



逢甲大學學生報告 ePaper

來台觀光旅客人數預測分析

作者：廖哲瑋、呂嘉祥、涂凱妮、林君涵、許閔傑、曾鴻儒

系級：統計三乙

學號：D9560996、D9561067、D9527767、D9561377、D9561006、D9561320

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：97 學年度 第二學期

摘要

台灣面臨了東南亞各國的競爭，尤其即將面臨來台觀光人數減少之危機，但危機也就是轉機，觀光業為能永續經營，應積極謀求改善策略。如何在國內觀光市場結構即將面臨重組之際，尋求台灣最理想的發展方向，實為台灣當前的重要課題。因此我們希望能夠應用課程所學之統計預測方法，加以適當分析，以期能預測出未來一年內來台旅遊觀光的人數，並希望這份報告能夠作為觀光局參考。

資料來源為AREMOS經濟統計資料庫，時間為西元1982年1月至2008年12月，並保留最後12筆做預測。首先描述時間序列圖是否呈現趨勢，變異數與平均數是否平穩。接著使用時間序列迴歸法、指數平滑法、分解法及Autoregressive Integrated Moving Average Model (ARIMA)與介入分析配適模型。最後使用Mean Squared Error(MSE)、Mean Absolute Deviation(MAD)、Mean Percentage Error(MPE)及Mean Absolute Percentage Error (MAPE)四個準則，來評估這四種方法何者最佳，並選出最佳模型。分析結果顯示，ARIMA之介入分析的配適模型為最佳模型。

關鍵字：介入分析、ARIMA、觀光、統計預測

目 次

第一章 緒論	5
第一節 研究動機	5
第二節 研究目的	5
第三節 研究背景	6
第四節 資料描述	6
第五節 研究流程	8
第二章 研究方法	9
研究模型之探討：	9
第一節 時間序列迴歸法 Time Series Regression	9
第二節 指數平滑法 Exponential Smoothing	15
第三節 分解法 Decomposition method	19
第四節 ARIMA 之介入分析	25
第五節 最佳模型	36
第三章 結論與建議	37
參考文獻	40

表 附 錄

(表 1-1)模型的參數估計表.....	11
(表 1-2) 配適一階自我相關模型的參數估計表	12
(表 1-3) 時間序列法實際值與預測表.....	14
(表 2-1) Winter Method –Additive 參數估計表	17
(表 2-2)指數平滑法實際值與預測表	18
(表 3-1) 去季節因子的參數估計表	20
(表 3-2)DY 模型的參數估計表	22
(表 3-3)DY 配適一階自我相關模型的參數估計表	22
(表 3-4)DY 配適二階自我相關模型的參數估計表	22
(表 3-5) DY 配適一階自我相關模型與 T^2 項的參數估計表	23
(表 3-6) DY 配適二階自我相關模型與 T^2 項的參數估計表	23
(表 3-7)分解法實際值與預測表.....	24
(表 4-1) Ljung-Box 自我相關檢測圖	30
(表 4-2)參數估計值.....	31
(表 4-3) ARIMA 之介入分析-Ljung-Box 自我相關檢測圖	33
(表 4-4) ARIMA 之介入分析-參數估計表	34
(表 4-5) ARIMA 之介入分析-實際值與預測表	35
(表 5-1)模式比較	37

圖 附 錄

(圖 1)原始時間序列圖.....	7
(圖 2)來台觀光人口	13
(圖 1-1)時間序列法實際值與預測圖	14
(圖 2-1) 指數平滑法 ACF 與 PACF 圖	16
(圖 2-2) 指數平滑法實際值與預測圖.....	18
(圖 3-1) 去季節因子的時間序列圖	19
(圖 3-2) 趨勢—循環時間序列圖	20
(圖 3-3)季節因子之時間序列圖.....	20
(圖 3-4)不規則項之時間序列圖.....	21
(圖 3-5) 分解法實際值與預測圖	25
(圖 4-1)原始的 ACF 圖與 PACF 圖	26
(圖 4-2)季節差分之 ACF 圖與 PACF 圖.....	27
(圖 4-3)配完模式後之 ACF 圖與 PACF 圖	28
(圖 4-4)White Noise 檢定 (圖 4-5)單根檢定	30
(圖 4-6)ARIMA 之介入分析-原始時間序列圖	31
(圖 4-7)ARIMA 之介入分析-配完模式後之 ACF 圖與 PACF 圖.....	33
(圖 4-8)ARIMA 之介入分析-實際值與預測表	35

第一章 緒論

第一節 研究動機

觀光在台灣的經濟這一環佔了很重要的一部份，觀光業可以說是目前最富生機的產業，一個地區是否有觀光客在於交通的便利性與觀光品質，這些都是不可或缺的。自民國八十三年台灣對日、美、英等十五國實施免簽證措施，來台觀光人數突破兩百萬，但突然的九二一大地震的影響，使得來台觀光人數受到衝擊，經過一段時間的沉寂，很快又回穩了來台觀光的情況，台灣一直都在積極的重建，加強國際觀光宣傳，但是在民國九十二年遇到了SARS，嚴重影響了來台觀光人數。為了復甦受到SARS疫情嚴重影響之觀光產業，政府也提出了許多計畫，來挽救來台觀光人數，這些計畫在之後也有明顯的成效。近年來大陸人民海外旅遊風氣漸盛，兩岸文化、與經貿之往來，可能再創來台觀光人數的高峰。

第二節 研究目的

台灣觀光資源非常豐富，政府也推出了觀光客倍增計畫，就是希望國際觀光客多來。但台灣觀光還面臨了東南亞各國的競爭，東南亞各國會用低廉的價錢搶佔台灣的市場，所以如何在這些國家中使來台觀光人數增加，讓觀光有個新的定位，實為觀光的當前課題。我們希望能用統計預測方法所學到的，選擇合適的時間序列之統計模式，並希望可以將這份報告作為觀光事業的參考標準，使其能預先知道未來一年來台觀光人數的情形，提早想出對策，讓來台觀光的人數能有所提升。

第三節 研究背景

現今互動頻繁、交通便利的時代。台灣，從以前四面環海的封閉小島慢慢變成對外交易頻繁的國家，也吸引了各國的人民，想要踏上這塊曾被譽為「福爾摩沙」的土地。隨著近年來與各國之間貿易頻繁、出遊費用大打折扣的情況下，人們有多餘的金錢想要追求身心上的放鬆，所以來台觀光的人數來有漸漸上升的趨勢，也為台灣帶來不少的收入、讓世界可以多認識這塊土地的美麗。此外，影響各國人民來台旅遊的其他因素也包含了天災及生活環境等因素，我們發現在本研究資料的期間內發生了兩件重大事件：1999年9月921大地震、2003年4月SARS肆虐，由於地震的影響，導致房屋倒塌、人民傷亡、旅遊景點被損壞，而影響旅客來台觀光的興致。2003年的SARS事件更是影響了所有大眾運輸工具的載客數，使得各國來台觀光的人劇減，因此，讓我們想深入去探討這兩起事件到底影響來台觀光的人數有多嚴重。

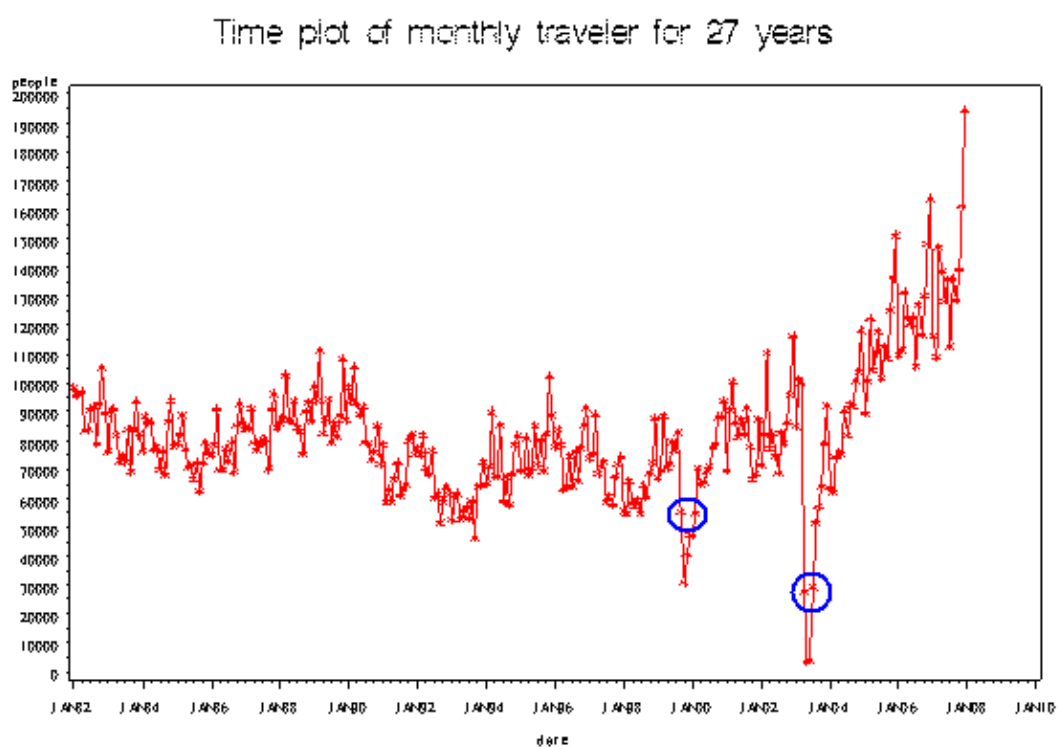
第四節 資料描述

此筆資料主要是記錄每月來台觀光旅客的人數。分析時間為西元1982年1月至2008年12月，並保留最後12筆做預測。資料來源為AREMOS經濟統計資料庫。如果想要了解更詳細的相關資料，可參考AREMOS經濟統計資料庫網站(http://cache.moe.edu.tw/aremos_ly/search.html)。

首先描述時間序列圖(圖1)，座標X軸為日期，從1982年1月開始，至2008年12月為止；座標Y軸為來台觀光旅客人數，以萬人為單位。可看出此台觀光旅客人數依時間有逐漸增加的長期變化之趨勢，表示在經過921大地震與SARS後，來台觀光的人數明顯的呈現上升的趨勢，圖中有兩個時間點較特別，一個是1999年9月受921大地震的影響，一個是2003年4月受到SARS

