

# 台灣鐵路客運指數預測分析



作者：溫士奇 D9551610

李靜芬 D9453207

楊珮雯 D9453298

吳思嫻 D9453429

系級：統計三乙

開課老師：陳婉淑 教授

課程名稱：統計預測分析

開課系所：統計系

開課學年：96 學年度 第二學期

## 摘要

由於國內運輸環境變遷，台鐵之經營日益艱困，尤其即將面臨高鐵通車之危機，但危機也就是轉機，台鐵為能永續經營，應積極謀求改善策略。如何在國內運輸市場結構即將面臨重組之際，尋求台鐵最有利的重新定位點，實為台鐵當前的重要課題。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搭乘鐵路客運的人數，並希望這份報告能夠作為台鐵營運參考。

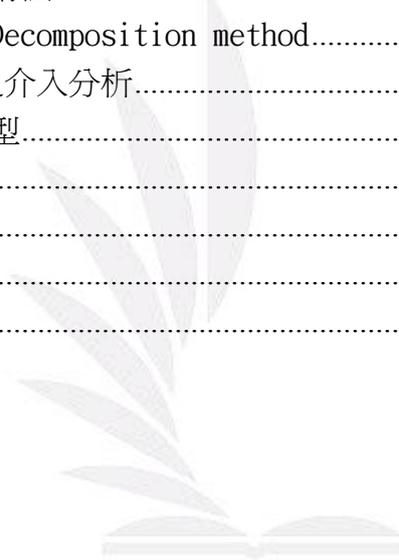
報告首先描述時間序列圖是否呈現趨勢，或變異數與平均數是否平穩。接著使用時間序列迴歸法、指數平滑法、分解法及 ARIMA 之介入分析配適模型。時間序列迴歸法是依照變數本身過去的資料，所存在的變異型態來建立模型；指數平滑法其計算方法是將前一期的預測值為基礎，再加上預測值與實際值之間差額的百分比，再選擇平滑係數，以預測之誤差結果來指引決策；分解法將時間序列分解成四個成份(Component)，趨勢、循環效應、季節效應、隨機性，並選擇加法模型或乘法模型來配適模型；ARIMA 為 AR 模式、差分與 MA 模式的組合模式，而介入分析是用來分析某一件事或政策(稱之介入因子)對於時間數列模型之影響。最後使用 MSE、MAE、MPE 及 MAPE 四個準則，來評估這四種方法何者最佳，並選出最佳模型。以 MSE、MAE 及 MPE 準則而言，其值愈小表現愈良好，反之則否；而 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。分析結果顯示，ARIMA 之介入分析的配適模型為最終模型。

從一般的角度看來，台灣鐵路客運營運不但受到航空客運威脅，還面臨公路客運的搶佔，加上高鐵的完工通車，及都會捷運系統的競爭，導致其營運指數時而成長時而停滯，但從分析結果看來，天然災害如颱風或流行病學如 SARS，都可能使台鐵營運陷入低潮，而如何從低潮中走出來，就是台鐵當局應該思考的問題。

**關鍵字：** ARIMA、介入分析、台鐵、時間序列、預測

# 目 錄

圖目錄.....	3
表目錄.....	3
第一章 緒論.....	4
第一節 研究動機.....	4
第二節 研究目的.....	4
第三節 研究背景.....	5
第四節 資料描述.....	5
第五節 研究方法.....	5
第六節 研究流程.....	6
第二章 研究主體.....	7
第一節 時間序列迴歸法Time Series Regression.....	8
第二節 指數平滑法.....	13
第五節 分解法Decomposition method.....	18
第五節 ARIMA之介入分析.....	25
第五節 最佳模型.....	33
第三章 結論與建議.....	34
第一節 結論.....	34
第二節 建議.....	34
參考文獻.....	35



## 圖目錄

圖 二-1 原始時間序列圖 .....	7
圖 二-2 時間序列迴歸法-修正後線性迴歸模型的 95%信賴水準之預測圖 .....	12
圖 二-3 時間序列迴歸法-修正後線性迴歸模型的未來一年之預測圖 .....	13
圖 二-4 指數平滑法-配模後之 ACF、PACF 圖 .....	14
圖 二-5 指數平滑法-WHITE NOISE 檢定與單根檢定 .....	16
圖 二-6 指數平滑法-預測圖 .....	17
圖 二-7 分解法-去季節因子的時間序列圖 .....	19
圖 二-8 分解法-趨勢循環之時間序列圖 .....	20
圖 二-9 分解法-季節因子之時間序列圖 .....	20
圖 二-10 分解法-不規則項之時間序列圖 .....	21
圖 二-11 分解法-加法模型的預測值描述趨勢循環圖 .....	22
圖 二-12 分解法-95%信賴水準預測圖 .....	23
圖 二-13 分解法-調整後 95%的預測區間 .....	24
圖 二-14 ARIMA 之介入分析-原始時間序列圖 .....	27
圖 二-15 ARIMA 之介入分析-原始 ACF、PACF 圖 .....	28
圖 二-16 ARIMA 之介入分析-季節差分後之 ACF、PACF 圖 .....	29
圖 二-17 ARIMA 之介入分析-配適最終模型之樣本 ACF 圖與 PACF 圖 .....	30
圖 二-18 ARIMA 之介入分析-95%信賴水準預測圖 .....	32

## 表目錄

表 二-1 時間序列迴歸法-預設模型的參數估計表 .....	8
表 二-2 時間序列迴歸法-預設模型的統計量 .....	9
表 二-3 時間序列迴歸法-修正後線性迴歸模型的統計量 .....	10
表 二-4 時間序列迴歸法-模型的自迴歸參數 .....	11
表 二-5 時間序列迴歸法-修正後線性迴歸模型的參數估計表 .....	11
表 二-6 時間序列迴歸法-實際值與預測值的差 .....	12
表 二-7 指數平滑法-參數估計表 .....	15
表 二-8 指數平滑法-預測值 .....	17
表 二-9 分解法-去季節因子的參數估計表 .....	19
表 二-10 分解法-加法模型的自迴歸參數 .....	22
表 二-11 分解法-加法模型參數估計表 .....	22
表 二-12 分解法-調整後實際值與預測值 .....	24
表 二-13 ARIMA之介入分析-Q統計量 .....	31
表 二-14 ARIMA之介入分析-參數估計表 .....	31
表 二-15 ARIMA之介入分析-實際值與預測值的差 .....	33

