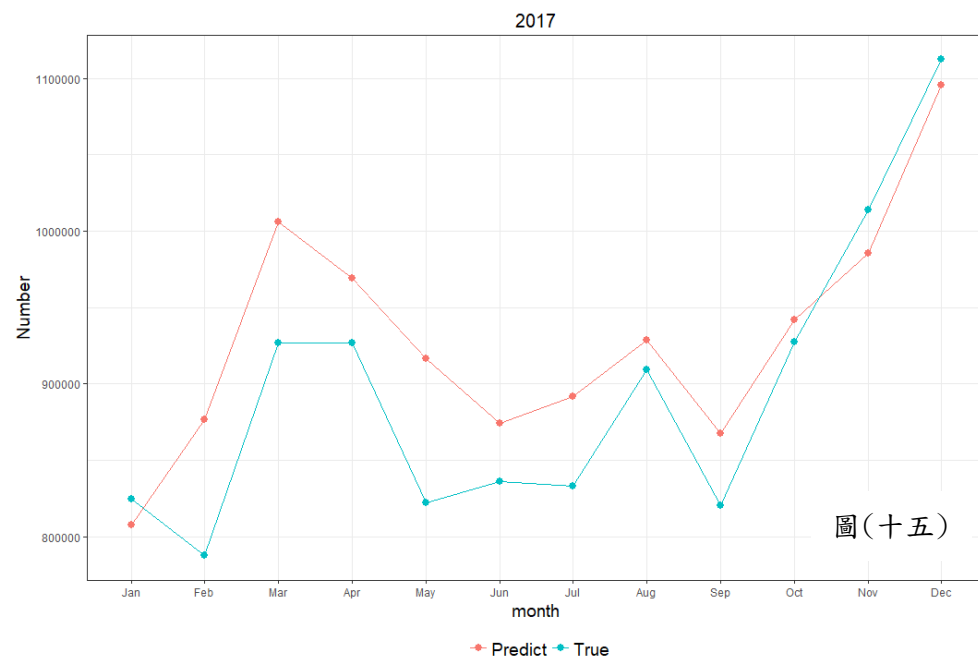
(2) 預測

我們先對 2017 年成長率做預測。



圖(十四)為 2017 年來台觀光人數月成長率的預測結果，真實值皆落在信賴區間內，趨勢也符合真實情況，MSE 為 0.002956672，準確度還不錯。

當得到月成長率的預測值後，我們可以透過 2016 年 12 月的來台觀光人數估算出 2017 年的人數。



圖(十五)

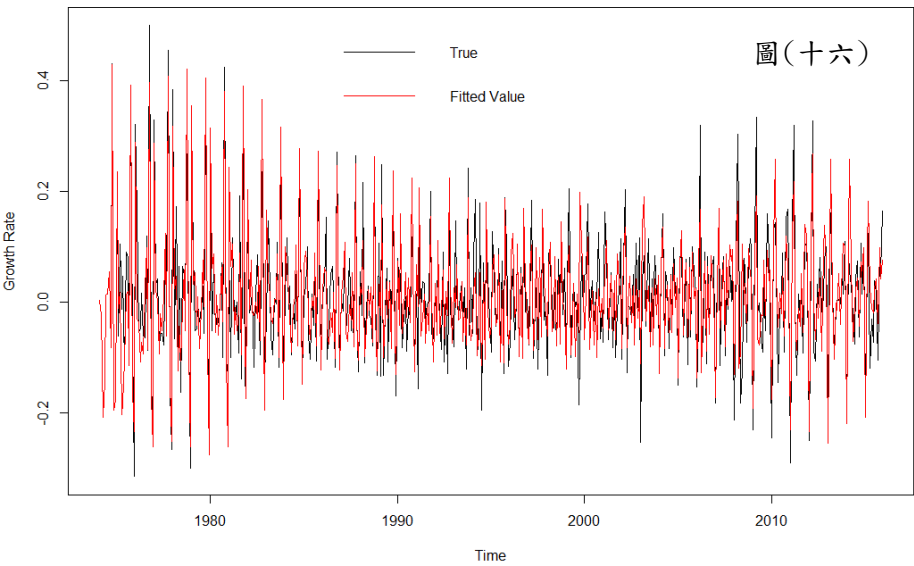
因為成長率預測區間的上界幾乎都大於 0，而下界也幾乎小於 0，導致觀光人數的預測區間會越來越廣，因此圖(十五)不顯示信賴區間。圖(十五)的 MAPE 為 12.29006；MPE 為 -6.920505。