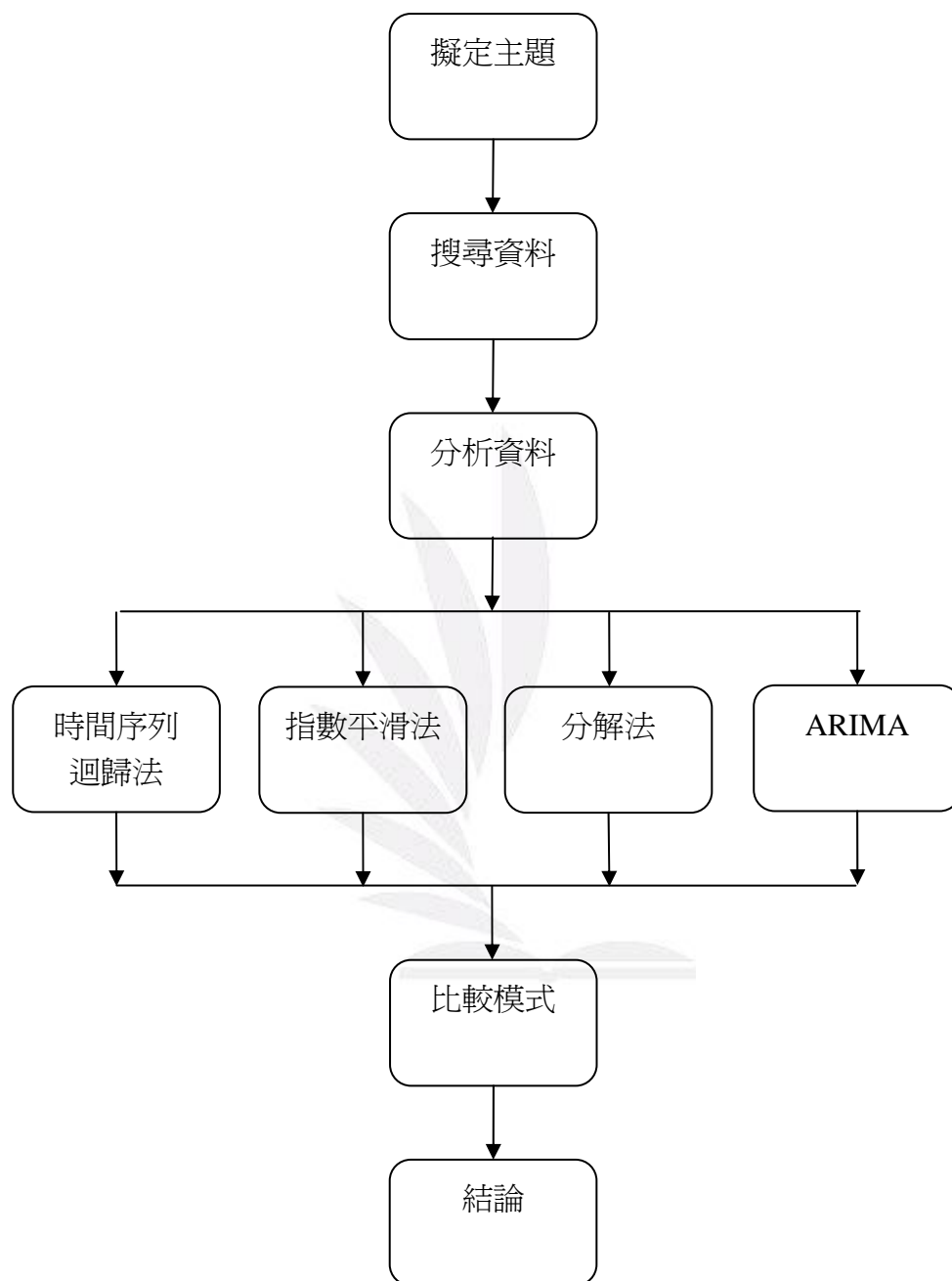
來台觀光旅客人數預測分析

疫情的影響，使得來台觀光的人數縮減，相對於以往相同月份的人數低很多。



(圖 1)原始時間序列圖

第五節 研究流程



第二章 研究方法

研究模型之探討：

第一節 時間序列迴歸法 Time Series Regression

本研究的資料是來台旅客觀光人數，於(圖1)中可判斷出變異數呈現平穩的狀態，不需再多做轉換，在人數方面可看出在每年三月與十二月左右較其他月份高，所以該時間序列圖是存在季節因子。在1999年9月受到921大地震與2003年4月受到SARS疫情的影響，使得來台觀光的人數縮減，所以我們將設季節虛擬變數和虛擬變 P_1 、 P_2 加入我們的模式中。

建立模型

$$y_t = TR_t + SN_t + \varepsilon_t \quad \text{其中 } y_t = \text{時間序列在時間 } t \text{ 的觀察值}$$

$TR_t = \text{時間 } t \text{ 的趨勢項}$

$SN_t = \text{時間 } t \text{ 的季節因子}$

$\varepsilon_t = \text{時間 } t \text{ 的誤差項(不規則因子)}$

$$y_t = \beta_0 + \beta_1 t + \beta_2 M_1 + \beta_3 M_2 + \beta_4 M_3 + \beta_5 M_4 + \beta_6 M_5 + \beta_7 M_6 + \beta_8 M_7 \\ + \beta_9 M_8 + \beta_{10} M_9 + \beta_{11} M_{10} + \beta_{12} M_{11} + \beta_{13} P_1 + \beta_{14} P_2 + \varepsilon_t$$

季節虛擬變數 M ，定義如下：

t 為時間， y_t 為在連續時間下，第 t 點所觀察到的值

M_1, M_2, \dots, M_{11} 為季節虛擬變數

$$M_1 = \begin{cases} 1 & \text{if period } t \text{ is January} \\ 0 & \text{otherwise} \end{cases} \quad \dots \quad M_{11} = \begin{cases} 1 & \text{if period } t \text{ is November} \\ 0 & \text{otherwise} \end{cases}$$

P_1, P_2 為重大事件影響點

$$P_1 = \begin{cases} 1 & \text{if period } t \text{ is September 1999} \\ 0 & \text{otherwise} \end{cases} \quad P_2 = \begin{cases} 1 & \text{if period } t \text{ is April 2004} \\ 0 & \text{otherwise} \end{cases}$$

Durbin-Watson檢定

Durbin-Watson檢定 (D-W統計量)是用來檢定資料是否有自我相關的問題，亦即殘差是否為獨立、抑或具有某種規則性趨勢，如果殘差獨立無任何趨勢存在時，我們說誤差項是隨機的。由殘差項的自我相關檢測表(表1-1)可看出， $Pr > DW$ 的P-value都大於0.05水準，但 $Pr < DW$ 的P-value值小於0.05水準，所以此模式不恰當，故配適一階自我相關模型，而參數估計 P_1 的P-value大於0.05水準，所以不顯著，我們將其參數刪除。

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho > 0 \quad \text{or} \\ H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho < 0$$

$$d(\equiv DW) = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2\sum e_t e_{t-1}}{\sum e_t^2}$$

$$\cong 2(1 - \hat{\rho}) \quad (\text{if } \sum e_t^2 = \sum e_{t-1}^2)$$

$$0 \leq |\rho| < 1 \quad d \cong 2(1 - \hat{\rho})$$

$$\text{when } \hat{\rho} = 1 \quad \Rightarrow \quad d = 0$$

$$\text{when } \hat{\rho} = -1 \quad \Rightarrow \quad d = 4$$

$$\text{when } \hat{\rho} = 0 \quad \Rightarrow \quad d = 2$$

(表 1-1)模型的參數估計表

The AUTOREG Procedure

Dependent Variable y
 people

Ordinary Least Squares Estimates

SSE	1.29661E11	DFE	297
MSE	436568553	Root MSE	20894
SBC	7163.26039	AIC	7107.11534
Regress R-Square	0.1871	Total R-Square	0.1871
Durbin-Watson	0.2632	Pr < DW	<.0001
Pr > DW	1.0000		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	82677	4621	17.89	<.0001
t	1	69.3296	13.1795	5.26	<.0001
d1	1	-14477	5797	-2.50	0.0131
d2	1	-11313	5797	-1.95	0.0519
d3	1	-168.0340	5796	-0.03	0.9769
d4	1	-12828	5855	-2.19	0.0292
d5	1	-17349	5796	-2.99	0.0030
d6	1	-12247	5796	-2.11	0.0354
d7	1	-20056	5795	-3.46	0.0006
d8	1	-11973	5795	-2.07	0.0397
d9	1	-18993	5853	-3.25	0.0013
d10	1	-10362	5795	-1.79	0.0748
d11	1	-1723	5795	-0.30	0.7665
p1	1	-23038	21321	-1.08	0.2808
p2	1	-59660	21354	-2.79	0.0055

配適一階自我相關後，發現Pr<DW和Pr>DW的P-value都大於0.05水準，所以無自我相關(表1-2)。故我們最後模式如下：其模式為：

$$y_t = \beta_0 + \beta_1 t + \beta_2 M_1 + \beta_3 M_2 + \beta_4 M_3 + \beta_5 M_4 + \beta_6 M_5 + \beta_7 M_6 + \beta_8 M_7 + \beta_9 M_8 + \beta_{10} M_9 + \beta_{11} M_{10} + \beta_{12} M_{11} + \beta_{13} P_2 + \varepsilon_t$$

(模型 1.)

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + a_t$$

預測方程式如下：

$$y_t = 81461 + 86.48t - 12511M_1 - 9840M_2 + 898.1483M_3 - 13671M_4 - 16881M_5 - 11987M_6 - 19953M_7 - 11979M_8 - 19950M_9 - 10453M_{10} - 1790M_{11} - 18634P_2 + \varepsilon_t$$

$$\varepsilon_t = 0.840603\varepsilon_{t-1} + a_t$$

(表 1-2) 配適一階自我相關模型的參數估計表

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.840603	0.031430	-26.75

Yule-Walker Estimates

SSE	2.88719E10	DFE	297
MSE	87211950	Root MSE	9880
SBC	6695.8465	AIC	6699.70145
Regress R-Square	0.4223	Total R-Square	0.8190
Durbin-Watson	1.8848	Pr < DW	0.1855
Pr > DW	0.8345		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