# 第一章 緒論

此章將為各位說明我們做這份報告之動機與目的，並同時點出整份報告所要探討的問題為何，以及解決這個問題的重要性。

## 第一節 研究動機

鐵路為台灣內陸交通運輸大動脈，百年來肩負本島陸上客貨運輸之重責大任，對國家社會經濟發展以及國防上均有重大貢獻。自民國67年中山高速公路通車後，公路客貨運輸發展迅速，致台鐵客貨運乃遭受嚴重衝擊。但隨著國內經濟成長，國民所得持續增加，自用小客車逐年激增，高速公路部分路段自75年起其尖峰時間之交通量已超過設計服務容量，遇有連續假期則呈現嚴重壅塞現象，又因都會區大型公共工程頻頻開工，市區交通壅塞，台鐵陸續引進通勤電車，中短程旅客乃逐漸回流，尤其短程旅客更是逐年成長。

長途旅客方面，由於政府開放航空後，國內客運市場結構隨之銳變，國內航空客運蓬勃發展，公路客運日漸式微，而由航空取而代之。各航空公司在快速擴充機隊及增開國內航線下，因競爭激烈又紛紛推出票價折扣吸引旅客。民國84年9月台鐵調漲票價，但高速公路及航空客運並未調價，面對國內航空挾其時間及票價的優勢，台鐵長途旅客復大量移轉航空，嚴重影響營收。自86年起台鐵新購400輛推拉式自強號陸續引進，加入西部幹線營運，運能大幅提昇，加上航空票價於88年12月及89年2月分二階段調漲，長途旅客遂又有回流台鐵的跡象。

## 第二節 研究目的

由於台鐵的運輸功能主要定位在「城際鐵路運輸服務」，在台鐵運輸結構中，七成三為短程旅次，係台鐵主要客源，但客票收入超過七成六卻係來自中長程旅運。目前台鐵除了有航空競爭外，還面臨公路客運的搶佔，加上高鐵的完工通車及都會捷運系統的競爭。如何在國內運輸市場結構即將面臨重組之際，尋求台鐵最有利的重新定位點，實為台鐵當前的重大課題。

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搭乘鐵路客運的人數，並希望這份報告能夠作為台鐵營運參考，讓台鐵的經營能預先知道未來的狀況、提早想出因應的對策，以達到永續經營的目標。

### 第三節 研究背景

現今交通愈來愈發達，交通工具的種類愈來愈多。從以前徒步、馬車、船、汽車、火車、郵輪、飛機、捷運甚至是高鐵，交通工具是愈來愈先進。台灣在二十幾年前公路客運並非盛行時，長途搭乘幾乎是搭乘台灣鐵路客運。但在那個經濟並不富裕的時代，火車的種類並沒有像現在這麼多，只有平快車與復興號可以選擇，既不便宜又不舒適，而且速度較現今的火車慢，所以當時搭乘鐵路的旅客相對於現在來說並不多。但隨著時代的進步，鐵路客運愈來愈快速與舒適，所以鐵路客運搭乘人數漸漸上升；但在這兩年，公路長途客運愈來愈發達，且價格又比鐵路客運來得便宜許多，雖然公路客運並不比鐵路客運來的快，但是以經濟學原理，多數人會較喜歡寧可多花一點點時間，來換取少許金錢付出，也有少數人們會想多花一點點的金錢，來換取少許時間付出，而改搭高速鐵路客運。所以在最近兩年，搭乘鐵路客運的人數似乎有漸漸下降的趨勢。此外，影響鐵路運輸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：2001年9月納莉風災、2003年5月SARS肆虐，由於風災的影響，致使鐵公路運輸被迫停擺，而影響到鐵路的營運。2003年的SARS事件更是影響了所有大眾運輸工具的載客率，因此，讓我們更想深入去探討這兩起事件到底影響台灣鐵路客運有多深。

### 第四節 資料描述

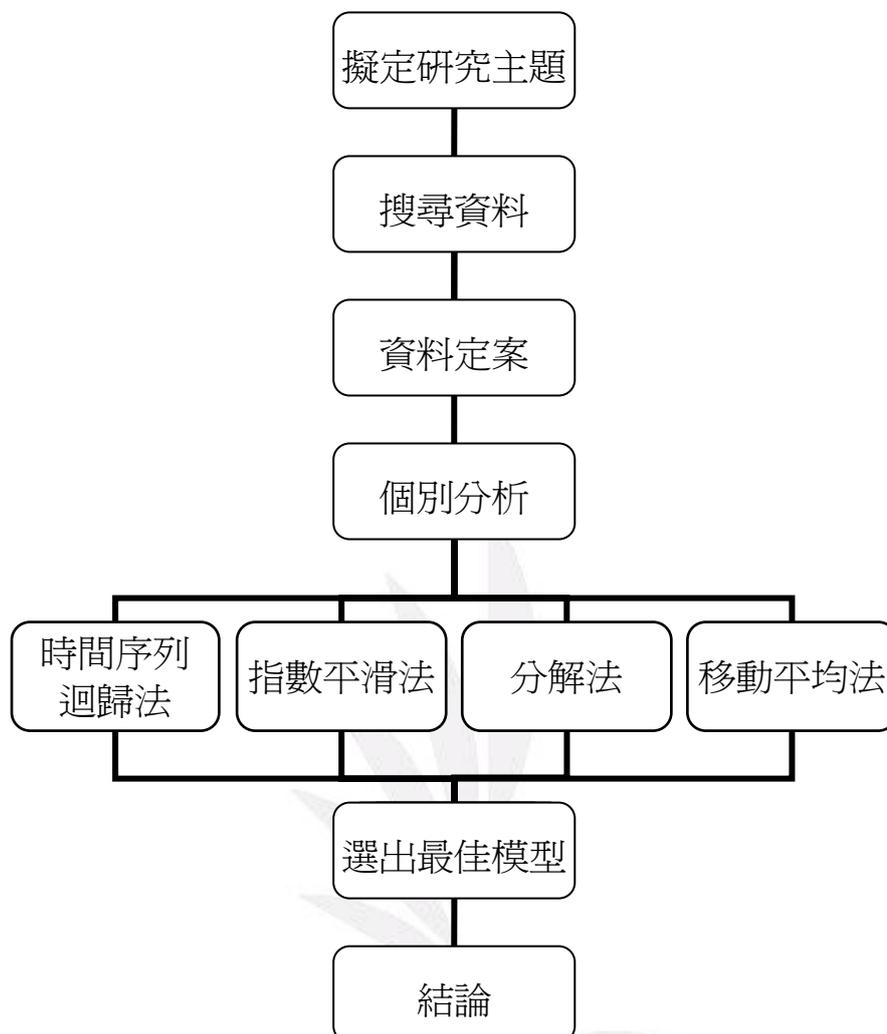
此筆資料主要是記錄每月全台灣鐵路之載客人數，其單位為指數(此筆資料是以2001年之平均載客人數當作基底(100)，再將各月份之載客人數以此為基底一一轉換而來)。分析時間為西元1989年1月至2008年2月，並保留最後12筆做預測。資料來源為AREMOS經濟統計資料庫。想了解更詳細的相關資料，可參考AREMOS經濟統計資料庫網站([http://cache.moe.edu.tw/aremos\\_ly/search.html](http://cache.moe.edu.tw/aremos_ly/search.html))。