方法四：月成長率的 ARIMA+事件分析

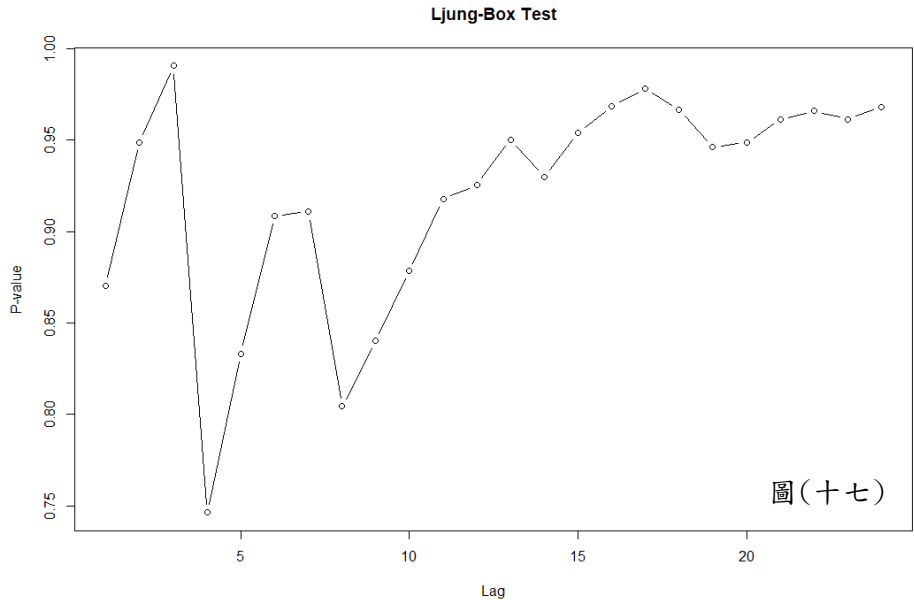
如方法三加入政治因素，將執政黨當作「事件」

模型為 $ARIMA(1,0,2)(1,1,2)_{12}$ ，加上「事件」變數。

(1) 配適檢定



圖(十六)



圖(十七)

從圖(十六)和圖(十七)顯示ARIMA(1,0,2)(1,1,2)₁₂ 加上事件變數是合適的模型。

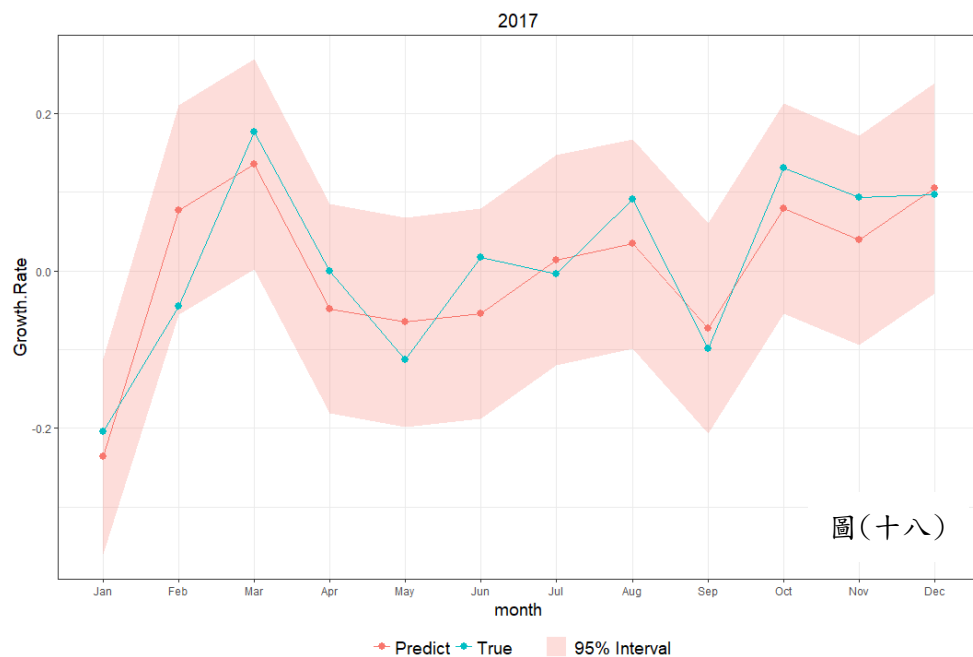
係數為：

	ar1	ma1	ma2	sar1	sma1	sma2
	0.6932	-1.0822	0.2191	-0.3218	-0.3462	-0.1517
s.e.	0.3631	0.3794	0.1937	0.3654	0.3679	0.2411