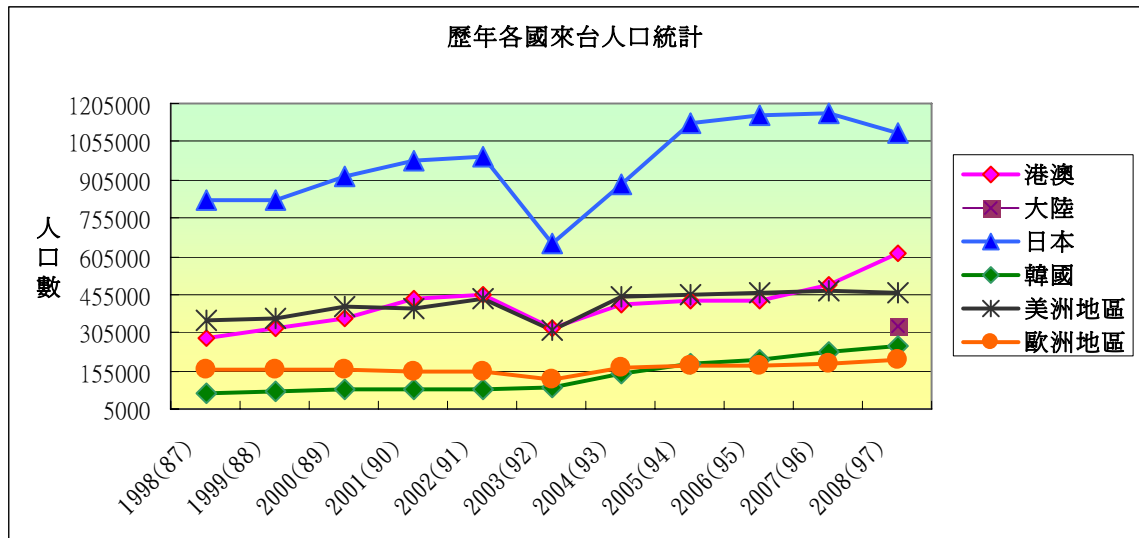
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	81461	7041	11.57	<.0001
t	1	86.4800	37.0024	2.34	0.0201
d1	1	-12511	2021	-6.19	<.0001
d2	1	-9840	2671	-3.68	0.0003
d3	1	898.1483	3066	0.29	0.7698
d4	1	-13671	3324	-4.11	<.0001
d5	1	-16881	3446	-4.90	<.0001
d6	1	-11987	3488	-3.44	0.0007
d7	1	-19953	3441	-5.80	<.0001
d8	1	-11979	3302	-3.63	0.0003
d9	1	-19950	3051	-6.54	<.0001
d10	1	-10453	2649	-3.95	<.0001
d11	1	-1790	1887	-0.90	0.3682
p2	1	-18634	7697	-2.42	0.0161

預測

我們事先保留了12筆的真實值，我們將預測出來12筆預測值、12筆真實值與95%信賴區間的上下界限整理如表1-3。

當中預測方程式可看出每月的觀光人口都有下降的趨勢，但是在三月的影響力卻高達到900多，其原因在亞洲鄰近國家，如日本春假是在二月上旬至四月上旬，2007年來台旅遊的日本觀光客達138萬人，在日本，平均每5位觀光客就有1位是台灣人，台日在地理位置及文化交流上比起各國更為親近，因為假期較短，人們可能會選擇較鄰近的國家，美麗且擁有著名小吃的台灣當然就會在考慮當中；而對於其他國家來說三、四月在寒假過後就是旅遊淡季，旅遊業價格也相對較便宜，且台灣又屬於氣候宜人的地方，在短期的旅遊中，應該會是旅遊考慮的要素，判斷可能因為這些原因，使得三月份來台

觀光人口數為正。



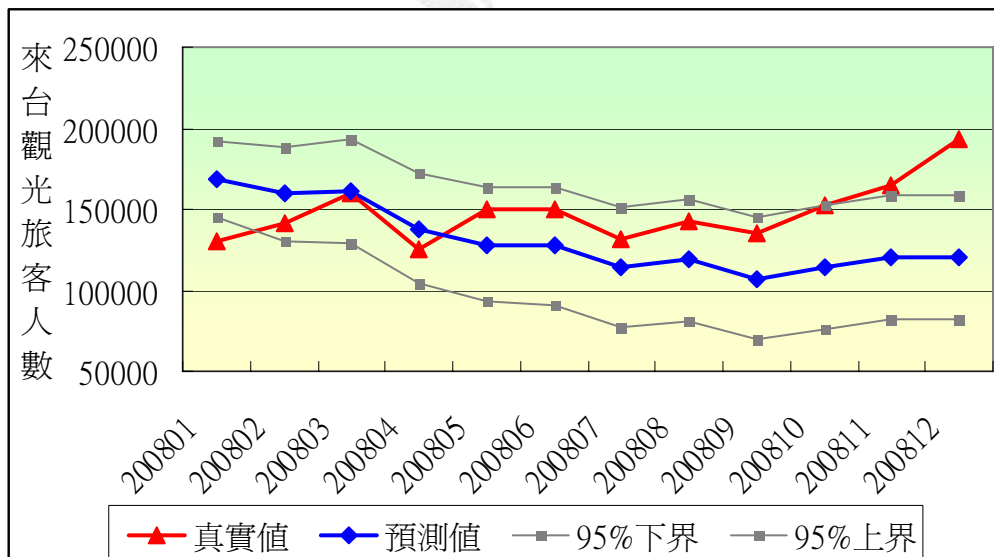
(圖 2)來台觀光人口

由內政部主計處的過去統計資料顯示，來台觀光人口，的確是以日本地區人民佔大半比例，而中國大陸部份在三通後，在觀光業也開始準備嶄露頭角，這也是觀光業目前注意的動向。

從圖1-1的預測圖可以看出真實值大多都在95%信賴區間內，但是第一筆與最後兩筆觀察值是落在區間之外的，為2008年1月及2008年11月，2008年12月，還有一筆觀測值是落在95%信賴區間的邊界附近，為2008年10月，從整體看來，預測值較真實值為低，真實值趨勢是上升，但預測值趨勢是下降，綜合以上幾點，此模式的預測能力可能還有待確定。

(表 1-3) 時間序列法實際值與預測表

	真實值	預測值	95%下界	95%上界
200801	129827	168372.9	144529.4	192216.4
200802	141667	159597.2	130669.4	188525
200803	159366	160727.4	128684.7	192770.2
200804	124999	138094.9	104011.2	172178.7
200805	149415	128121.4	92672.34	163570.4
200806	149473	127343	90951.45	163734.6
200807	131359	114623.1	77575.5	151670.7
200808	143140	118614.4	81105.46	156123.4
200809	134740	107309.9	69473.4	145146.4
200810	152998	114017.7	75946.18	152089.2
200811	164530	120350	82107.59	158592.3
200812	193715	120195.1	81826.48	158563.7



(圖 1-1)時間序列法實際值與預測圖

第二節 指數平滑法 Exponential Smoothing

指數平滑法是利用過去時間序列的加權平均值用來平滑資料的方法，並利用該加權平均值做為下一期的預測值。如果季節的波動沒有變異、保持一致時，通常會對資料配適加法季節模式；但若遇上季節波動有越來越大或越來越小的趨勢時，通常會對資料配適乘法季節模式。

由於我們的資料具有季節變動，而且變異數變動判斷趨近一致，因此我們選擇 Additive Holt-Winters model。

預設模式如下：

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

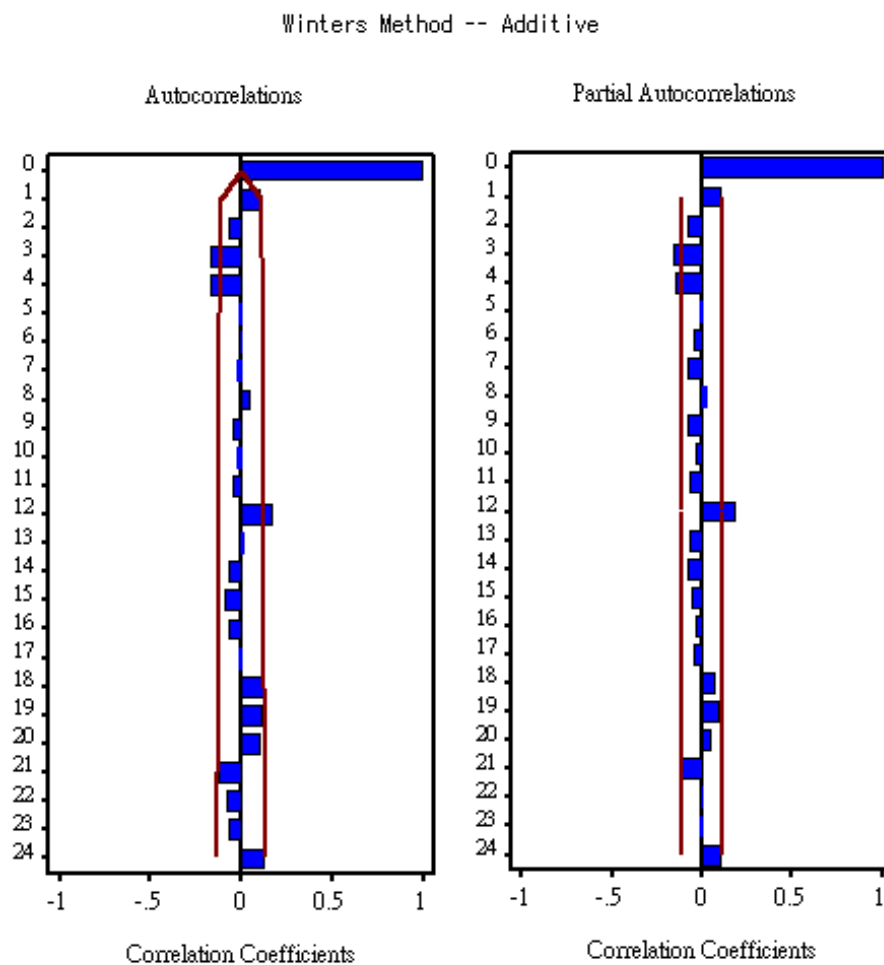
$$b_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)b_{t-1}$$

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

$$F_{t+m} = (L_t + b_t m) S_{t-s+m} \quad (\text{模型 2.})$$

從圖2-1大部分的值皆在兩倍標準差以內，但在lag12的地方有超過，所以我們要再做進一步的分析，才能判定此模式是否適合。從表2-1可以知道參數估計如下：

$$\hat{\alpha} = 0.85936 \quad \hat{\gamma} = 0.001 \quad \hat{\delta} = 0.999$$



(圖 2-1) 指數平滑法 ACF 與 PACF 圖

(表 2-1) Winter Method –Additive 參數估計表

Model Parameter	Estimate	Std. Error	T	Prob> T
LEVEL Smoothing Weight	0.85936	0.0385	22.3417	<.0001
TREND Smoothing Weight	0.00100	0.0157	0.0635	0.9494
SEASONAL Smoothing Weight	0.99900	0.3084	3.2396	0.0013
Residual Variance (sigma squared)	83410512	.	.	.
Smoothed Level	174233	.	.	.
Smoothed Trend	161.07018	.	.	.
Smoothed Seasonal Factor 1	-13192	.	.	.
Smoothed Seasonal Factor 2	-1854	.	.	.
Smoothed Seasonal Factor 3	18588	.	.	.
Smoothed Seasonal Factor 4	-3959	.	.	.
Smoothed Seasonal Factor 5	-8965	.	.	.
Smoothed Seasonal Factor 6	-1432	.	.	.
Smoothed Seasonal Factor 7	-9590	.	.	.
Smoothed Seasonal Factor 8	5837	.	.	.
Smoothed Seasonal Factor 9	-5653	.	.	.
Smoothed Seasonal Factor 10	-676.06302	.	.	.
Smoothed Seasonal Factor 11	8557	.	.	.
Smoothed Seasonal Factor 12	20282	.	.	.