### 第五節 研究方法

此研究我們使用了四種統計預測方法來做分析，分別為：

1. 時間序列迴歸法：對時間序列之研究，稱為時間序列分析。
2. 指數平滑法
3. 分解法
4. ARIMA之介入分析

## 第六節 研究流程



## 第二章 研究主體

本章首先描述時間序列圖是否呈現趨勢，或變異數與平均數是否平穩，再使用時間序列迴歸法、指數平滑法、分解法及 ARIMA 之介入分析配適模型，最後使用 MSE、MAE、MPE 及 MAPE 四個準則，來評估這四種方法何者最佳，並選出最佳模型。

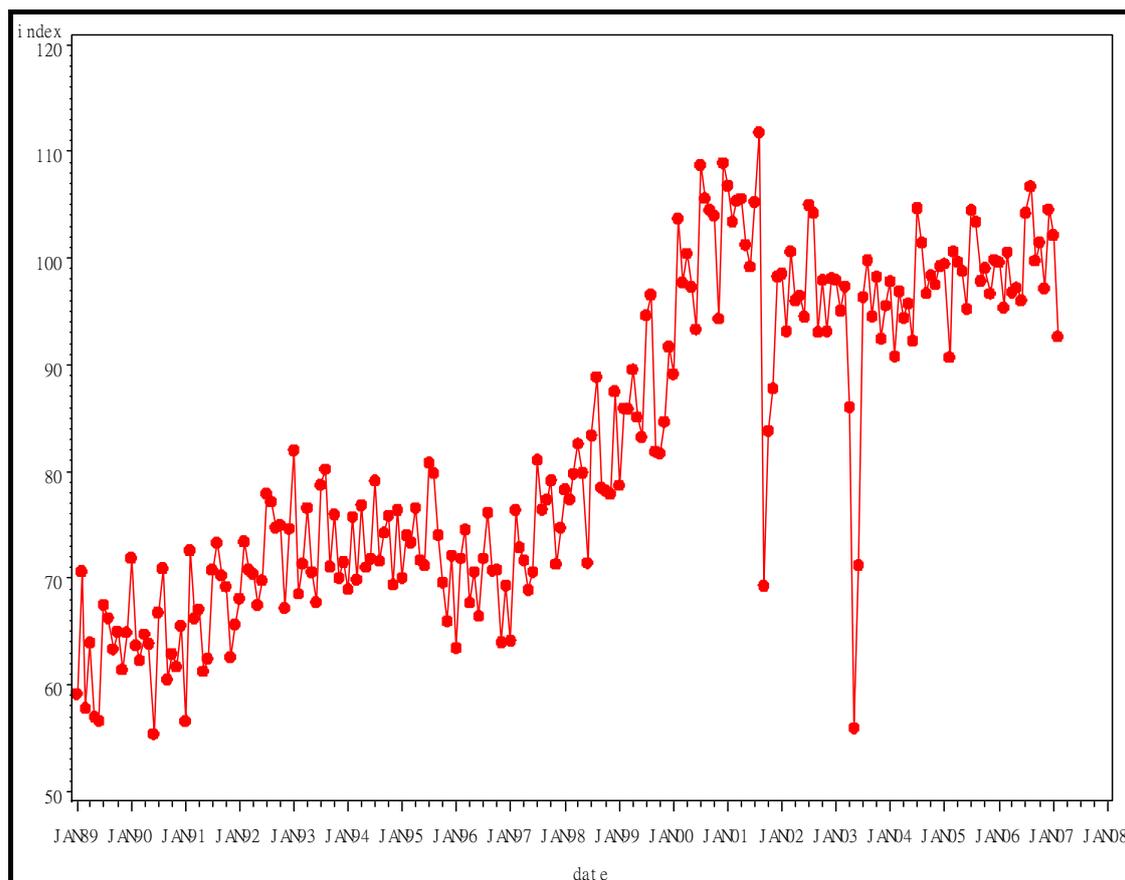


圖 二-1 原始時間序列圖

由圖二-1 可看出此鐵路客運營運指數乘客依時間有逐漸增加的長期變化之趨勢，且在每年二月、七月、八月時更是達到高峰，可知其存在季節性變動，可能受到寒假返鄉過節與暑假旅遊的影響而形成。圖中還有兩個時間點的指數較特別，一個是 2001 年 9 月受到納莉颱風侵台的影響，一個是 2003 年 5 月受到 SARS 疫情的影響，使得當時乘坐火車的乘客相對大量縮減，其指數相對於以往相同月份的指數低很多。原始時間序列圖變異數呈現蠻平穩的狀態，因此不需做變數轉換。

## 第一節 時間序列迴歸法 Time Series Regression

對時間序列之研究，稱為時間序列分析。時間序列的資料往往不能以迴歸分析的方法來建立模型加以分析，因為迴歸分析想要建立的是因果模型。而時間序列中之各觀測值間通常都存在相關性，時間相隔越短之兩觀測值，其相關性越大，時間序列並不滿足所謂“各觀測值為獨立”的必要假設。因此，時間序列分析和其他傳統分析不同的是，它不需藉助預測變數，僅依照變數本身過去的資料所存在的變異型態來建立模型。