	var1	var2
	0.0132	0.0027
s.e.	0.0066	0.0160

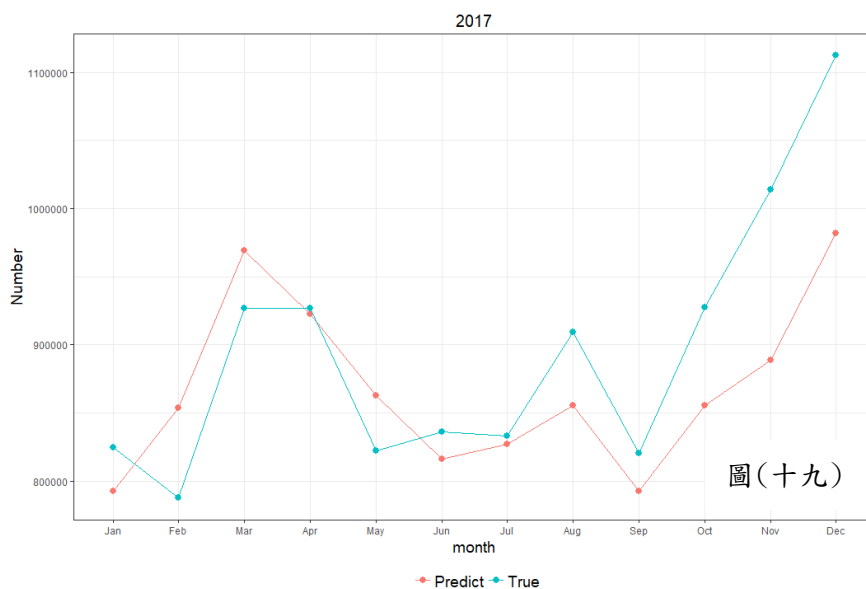
(2) 預測

我們先對 2017 年成長率做預測。



圖(十八)

圖(十八)為 2017 年來台觀光人數月成長率的預測結果，真實值皆落在信賴區間內，趨勢也符合真實情況，MSE 為 0.003094827，準確度比方法三差。



圖(十九)

我們可以透過 2016 年 12 月的來台觀光人數估算出 2017 年的人數。因為成長率預測區間的上界幾乎都大於 0，而下界也幾乎小於 0，導致觀光人數的預測區間會越來越廣，因此圖(十九)不顯示信賴區間。圖(十九)的 MAPE 為 11.57578；MPE 為 -0.0317099。

四、總結

	方法一	方法二	方法三	方法四	最佳模型
MAPE	7.183217	7.14725	12.29006	11.57578	方法二
MPE	-7.091451	-6.975647	-6.920505	-0.0317099	方法四

如果是以 MAPE 衡量，直接預測人數的 ARIMA+事件分析模型表現較好；而如果是 MPE 衡量，透過預測月增率估計人數的 ARIMA+事件分析模型表現較好，端看使用者的選擇。

估計月增率的好處在於可以逐步預測，當 2017 年 1 月的觀光人數為已知，就可以透過月增率預測下個月份，彈性比直接預測人數來得要好；缺點是逐步預測期數較短，而且逐步預測建議還是要重新建模做預測。

五、延伸探討

嘗試以上 4 種 ARIMA 預測方法，雖然預測趨勢有符合期待，但大部份的預測結果都有高估的現象發生。陸客為外國旅客的主要來源，其次是日本，雖然已用事件分析考慮中國因應台灣執政黨所帶來的觀光衝擊，但效果並不顯著。我認為是因為還沒探討其他即時性的因素，像是匯率，也沒考慮到其他主要人數來源國家的經濟狀況，但這些變數也需要預測，如果這些變數估計不準確反而會導致觀光人數的預測更不準確，這是最困難也是需要研究最深的地方。

六、資料來源

AREMOS 經濟統計資料庫網路檢索系統：

<http://net.aremos.org.tw/>