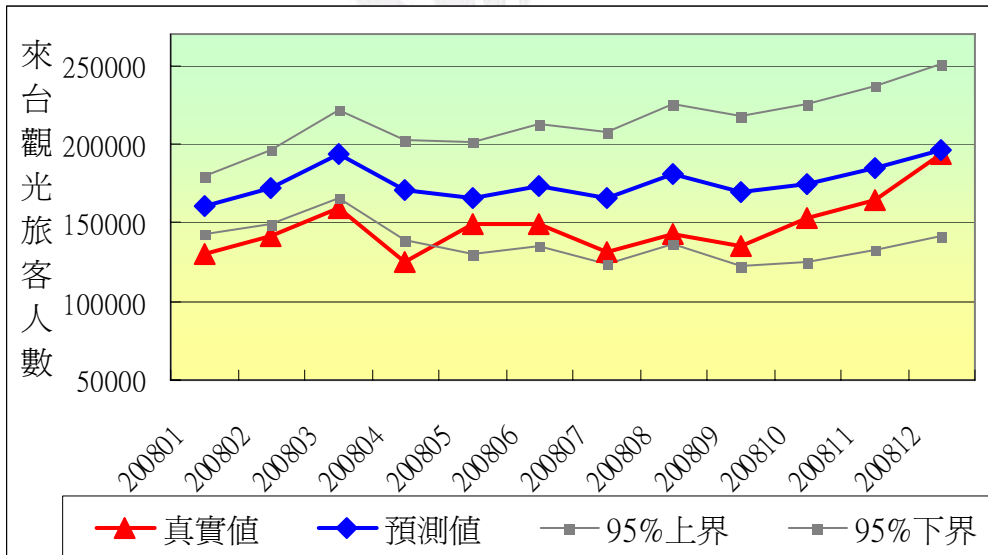
預測

我們事先保留了12筆的真實值，我們將預測出來12筆預測值、12筆真實值與95%信賴區間的上下界限整理如表2-2。

從圖2-2的預測圖可以看出實際值大多都在95%信賴區間內，但是前四筆觀察值是落在區間之外的，分別為2008年1月、2月、3月與4月還有兩筆觀測值是落在95%信賴區間的邊界附近，分別為2008年7月與8月；另外，再從整體看來，預測值較實際值為高。綜合以上幾點，此模式的預測能力可能還有待確認。

(表 2-2)指數平滑法實際值與預測表

	真實值	預測值	95%下界	95%上界
200801	129827	161202	143302	179102
200802	141667	172701	149089	196313
200803	159366	193304	165107	221502
200804	124999	170919	138776	203061
200805	149415	166073	130412	201734
200806	149473	173767	134900	212635
200807	131359	165771	123936	207606
200808	143140	181359	136748	225969
200809	134740	170030	122801	217258
200810	152998	175167	125454	224881
200811	164530	184561	132476	236646
200812	193715	196447	142090	250805

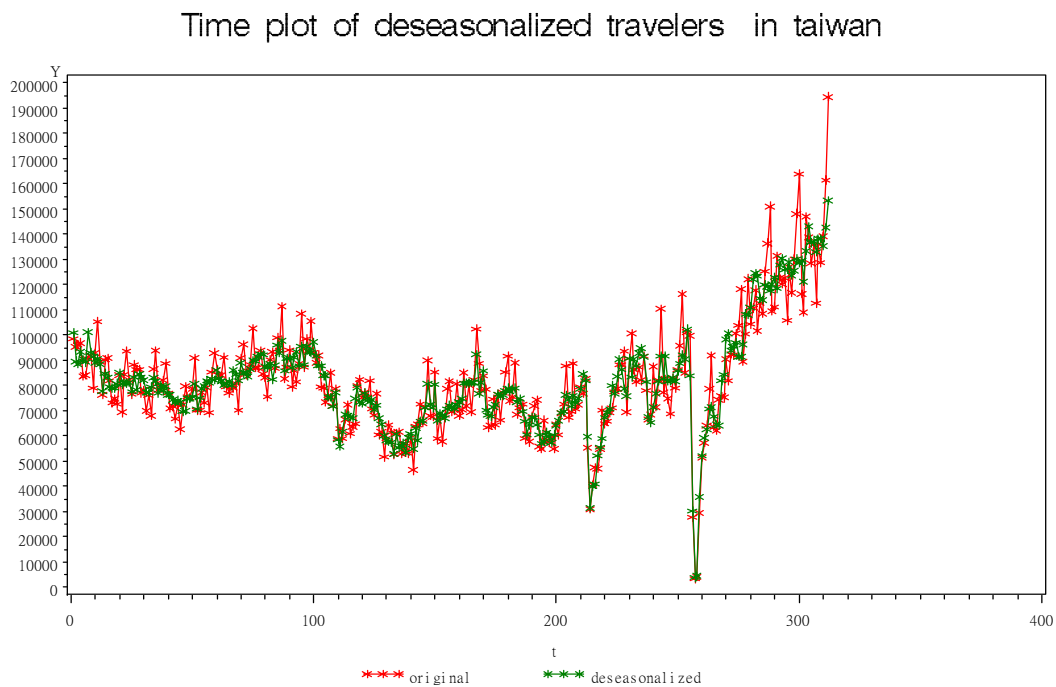


(圖 2-2) 指數平滑法實際值與預測圖

第三節 分解法Decomposition method

我們將時間序列分解成四個構成要素，趨勢－循環成分、季節成分，不規則成分，四個成分都是彼此獨立；我們利用X11方法來分解時間序列資料，使預測更準確。分解法又可分為加法模型和乘法模型，如果變異數一致，代表變異數保持平穩，我們會使用加法模式，但變異數越來越大，代表變異數不一致，則使用乘法模式。

由圖3-1可以看出去掉季節因子後的時間序列有隨著時間逐漸增加的趨勢，再由表3-1亦可判斷出時間和來台觀光人數呈現正向趨勢的關係；結果代表沒有季節因素的影響則來台觀光人數也會逐漸增加。趨勢-循環項的時間序列圖(圖3-2)顯示資料是有升現象的趨勢循環。圖3-3為季節因子之時間序列圖，由圖可以看出，此份資料有明顯的季節現象，圖3-4則為不規則項之時間序列圖

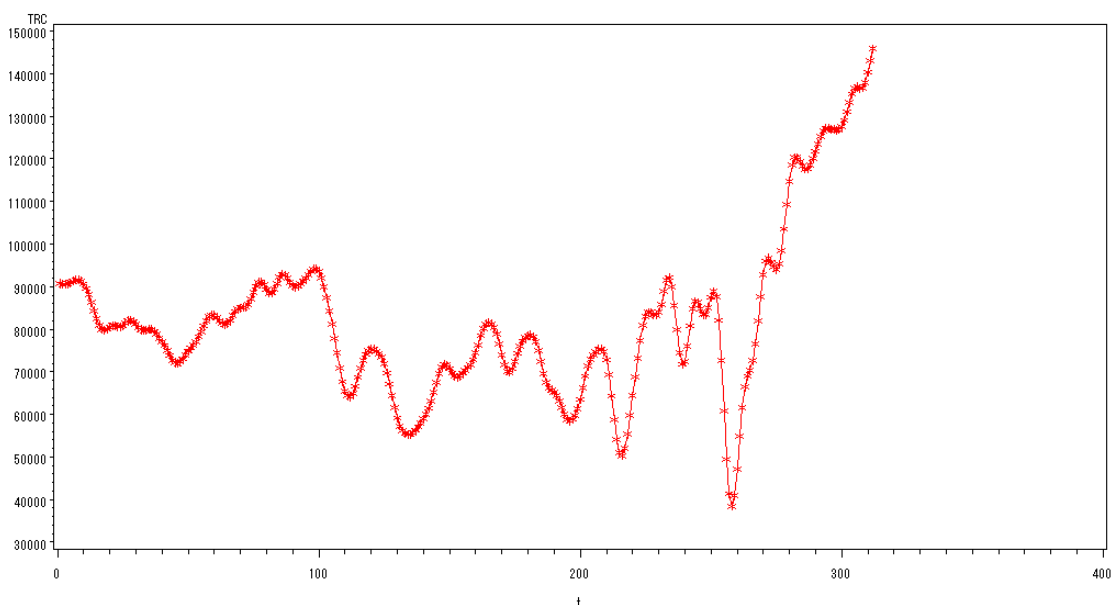


(圖 3-1) 去季節因子的時間序列圖

(表3-1) 去季節因子的參數估計表

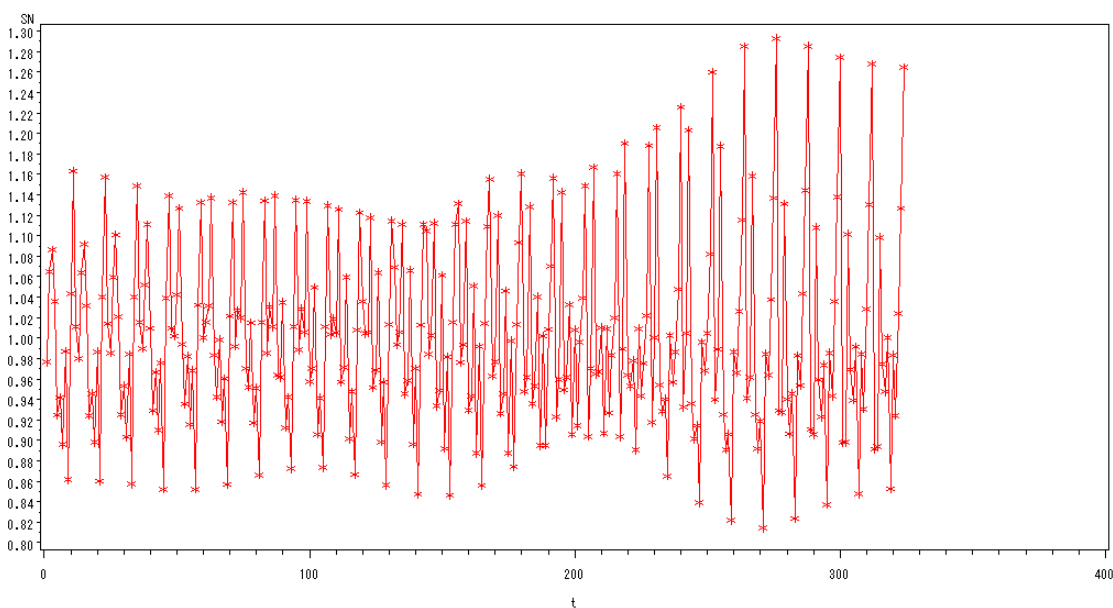
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	72035	2271	31.73	<.0001
t	1	64.6686	12.5744	5.14	<.0001