### 一、 預設模型

Parameter Estimates					
Variable	Parameter Estimate	Standard Error	t Value	Pr >  t	Variance Inflation
Intercept	61.77003	1.85944	33.22	<.0001	0
t	0.19849	0.00754	26.34	<.0001	1.00173
d1	-1.64639	2.30154	-0.72	0.4752	1.8769
d2	-0.67331	2.30143	-0.29	0.7702	1.87672
d3	-0.13967	2.33311	-0.06	0.9523	1.83641
d4	-0.18594	2.3329	-0.08	0.9366	1.83609
d5	-4.63332	2.33272	-1.99	0.0483	1.8358
d6	-6.04015	2.33256	-2.59	0.0103	1.83555
d7	4.48191	2.33243	1.92	0.056	1.83534
d8	4.7812	2.33232	2.05	0.0416	1.83517
d9	-3.08063	2.33223	-1.32	0.188	1.83503
d10	-1.39412	2.33217	-0.6	0.5506	1.83494
d11	-5.54595	2.33214	-2.38	0.0183	1.83488

表 二-1 時間序列迴歸法-預設模型的參數估計表

由表二-1 可知，因為  $\beta_1$  為正，所以資料存在上升的線性趨勢，亦可由原始時間序列圖印證，故可以使用時間序列迴歸法來配適模型，預設模型為：

(台灣鐵路客運指數預測分析)

$$y_t = TR_t + SN_t + \varepsilon_t$$
$$= \beta_0 + \beta_1 t + \beta_2 M_1 + \beta_3 M_2 + \beta_4 M_3 + \beta_5 + \dots + \beta_{12} M_{11} + \varepsilon_t$$

$t$  為時間

$y_t$  為在連續的時間下，第  $t$  點所觀察到的值

$M_1, M_2, \dots, M_{11}$  為季節虛擬變數

For example :  $M_1 = \begin{cases} 1, & \text{if period } t \text{ is January.} \\ 0, & \text{otherwise.} \end{cases}$

## 二、 診斷模型是否適合

Ordinary Least Squares Estimates			
SSE	10034.5865	DFE	205
MSE	48.9492	Root MSE	6.99637
SBC	1523.4426	AIC	1479.44416
Regress R-Square	0.7835	Total R-Square	0.7835
Durbin-Watson	0.6929	Pr < DW	<.0001
Pr > DW	1		

表 二-2 時間序列迴歸之預設模型的統計量

由表二-2 可得知，Durbin-Watson 值為 0.6929，所以此模型的殘差有一階正自我相關；再看 Pr<DW 的部分，若 Pr<DW 為顯著，則表示殘差有一階正自我相關，若 Pr>DW 為顯著，則表示殘差有一階負自我相關。由上述可判斷此線性迴歸模型的殘差有一階正自我相關，所以需要修正預設模型。

### 三、修正後之模型

因為預設模型的殘差有一階正自我相關，故需要修改原本預設模型，修正後的線性迴歸模型為：

$$y_t = \beta_0 + \beta_1 t + \beta_2 M_1 + \beta_3 M_2 + \beta_4 M_3 + \beta_5 + \cdots + \beta_{12} M_{11} + \varepsilon_t$$

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

$t$  為時間

$y_t$  為在連續的時間下，第  $t$  點所觀察到的值

$M_1, M_2, \dots, M_{11}$  為季節虛擬變數

For example :  $M_1 = \begin{cases} 1, & \text{if period } t \text{ is January.} \\ 0, & \text{otherwise.} \end{cases}$

### 四、診斷修正後模型是否適合

Yule-Walker Estimates			
SSE	5776.81856	DFE	204
MSE	28.31774	Root MSE	5.32144
SBC	1408.99242	AIC	1361.60949
Regress R-Square	0.5664	Total R-Square	0.8754
Durbin-Watson	2.2427	Pr < DW	0.69597
Pr > DW	0.3403		

表 二-3 時間序列迴歸法-修正後線性迴歸模型的統計量

由表二-3 可得知，Durbin-Watson 值為 2.2427，表示此模型的殘差有沒有一階正負自我相關；再看 Pr<DW 的部分，若 Pr<DW 為顯著，則表示殘差有一階正自我相關，若 Pr>DW 為顯著，則表示殘差有一階負自我相關。由上述可判斷此線性迴歸模型的殘差無一階正自我相關，亦無一階正自我相關，表示此修正後模型為合適的。

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.646665	0.053405	-12.11

表 二-4 時間序列迴歸法-模型的自迴歸參數

The AUTOREG Procedure					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	61.9888	2.389	25.95	<.0001
t	1	0.1961	0.0158	12.4	<.0001
d1	1	-1.6056	1.374	-1.17	0.2439
d2	1	-0.6285	1.7439	-0.36	0.7189
d3	1	-0.5168	1.9616	-0.26	0.7925
d4	1	-0.4154	2.0777	-0.2	0.8417
d5	1	-4.7669	2.1355	-2.23	0.0267
d6	1	-6.1116	2.1527	-2.84	0.005
d7	1	4.4503	2.1345	2.08	0.0383
d8	1	4.7744	2.0769	2.3	0.0225
d9	1	-3.0733	1.9645	-1.56	0.1193
d10	1	-1.3811	1.7617	-0.78	0.434
d11	1	-5.5351	1.3777	-4.02	<.0001

表 二-5 時間序列迴歸法-修正後線性迴歸模型的參數估計表

由表二-4 與表二-5 可以得知參數估計為下：

$$\begin{aligned} \hat{\beta}_0 &= 61.9888 & \hat{\beta}_1 &= 0.1961 & \hat{\beta}_2 &= -1.6056 & \hat{\beta}_3 &= -0.6285 \\ \hat{\beta}_4 &= -0.5168 & \hat{\beta}_5 &= -0.4154 & \hat{\beta}_6 &= -4.7669 & \hat{\beta}_7 &= -6.1116 \\ \hat{\beta}_8 &= 4.4503 & \hat{\beta}_9 &= 4.7744 & \hat{\beta}_{10} &= -3.0733 & \hat{\beta}_{11} &= -1.3811 \\ \hat{\beta}_{12} &= -5.5351 & \hat{\phi}_1 &= -0.646665 & \hat{\sigma} &= 5.32144 \end{aligned}$$

## 五、 預測

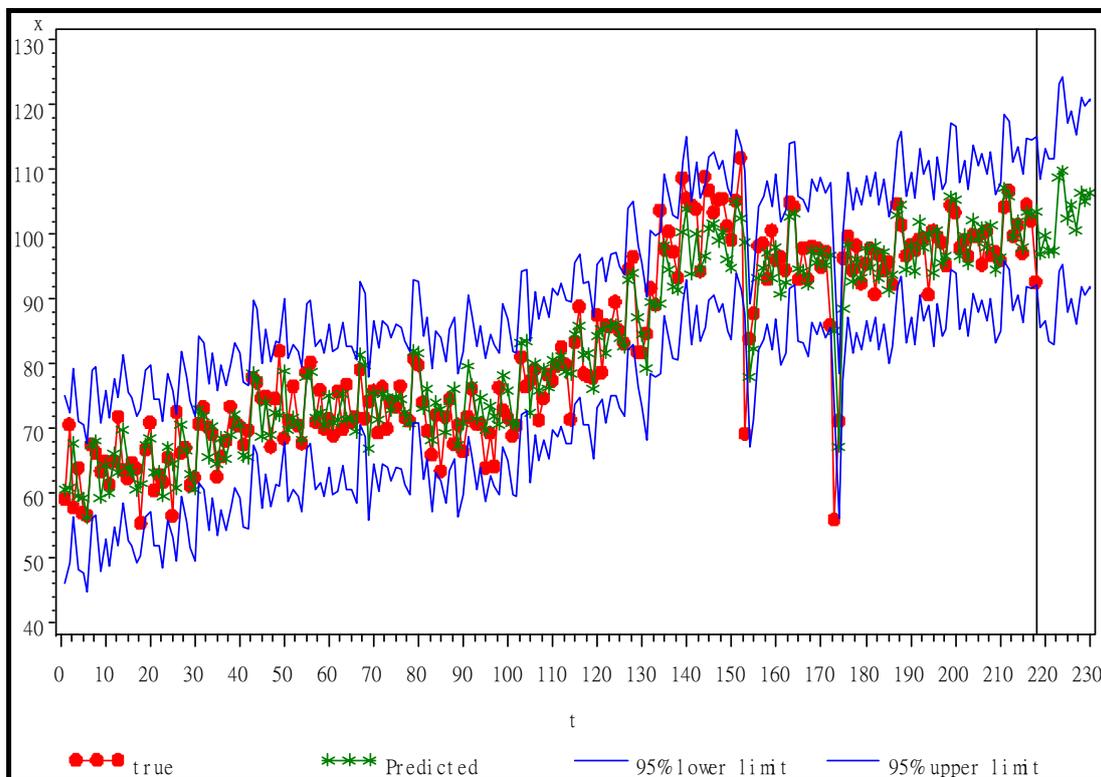


圖 二-2 時間序列迴歸法-修正後線性迴歸模型的 95%信賴水準之預測圖

日期	實際值	預測值	95%信賴區 間下限	95%信賴區 間上限	實際值與預 測值之差
07年03月	104	99.4717	89.1905	109.753	4.5283
07年04月	97.67	99.1005	87.9006	110.3005	-1.4305
07年05月	98.08	94.7783	82.644	106.9126	3.3017
07年06月	96.36	93.4916	80.4066	106.5767	2.8684
07年07月	107.94	104.0276	89.9751	118.0801	3.9124
07年08月	99.51	104.1752	89.1384	119.2119	-4.6652
07年09月	99.9	96.1922	80.1542	112.2303	3.7078
07年10月	97.42	98.088	81.0316	115.1444	-0.668
07年11月	97.03	94.0055	75.9138	112.0972	3.0245
07年12月	103.33	99.4746	80.3307	118.6186	3.8554
08年01月	100.45	97.6702	77.4571	117.8832	2.7798
08年02月	93.24	98.4586	77.1596	119.7575	-5.2186