Trend-cycle component

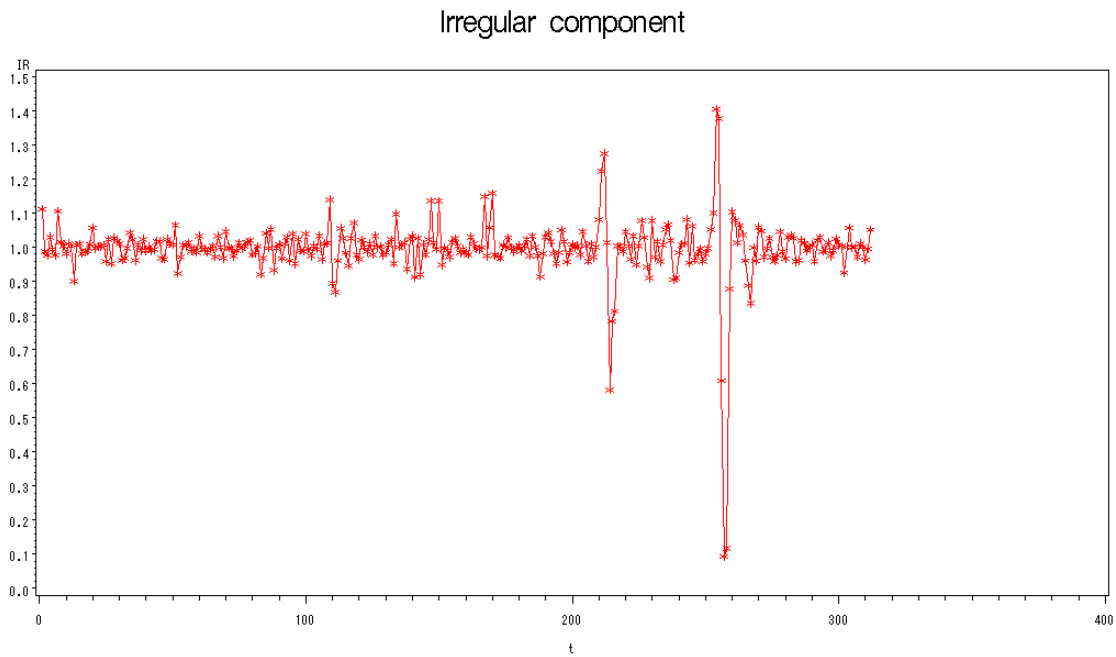


(圖 3-2) 趨勢—循環時間序列圖

Seasonal component



(圖 3-3) 季節因子之時間序列圖



(圖 3-4)不規則項之時間序列圖

我們將使用加法模式來討論。

加法模式的公式為： $y_t = TR_t + SN_t + CL_t + IR_t$

趨勢預測估計式為： $dy_t = \beta_0 + \beta_1 t + \varepsilon_t$

我們檢視殘差項是否自我相關可看出表3-2， $Pr < DW$ 為 < 0.0001 ，表示有嚴重的正自我相關。所以在後面的步驟，我們將會對殘差配適一階自我相關。表3-3為對殘差配適一階自我相關之後。DW檢定中，發現殘差依然存在著正自我相關，可能需要二階自我相關的配適，殘差配適二階自我相關(表3-4)，發現依然存在正自我相關。

(表 3-2)DY 模型的參數估計表

The AUTOREG Procedure
Dependent Variable DY

Ordinary Least Squares Estimates

SSE	1.24055E11	DFE	310
MSE	400178031	Root MSE	20004
SBC	7074.81228	AIC	7067.32627
Regress R-Square	0.0786	Total R-Square	0.0786
Durbin-Watson	0.1201	Pr < DW	<.0001
Pr > DW	1.0000		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	72035	2271	31.73	<.0001
t	1	64.6686	12.5744	5.14	<.0001

(表 3-3)DY 配適一階自我相關模型的參數估計表

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.921510	0.022093	-41.71

Yule-Walker Estimates

SSE	1.47928E10	DFE	309
MSE	47873318	Root MSE	6919
SBC	6418.9533	AIC	6407.72429
Regress R-Square	0.0094	Total R-Square	0.8901
Durbin-Watson	1.6248	Pr < DW	0.0003
Pr > DW	0.9997		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	72010	9147	7.87	<.0001
t	1	84.9662	49.7060	1.71	0.0884

(表 3-4)DY 配適二階自我相關模型的參數估計表

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-1.006628	0.056737	-17.74
2	0.092367	0.056737	1.63

Yule-Walker Estimates

SSE	1.44927E10	DFE	308
MSE	47054369	Root MSE	6860
SBC	6418.31866	AIC	6403.34665
Regress R-Square	0.0101	Total R-Square	0.8924
Durbin-Watson	1.7807	Pr < DW	0.0227
Pr > DW	0.9773		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	72061	8434	8.54	<.0001
t	1	81.6152	45.9915	1.77	0.0770

在這裡我們推斷，配到了二階自我相關，殘差依然存在自我相關的問題，初步判斷可能資料是非線性關係，所以我們在這裡以曲線 $t2=t^2$ 來重新配適模式於(表3-5)中，以曲線方式配適一階自我相關，發現殘差仍然存在著正自我相關，在這裡以曲線方式配適二階自我相關，於(表3-6)中，我們發現其 $Pr < DW$ 0.1776 和 $Pr > DW$ 0.8222 皆大於顯著水準 $\alpha = 0.05$ ，所以已不存在自我相關模式且其截距項, t , $t2$ 中，其 P 值 < 0.0001 ，都是顯著的

(表 3-5) DY 配適一階自我相關模型與 T^2 項的參數估計表

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.892701	0.025678	-34.77

Yule-Walker Estimates

SSE	1.41722E10	DFE	308
MSE	46013574	Root MSE	6783
SBC	6411.02539	AIC	6396.05338
Regress R-Square	0.0769	Total R-Square	0.8947
Durbin-Watson	1.6349	Pr < DW	0.0004
Pr > DW	0.9996		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	102851	9527	10.80	<.0001
t	1	-541.2690	140.5388	-3.85	0.0001
t2	1	1.9838	0.4334	4.58	<.0001

(表 3-6) DY 配適二階自我相關模型與 T^2 項的參數估計表

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-1.034567	0.056348	-18.36
2	0.158918	0.056348	2.82

Yule-Walker Estimates

SSE	1.37012E10	DFE	307
MSE	44629204	Root MSE	6681
SBC	6406.27394	AIC	6387.55893
Regress R-Square	0.0923	Total R-Square	0.8982
Durbin-Watson	1.9085	Pr < DW	0.1778
Pr > DW	0.8222		

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	102732	8402	12.23	<.0001
t	1	-532.0882	124.0417	-4.29	<.0001
t2	1	1.9408	0.3829	5.07	<.0001

其模式為：

$$dy_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

$$\text{where } \varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + a_t \quad (\text{模型 3.})$$

$$dy_t = 102732 - 532.0882t + 1.9406t^2 + \varepsilon_t$$

$$\text{where } \varepsilon_t = 1.034567\varepsilon_{t-1} - 0.158918\varepsilon_{t-2} + a_t$$

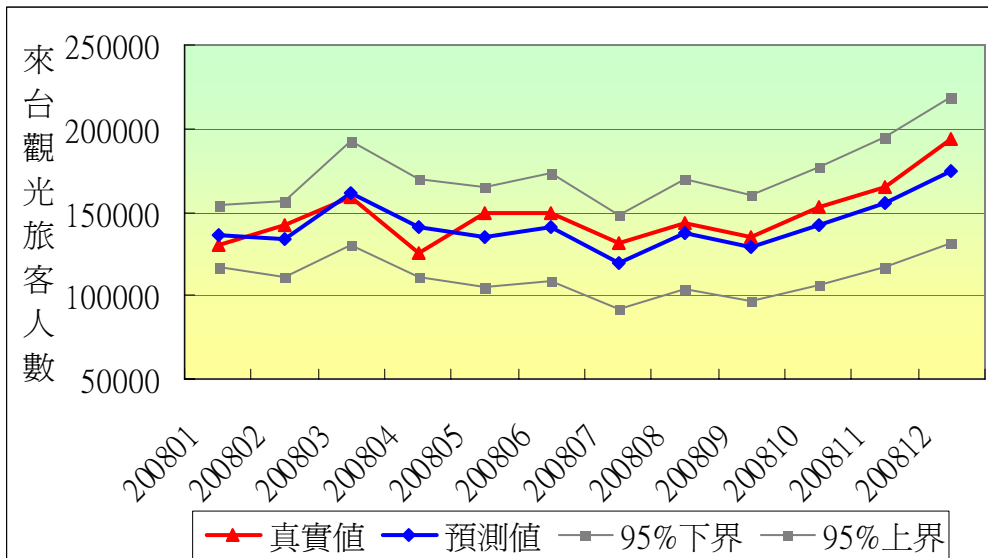
預測

我們事先保留了12筆的真實值，我們將預測出來12筆預測值、12筆真實值與95%信賴區間的上下界限整理如表3-7。

從圖3-5，所有的預測值皆與真實值相差不大，表示我們所建立出來估計式的預測能力是良好的。再由95%信賴水準的預測圖可以看出，所有的預測值皆在信賴區間的範圍之內，表示預測結果相當不錯。

(表 3-7)分解法實際值與預測表

	真實值	預測值	95%下界	95%上界
200801	129827	135724.5	116891.3	154557.7
200802	141667	133640.0	111043.0	156236.9
200803	159366	161208.8	130354.3	192063.2
200804	124999	140865.0	111613.5	170116.6
200805	149415	135282.9	105562.7	165003.1
200806	149473	141409.7	109068.0	173751.5
200807	131359	119630.8	91453.9	147807.7
200808	143140	137208.0	104178.9	170237.0
200809	134740	128464.8	97031.6	159897.9
200810	152998	141933.1	106775.1	177091.0
200811	164530	155932.8	116947.2	194918.4
200812	193715	174959.3	130910.0	219008.5