表 二-6 時間序列迴歸法-實際值與預測值的差

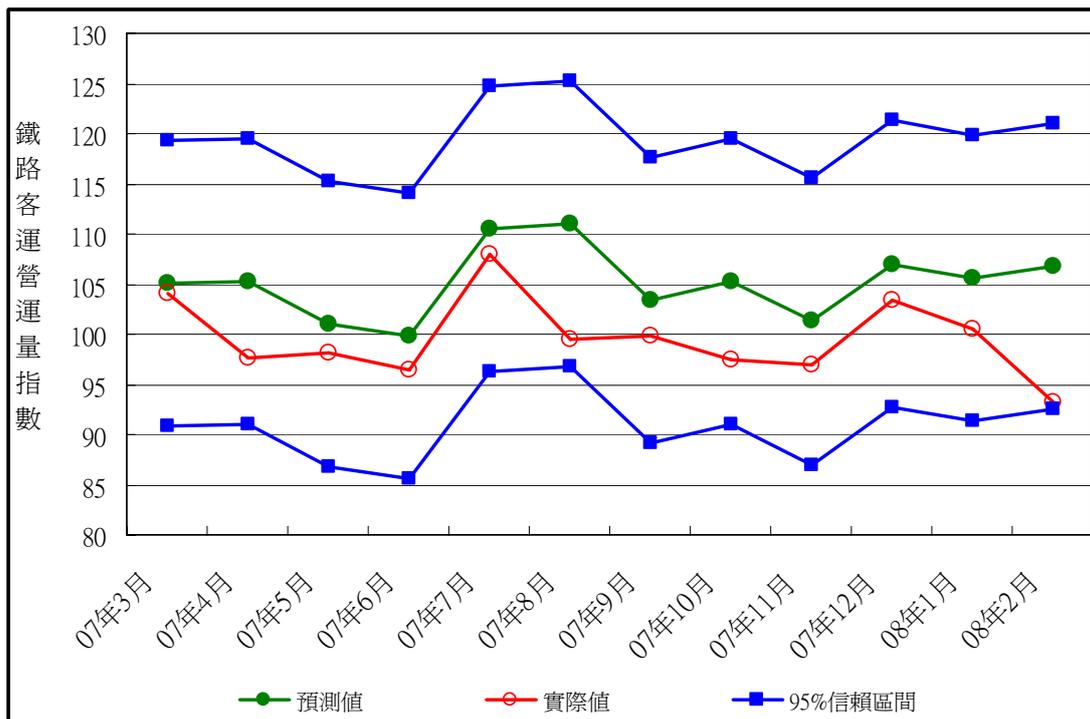


圖 二-3 時間序列迴歸法-修正後線性迴歸模型的未來一年之預測圖

從圖二-2 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，除了 2001 年 9 月受到納莉颱風侵台影響，與 2003 年 4 月到 7 月受到 SARS 疫情影響的兩個時間點，都比預測的信賴區間超出一點點，其餘預測值的結果與實際值相去不遠，表示預測結果相當不錯。再由圖二-3 未來一年的預測結果，實際值都維持在預測區間內，而預測值與真實值也都相對接近，可以看出此模型預測結果良好。

## 第二節 指數平滑法

指數平滑法比起其他複雜的加權平均法而言，具有容易使用與容易了解的特性。其計算方法是將前一期的預測值為基礎，再加上預測值與實際值之間差額的百分比。至於對平滑常數之要求及數值的特性與變化之意義為誤差值預測值愈陡峭對需求變數反應愈快；若誤差值預測值愈平滑對需求變數反應愈慢。

因此選擇平滑係數，基本上是以判斷或嘗試錯誤的方式，以預測之誤差結果來指引決策。而選擇平滑係數的目的，即在平滑係數變動的利益，以及當時發生時，能與反應真實變化的利益之間取得平衡。

## 二、預設模型

因為我們的資料具有季節變動，而且這變動保持一致，因此我們所使用的指數平滑法為 Additive Holt-Winters model，預設模型如下：

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

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

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

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

## 三、診斷配適模型後之 ACF 和 PACF 圖

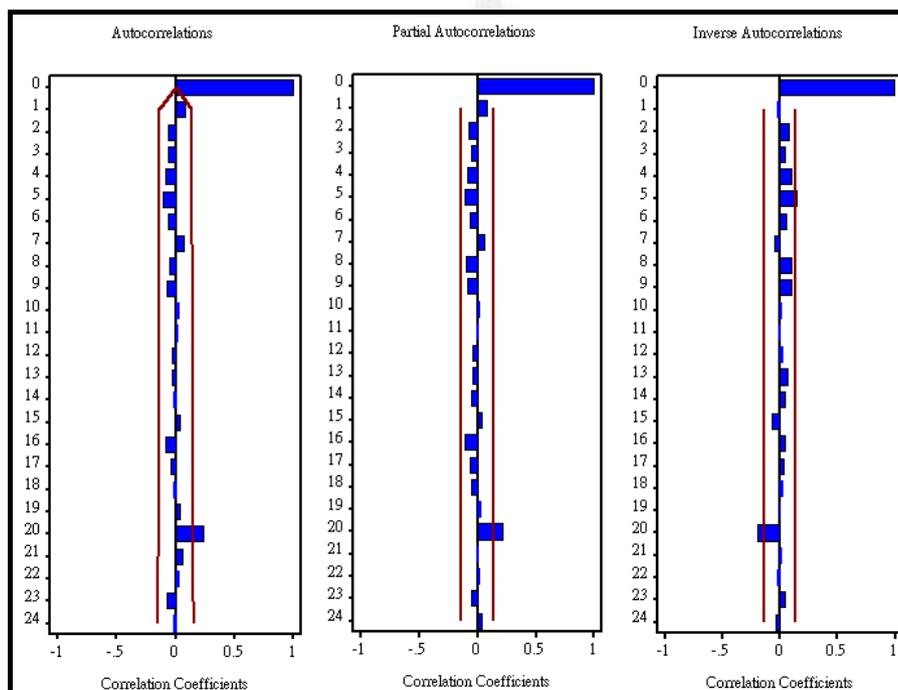


圖 二-4 指數平滑法-配模後之 ACF、PACF 圖

由圖二-4 我們可以看到，大部分相關係數的值皆在兩倍標準差以內，但在 lag20 的地方有超過，所以我們要再做進一步的討論，才能判定此模式是否適合。

#### 四、模型之配適值

Model Parameter	Estimate	Std. Error	T	Prob> T
LEVEL Smoothing Weight	0.40998	0.0386	10.6153	<.0001
TREND Smoothing Weight	0.05389	0.0271	1.9859	0.0483
SEASONAL Smoothing Weight	0.06224	0.0315	1.9747	0.0496
Residual Variance (sigma squared)	27.51666	.	.	.
Smoothed Level	98.39641	.	.	.
Smoothed Trend	-0.01946	.	.	.
Smoothed Seasonal Factor 1	-0.51226	.	.	.
Smoothed Seasonal Factor 2	0.2956	.	.	.
Smoothed Seasonal Factor 3	1.09478	.	.	.
Smoothed Seasonal Factor 4	0.74304	.	.	.
Smoothed Seasonal Factor 5	-3.55975	.	.	.
Smoothed Seasonal Factor 6	-4.82695	.	.	.
Smoothed Seasonal Factor 7	5.72845	.	.	.
Smoothed Seasonal Factor 8	5.89549	.	.	.
Smoothed Seasonal Factor 9	-2.06803	.	.	.
Smoothed Seasonal Factor 10	-0.15277	.	.	.
Smoothed Seasonal Factor 11	-4.21588	.	.	.
Smoothed Seasonal Factor 12	1.27276	.	.	.

表 二-7 指數平滑法-參數估計表

由上表二-7 可以知道參數估計如下：

$$\hat{\alpha} = 0.40998$$

$$\hat{\gamma} = 0.05389$$

$$\hat{\delta} = 0.06224$$

$$\hat{\sigma}^2 = 27.51666$$

其中  $\hat{\alpha}$  為 水平項的權重、 $\hat{\gamma}$  為趨勢項的權重、 $\hat{\delta}$  為季節項的權重；而我們看模型的平滑季節因子裡，5 月、6 月、11 月的數值相對於其他因子來的小因此我們知道這幾個月旅客人數較少；而 7 月、8 月的數值相對於其他因子來的大，因此我們可以得出這幾個月旅客人數較其他月分來的多。

## 五、診斷模型

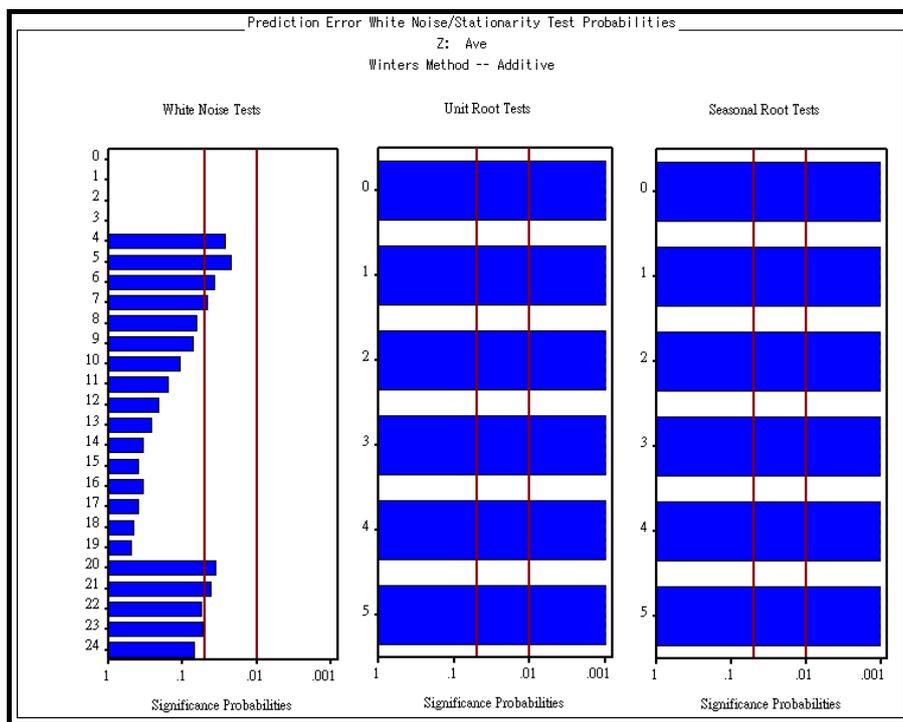


圖 二-5 指數平滑法-White Noise 檢定與單根檢定

由表二-7 可知季節與趨勢之 p-value 皆小於 0.05，即此兩個參數顯著，兩變數應留在模型當中，因此我們認為此模式恰當；而由圖二-5 White Noise 我們可以看到，在 Lag4、Lag5、Lag6、Lag7、Lag20、Lag21 的地方其 p-value 值皆大於 0.05，即此模式不符合殘差項假設；單根檢定的 p-value 的值皆小於 0.001，我們可知此模式為一個平穩的序列。而我們知道，殘差項假設為建立模型的基本假設，當我們做出的模型不符合此基本假設時，我們會認為此模型不合適。

## 六、預測

日期	實際值	預測值	95%信賴區 間下限	95%信賴區 間上限	實際值與預 測值之差
07年03月	104	99.4717	89.1905	109.753	4.5283
07年04月	97.67	99.1005	87.9006	110.3005	-1.4305
07年05月	98.08	94.7783	82.644	106.9126	3.3017
07年06月	96.36	93.4916	80.4066	106.5767	2.8684
07年07月	107.94	104.0276	89.9751	118.0801	3.9124
07年08月	99.51	104.1752	89.1384	119.2119	-4.6652
07年09月	99.9	96.1922	80.1542	112.2303	3.7078
07年10月	97.42	98.088	81.0316	115.1444	-0.668
07年11月	97.03	94.0055	75.9138	112.0972	3.0245
07年12月	103.33	99.4746	80.3307	118.6186	3.8554
08年01月	100.45	97.6702	77.4571	117.8832	2.7798
08年02月	93.24	98.4586	77.1596	119.7575	-5.2186