(圖 3-5) 分解法實際值與預測圖

第四節 ARIMA之介入分析

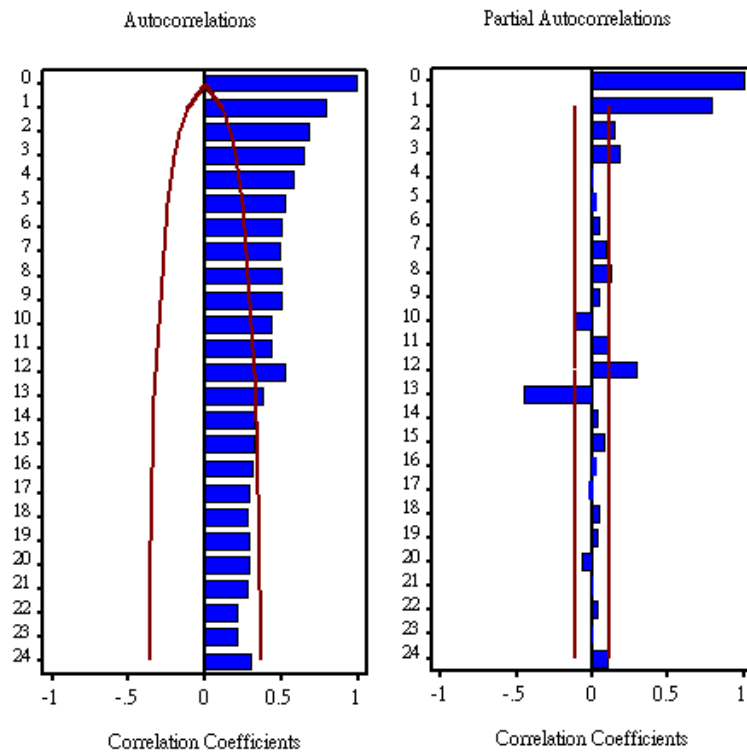
在研究方法中，我們是利用由Box and Jenkin(1976)年提出自我迴歸整合移動平均模型(Autoregressive Integrated Moving Average Model; ARIMA)，介入分析是指時間序列在某個時點受到外在因素影響或政策使其改變，則須採介入因子分析此種類型的時間序列。若以介入模型分析時間序列時，必須注意：

- 1.介入因子時間的起點。
- 2.介入因子時間影響圖形。

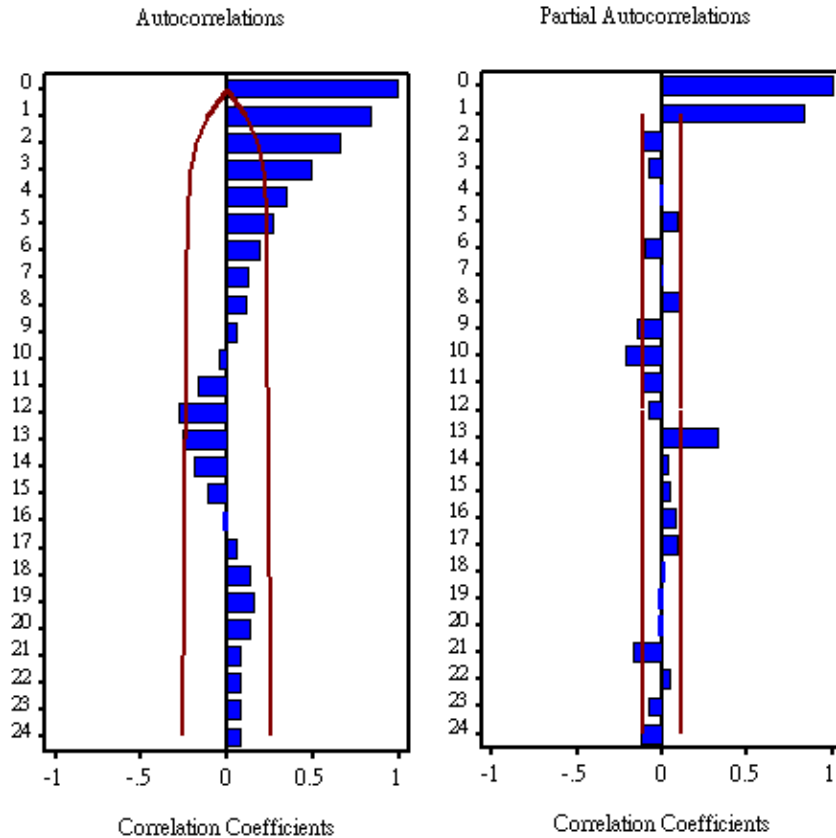
對於介入起點及圖形，均以虛擬變數的方法定義介入模型中。

一開始我們先檢視我們的原始時間序列圖(圖1)，由圖當中，我們得知資料的季節的波動保持一致的狀態，所以變異數是平穩的，但資料有上升的趨勢，代表平均數為不平穩，可以配合資料的ACF與PACF圖，來確定資料是否平穩。由原

始的ACF與PACF圖(圖4-1)，發現ACF圖呈現的是緩慢下降的狀態，更進一步可看出殘差Lag12、Lag24的部分都有著特別突出的現象，代表的此筆資料隨著時間的進行有擺盪的現象，確定了平均數為不平穩，所以我們要解決平均數不平穩的問題，解決方法為對資料做一次差分或季節差分，我們對資料做一次差分與季節差分，互相比較過後，發現做季節差分後ACF圖收斂較快(圖4-2)，但同時做一次差分與季節差分，卻又過度差分，所以判斷結果認為對資料做季節差分最為適當。在經過季節差分之後，時間序列的變異數與平均數皆已達到平穩的狀態，表示可以對資料進一步配適模型。



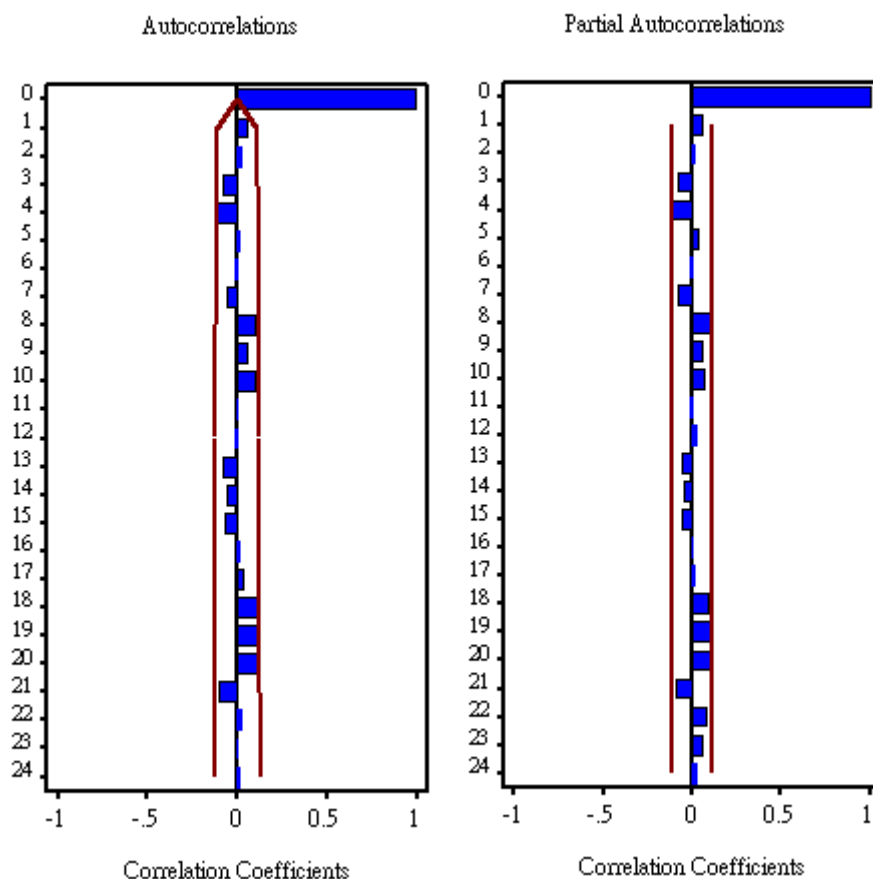
(圖 4-1)原始的 ACF 圖與 PACF 圖



(圖 4-2)季節差分之 ACF 圖與 PACF 圖

經過季節差分之後的ACF與PACF圖，我們認為PACF圖呈現出較好的cut off 表現，ACF圖呈現快速下降的狀態，所以我們決定要配適AR模式。由PACF 圖可以看出，殘差在lag1與lag12之處有較明顯的不顯著，且由於原始資料的變異數有變大的趨勢，所以我們將以乘法模式來處理，對資料配適ARIMA(1,0,0)(0,1,1)_s 的模式(圖4-3)。

ARIMA $p=(1)$ $d=(12)$ $q=(12)$ NOINT



(圖 4-3)配完模式後之 ACF 圖與 PACF 圖

在後面，我們將探討自我相關檢測，以及參數估計值，還有此模式的預測表現。我們對**White noise**、對單根的檢定和**Ljung-Box** 檢定，用以判定殘差項是否有自我相關。

一、White noise

我們先建立虛無與對立假設

$$\begin{cases} H_0 : \text{White Noise} \\ H_1 : \text{No White Noise} \end{cases}$$

檢定規則為若P-value 小於0.05，則拒絕虛無假設，即可以檢定出殘差項有不完全符合White Noise。(圖4-4)

二、單根的檢定

$$\begin{cases} H_0 : \text{時間序列不平穩} \\ H_1 : \text{時間序列平穩} \end{cases}$$

檢定規則為若P-value都小於顯著水準0.05，則拒絕 H_0 ，表示時間序列已經達到平穩的狀態(圖4-5)。

三、Ljung-Box

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 & K = 6.12.18.24.30.36.42.48 \\ H_1 : \text{至少一個不為0} \end{cases}$$

檢定規則為若P-value大於0.05，則不拒絕 H_0 ，表示殘差項沒有自我相關。由(表4-1)可以看出，所有的P-value在Lag24、Lag30與Lag48皆小於0.05，表示配適後的殘差有自我相關，也就是該模式的配適是不合適的。

其ARIMA模型為(表4-2)：

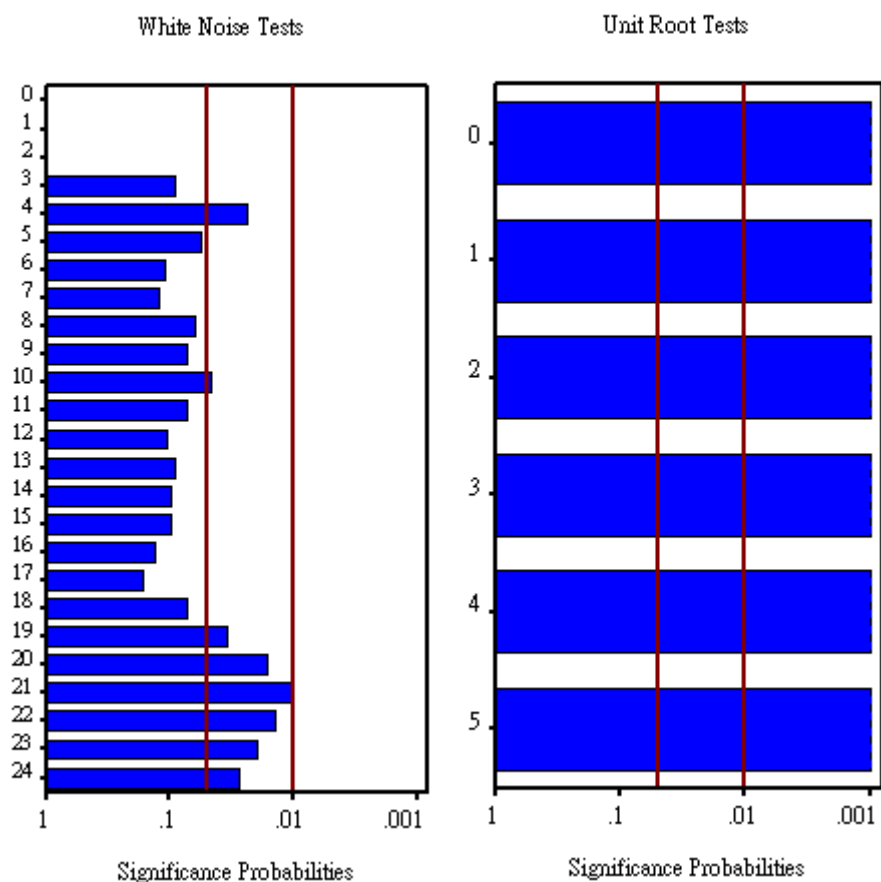
$$(1 - \phi_1 B)(1 - B^{12})y_t = (1 - \Theta B^{12})a_t$$

把我們參數估計值放入，其預測方程式如下：

$$(1 - 0.89793B)(1 - B^{12})y_t = (1 - 0.59773B^{12})a_t$$

$$\hat{\sigma}^2 = 79187473$$

ARIMA p=(1) d=(12) q=(12) NOINT



(圖 4-4)White Noise 檢定 (圖 4-5)單根檢定

(表 4-1) Ljung-Box 自我相關檢測圖

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	8.59	4	0.0721	0.063	0.020	-0.089	-0.123	0.018	-0.018
12	16.80	10	0.0789	-0.055	0.104	0.059	0.095	0.000	-0.013
18	25.75	16	0.0577	-0.073	-0.059	-0.064	0.007	0.030	0.120
24	37.91	22	0.0188	0.116	0.118	-0.097	0.024	-0.000	0.017
30	41.90	28	0.0443	0.022	-0.051	0.062	0.064	-0.032	-0.003
36	44.81	34	0.1016	-0.036	-0.051	0.048	0.001	-0.044	-0.022
42	53.93	40	0.0695	-0.046	-0.041	0.000	-0.047	0.032	0.138
48	73.87	46	0.0057	0.121	0.131	0.038	-0.121	-0.068	-0.060

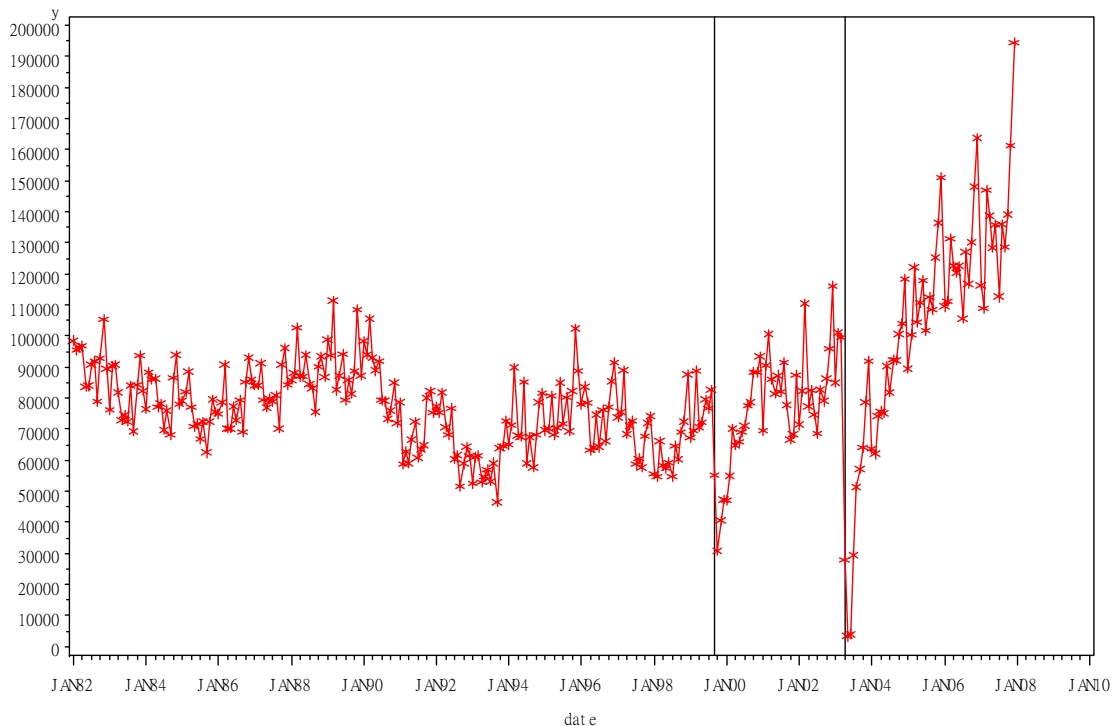
(表 4-2) 參數估計值

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.59773	0.05399	11.07	<.0001	12
AR1,1	0.89793	0.02740	32.77	<.0001	1

Variance Estimate	79187473
Std Error Estimate	8898.734
AIC	6316.142
SBC	6323.55
Number of Residuals	300

我們檢查原始時間序列圖中發現有兩點表現異常，這兩個時間點分別受到1999年921大地震影響，以及2003年4月到7月受到SARS疫情影響，所以才會導致來台觀光人數驟減，故將這兩個影響點加入模型(圖4-6)。



(圖 4-6) ARIMA 之介入分析-原始時間序列圖

如同ARIMA方法，在配適模型之前，檢查序列是否平穩，發現還是要對序列做季節差分，之後配適模型；

$$(1 - \phi_1 B^1 - \phi_2 B^2)(1 - B^{12})y_t = (1 - \Theta B^{12})a_t$$

$$y_t = \frac{C_1}{1 - \delta_0 B} P_{1t} + \frac{C_2}{1 - \delta_1 B^1 - \delta_2 B^2} P_{2t} + \varepsilon_t$$

$$P_{1t} = \begin{cases} 1 & \text{if } t = 213(\text{September 1999}) \\ 0 & \text{otherwise} \end{cases}$$

$$P_{2t} = \begin{cases} 1 & \text{if } t = 256(\text{April 2003}) \\ 0 & \text{otherwise} \end{cases}$$

$$Z_t = C_1 z_t(p_1) + C_2 z_t(p_2) + \varepsilon_t$$

$$Z_t = (1 - B^{12})y_t$$

$$z_t(p_1) = (1 - B^{12})P_{1t}$$

$$z_t(p_2) = (1 - B^{12})P_{2t}$$

$$(1 - \phi_1 B - \phi_2 B^2)\varepsilon_t = (1 - \Theta B^{12})a_t$$

(模型 4.)

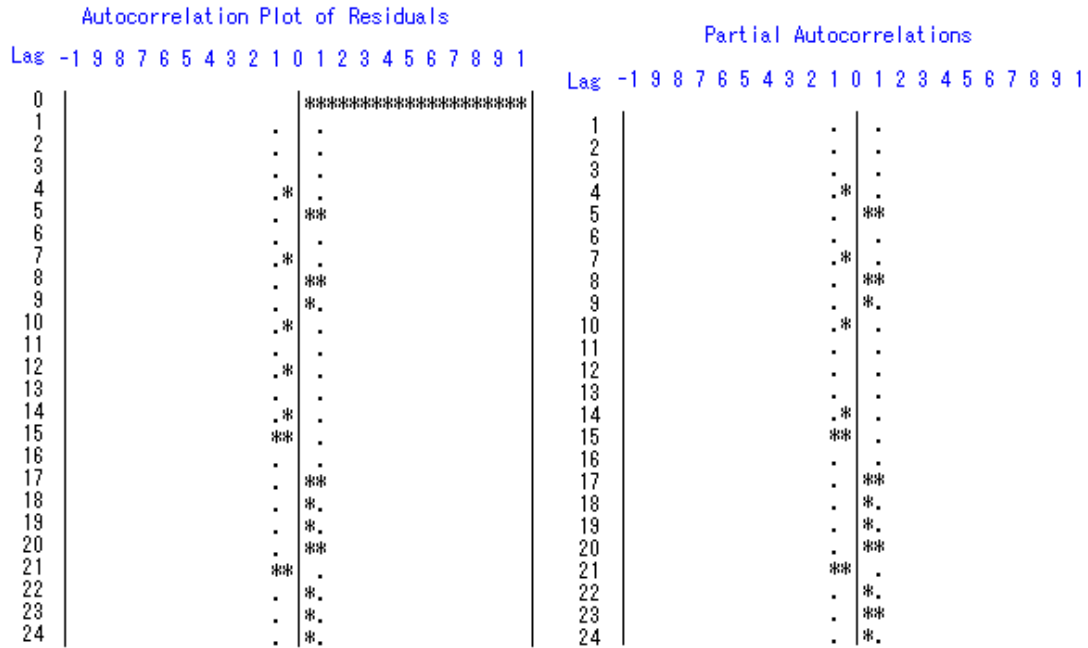
判斷模式

從圖中(圖4-7)，我們發現ACF與PACF皆在2倍標準差內，表示配適模型後的殘差皆無自我相關，在Q*統計量可以得知

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 & K = 6.12.18.2430.36.4248 \\ H_1 : \text{至少一個不為} 0 \end{cases}$$

檢定規則為若P-value 大於0.05，則不拒絕H₀，表示殘差項沒有自我相關。由(表4-3)可以看出，所有的P-value 皆大於0.05，表示配適後的殘差沒有自我相關，也就是該模式的配適是合適的。

來台觀光旅客人數預測分析



(圖 4-7)ARIMA 之介入分析-配完模式後之 ACF 圖與 PACF 圖

(表 4-3) ARIMA 之介入分析-Ljung-Box 自我相關檢測圖

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	2.90	3	0.4079	-0.005	-0.025	-0.011	-0.045	0.082	0.002
12	6.96	9	0.6412	-0.043	0.086	0.033	-0.041	-0.017	-0.030
18	14.17	15	0.5128	0.001	-0.043	-0.113	0.005	0.085	0.031
24	22.16	21	0.3904	0.032	0.083	-0.087	0.067	0.056	0.040
30	28.93	27	0.3644	0.019	-0.017	0.037	0.055	-0.036	0.118
36	34.89	33	0.3781	-0.036	-0.103	0.018	-0.038	-0.062	0.008
42	39.79	39	0.4349	-0.056	-0.076	0.017	-0.065	0.027	-0.000
48	45.83	45	0.4374	0.045	0.081	-0.041	-0.045	0.044	-0.054

檢查參數

可看出參數估計的P值皆小於顯著水準 $\alpha = 0.05$ ，因此我們拒絕其虛無假設，即參數顯著，應當留在模型裡(表4-4)。綜合以上的結果可推知，我們所配適的模式 $ARIMA(2,0,0)(0,1,1)_S$ ，再加入介入分析為一個合適的模型，而參數估計值如下：

(表 4-4) ARIMA 之介入分析-參數估計表

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	Variable	Shift
MA1,1	0.53031	0.05885	9.01	<.0001	12	y	0
AR1,1	0.67002	0.05777	11.60	<.0001	1	y	0
AR1,2	0.24228	0.05891	4.11	<.0001	2	y	0
NUM1	-30243.4	6284.0	-4.81	<.0001	0	p1	0
DEN1,1	0.96382	0.05047	19.10	<.0001	1	p1	0
NUM2	-63406.6	6080.6	-10.43	<.0001	0	p2	0
DEN1,1	1.24165	0.06456	19.23	<.0001	1	p2	0
DEN1,2	-0.44225	0.06669	-6.63	<.0001	2	p2	0
Variance Estimate				56262161			
Std Error Estimate				7500.811			
AIC				6176.648			
SBC				6206.225			
Number of Residuals				298			

$$\hat{C}_1 = -30243.4 \quad \hat{C}_2 = -63406.6 \quad \hat{\phi}_1 = 0.67002 \quad \hat{\phi}_2 = 0.24228 \quad \hat{\Theta}_{12} = 0.05303$$

$$\hat{\delta}_0 = 0.96382 \quad \hat{\delta}_1 = 1.24165 \quad \hat{\delta}_2 = -0.44225 \quad \hat{\sigma}^2 = 56262161$$

預測

我們事先保留了12筆的真實值，我們將預測出來12筆預測值、12筆真實值與95%信賴區間的上下界限整理如表4-5。

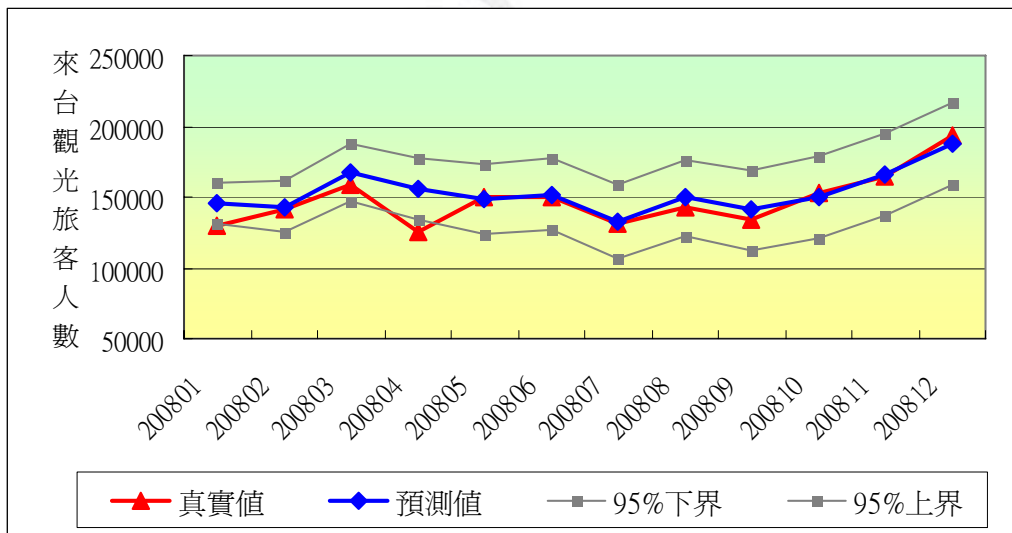
從圖4-8的預測圖可以看出實際值大多都在95%信賴區間內，代表此模型之預測能力相當不錯，還有兩筆觀測值是落在95%信賴區間的邊界附近，分別為2008年1月與4月，都比預測的信賴區間超出一些，其餘預測值的結果與實際值相去不遠，表示預測結果相當良好。

預測值在1月與4月與預測值差了一段距離，判斷全球股災發生在2008年1月中，因市場憂慮美國經濟衰退、美國次級房貸風暴無法控制以及法國興業銀行交易員熱羅姆·凱維埃爾涉嫌違規巨額交易歐洲股價期貨指數後的巨額平倉活動而引

發的全球金融市場震盪，加上中國南方遭遇大雪，而在2008年3月14爆發了震驚世界的”西藏事件”，原本一個藏人和平示威運動演變成騷亂、暴動甚至更嚴重的流血衝突，政府處理事情的方法也使正在準備北京奧運會的中國形象變的黯淡，當時世界各地因為這件事情爆發了反對的示威運動，可能因為這些原因，導致在中國鄰近的台灣受到影響。

(表 4-5) ARIMA 之介入分析-實際值與預測表

	真實值	預測值	95%下界	95%上界
200801	129827	145434.3	130733.0	160135.6
200802	141667	143374.4	125678.2	161070.6
200803	159366	166772.4	147366.1	188178.6
200804	124999	155761.5	133379.2	178143.7
200805	149415	148560.9	124575.1	172546.6
200806	149473	152078	126791.7	177364.3
200807	131359	132370.8	106009.0	158732.5
200808	143140	149538.3	122280.8	176796.5
200809	134740	140649.0	112637.8	168660.1
200810	152998	149983.8	121336.9	178630.8
200811	164530	165623.1	136437.0	194809.3
200812	193715	187638.4	157993.3	217283.4



(圖 4-8) ARIMA 之介入分析-實際值與預測表

第五節 最佳模型

我們利用時間序列迴歸法、指數平滑法、分解法 以及 ARIMA之介入分析以配適預測模式，再利用MAD、MSE、MPE和MAPE 四個準則來評估何者為最佳預測模式，MAD、MSE、MPE準則其值越小表示該模式為最佳，MAPE準則其值小於5表現良好，其值介於5至10表現不錯，大於10表現稍差。計算出MSE、MAD、MPE、MAPE 此四個評估值於表5-1，可以發現分解法與ARIMA介入分析法在所有準則之表現都是良好的，但就預測圖來看，發現ARIMA介入分析配適的結果，預測趨勢較符合真實資料，因此選擇ARIMA介入分析配適模型為我們的最佳模式。

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

(表 5-1)模式比較

分析方法	MSE	MAD	MPE	MAPE
時間序列迴歸法	1119526188	28310.75	9.65781	18.6041
指數平滑法	905504381.3	28006	-19.889182	19.889182
分解法	114942795.8	9681.84	3.49971	6.56501
ARIMA 介入分析	116258564.2	6953.81	-4.12218	5.06862
準則下的最佳模型	分解法	ARIMA 介入分析	分解法	ARIMA 介入分析

第三章 結論與建議

在該研究中，我們使用了時間序列迴歸法、指數平滑法、分解法以及ARIMA之介入分析這四種方法來分析來台觀光旅客人數資料。在這四個研究方法之中，分解法與ARIMA介入分析法在所有準則之表現都是良好的，但就預測圖來看，發現ARIMA介入分析配適的結果，預測趨勢較符合真實資料，因此選擇ARIMA介入分析配適模型為我們的最佳模式。

圖4-8的預測圖可以看出實際值大多都在95%信賴區間內，代表此模型之預測能力相當不錯，但兩筆觀測值是落在95%信賴區間的邊界附近，全球股災發生在2008年1月中，因市場憂慮美國經濟衰退、美國次級房貸風暴無法控制以及法國興業銀行交易員熱羅姆·凱維埃爾涉嫌違規巨額交易歐洲股價期貨指數後的巨額平倉活動而引發的全球金融市場震盪，加上中國南方遭遇大雪，而在2008年4月判斷在2008年3月14爆發了震驚世界的”西藏事件”，原本是一個藏人和平示威運動，卻演變成騷亂、暴動甚至更嚴重的流血衝突，政府處理事情的方法也使正在準備北京奧運會的中國形象變的黯淡，當時世界各地因為這件事情爆發了反對的示威運動，可能因為這些原因，導致台灣受到影響。

圖形中可看出信賴區間都比預測的超出一些，其餘預測值的結果與實際值相去不遠，表示預測結果相當良好；但從預測分析結果看來，天然災害如地震或流行病如SARS，都可能使來台觀光人數陷入低潮，而如何從低潮中走出來，就是有關當局應該思考的問題。

台灣觀光經過了大地震與SARS的影響，政府為了提升來台觀光人數，積極推動「後SARS復甦計畫」各項宣傳 促銷措施下：

『觀光局分三階段執行「後SARS觀光復甦計畫」：第一階段為全民量體溫結束後啟動（92年6月11日至6月底），目標為復甦國民旅遊市場，第二階段為世界衛生組織（WTO）將台灣自感染區解除後啟動（92年7月至9月），目標為復甦外人來台旅遊信心、商務旅行趨於正常、觀光旅遊恢復七成。第三階段為恢復主要市場來台旅遊人次（92年10月至12月），目標為12月底來台旅客數恢復達年初水準。』

希望來台旅客逐步恢復，不只要政府的政策，更需要民眾的配合，更透過參與國際旅展及各地區大型活動、結合國際大型賽會推廣行銷臺灣，就是希望招攬更多國際旅客來臺。

以觀光局的統計數字來看，2007年10月份大陸客以觀光目的來台的人數成長65.66%，看似驚人，實際上總人數只有1萬1797人次，1月到10月加起來也只有7萬255人次。同一時間10月份的日本觀光客掉到剩6萬460人次，負成長4.64%，1到10月加起來只有55萬422人次，比去年同期負成長8.56%，相當於少掉將近5萬人左右。

其實台灣是有相當好的觀光品質，從報章雜誌上也都可以看到外國人來台，都稱讚我們的小吃、夜市，令他們流連忘返，

雖然政府有在重視觀光業的發展，但是卻不懂得如何包裝，行銷到世界各國，由主計處的統計資料顯示，來台觀光人口也都以東南亞居多，歐洲、美洲占的比例相當低，才會讓觀光人口都無法再往上突破，在這樣下去，只會讓這瓶頸越來越明顯。

近幾年來，從新聞上都看到每個地區都在發展自己地方特色，文化、食物、風景、地方風俗，使自己別樹一格，來吸引觀光客的前來；在與大陸三通之後，大陸客也越來越多人包圍前往台灣，也製造了一筆商機，但這只是短暫的，若是希望它能長久持續下去，就要看政府與人民的政策和計畫與配合，發展地方特色、國家形象和良好的治安與環境，這樣才是使國內觀光能歷久不衰的方法。



參考文獻

Bowerman ,O'Connell and Koehler ,Forecasting. Time series ,and Regression 4th

陳婉淑教授 統計預測方法講義

AREMOS經濟統計資料庫網站：

http://cache.moe.edu.tw/aremos_ly/search.html

92 年觀光年報：

http://admin.taiwan.net.tw/auser/B/Annual_2003/chinese/c_chap0_4.htm

互動百科：<http://www.hudong.com/wiki/2008%E5%B9%B4>

行政院主計處：<http://www.dgbas.gov.tw/mp.asp?mp=1>