表 二-8 指數平滑法-預測值

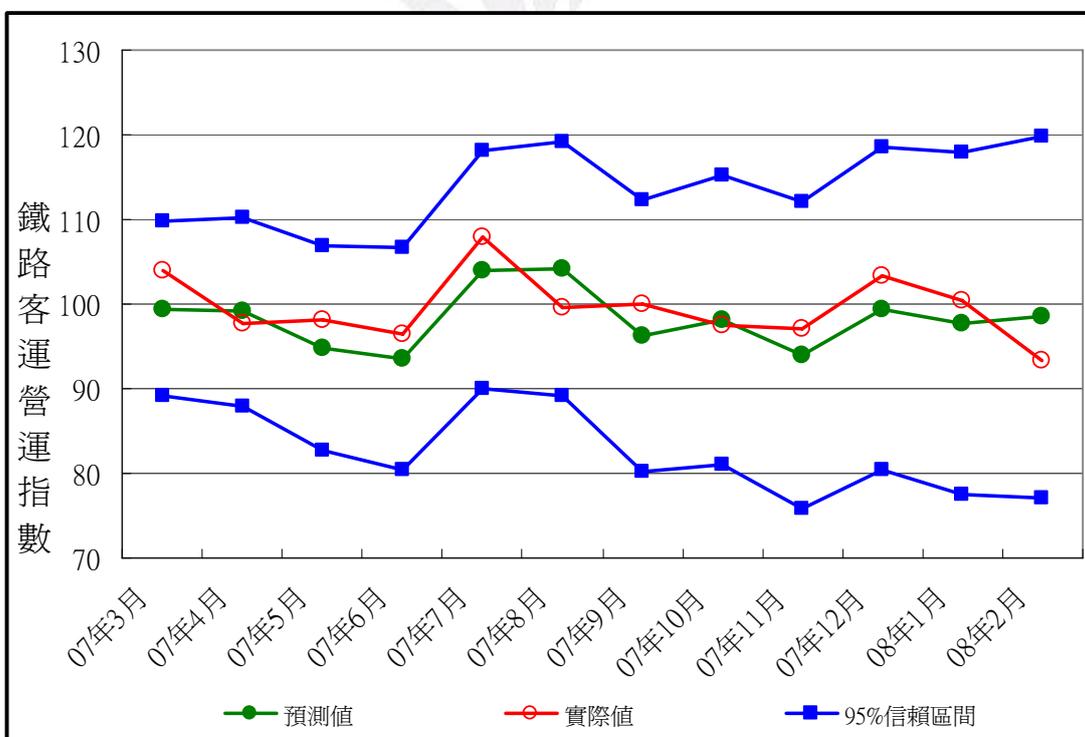


圖 二-6 指數平滑法-預測圖

從圖二-6 的預測圖我們可以看出，95%的信賴區間一開始比較小，隨著時間的進行越來越大，表示出越到後期，模型的預測準確度越低。而看表二 9 的實際值和預測值的差距那一欄，發現除了 2007 年 8 月和 2008 年 2 月的差距比較多以外，其他的點都預測的很好。在查過相關文獻之後發現，2007 年 8 月時由於強度颱風聖帕(SEPAT) 侵台造成火車停駛、台鐵七堵站 28 日發生號誌故障，造成 19 列客車停駛，67 列客車延誤達 2,915 分鐘因此當月份累計搭乘指數較以往來的低；而 2008 年 2 月則是往年搭乘台鐵的人轉而搭乘高鐵。這個影響我們從高鐵的搭乘人次從一月的一百一十五萬人次增加到二月的兩百零九萬人次就可以看出。因此從整體上來說，扣除掉某些特殊點以外，此模型之預測能力相當不錯。

## 第五節 分解法 Decomposition method

一般將時間序列分解成四個成份(Component)，趨勢、循環效應、季節效應、隨機性；有許多不同的方式可以用來分解時間序列資料，若將每一成份獨立出來，會使預測更準確。分解法可分為加法模型和乘法模型，依據此兩種模型，分析時間序列的方法亦有兩種：

1. 如果時間序列屬於相加模型，則可從序列中減去某種影響成份的變動，來求出另一種成份的變動。
2. 如果時間序列屬於相乘模型，則可將其他成份除時間序列，來求出某種影響成份的變動。然而，在時間序列分析中，大都採用乘法模型的假設，因乘法模型比加法模型更能正確代表一時間序列。

# 一、 分析

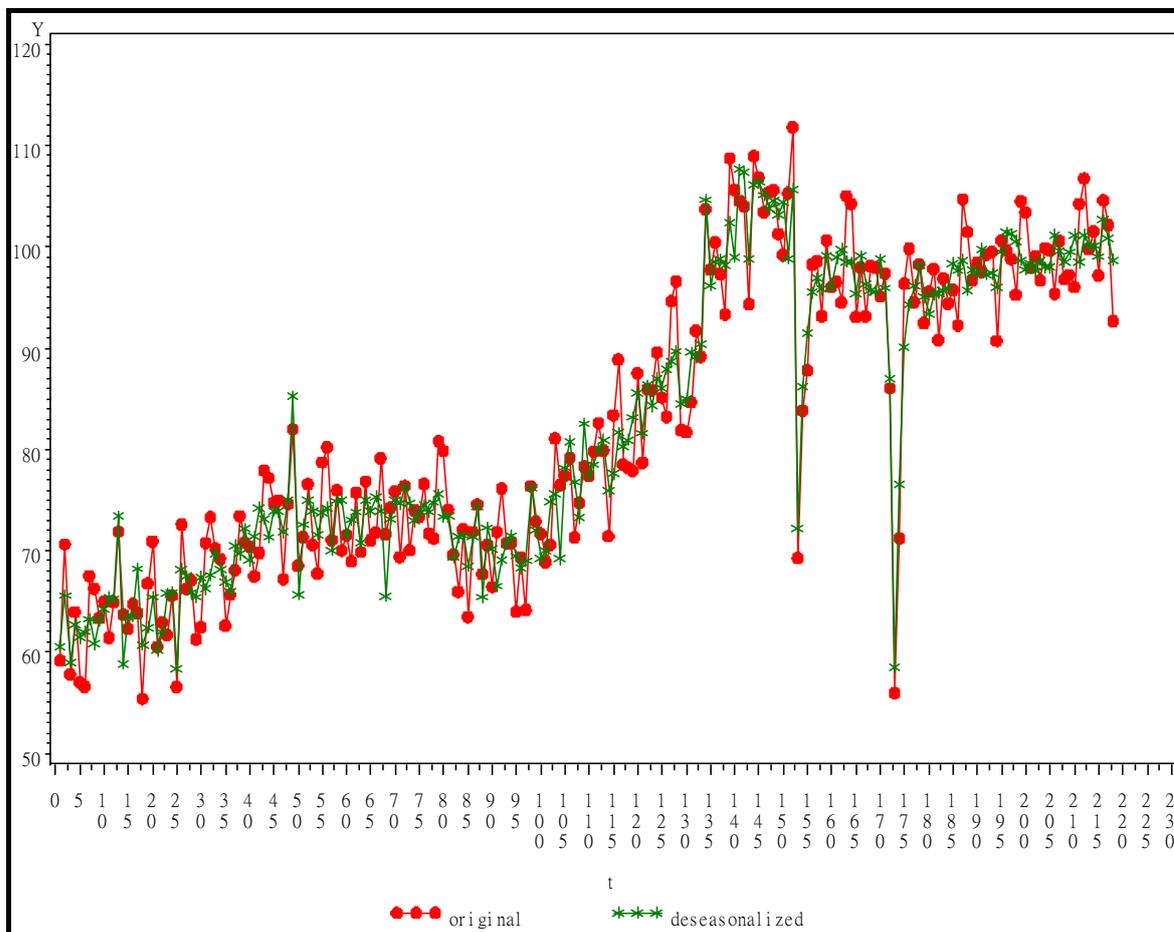


圖 二-7 分解法-去季節因子的時間序列圖

Parameter Estimates					
Variable	D F	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	60.58075	0.89739	67.51	<.0001
t	1	0.19865	0.00711	27.96	<.0001

表 二-9 分解法-去季節因子的參數估計表

由圖二-7 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，再由表二-9 亦可判斷出時間和鐵路客運營運量指數呈現正向趨勢的關係；此判斷結果代表即使沒有季節因素的影響鐵路客運營運量指數也會隨著時間而逐漸增加。

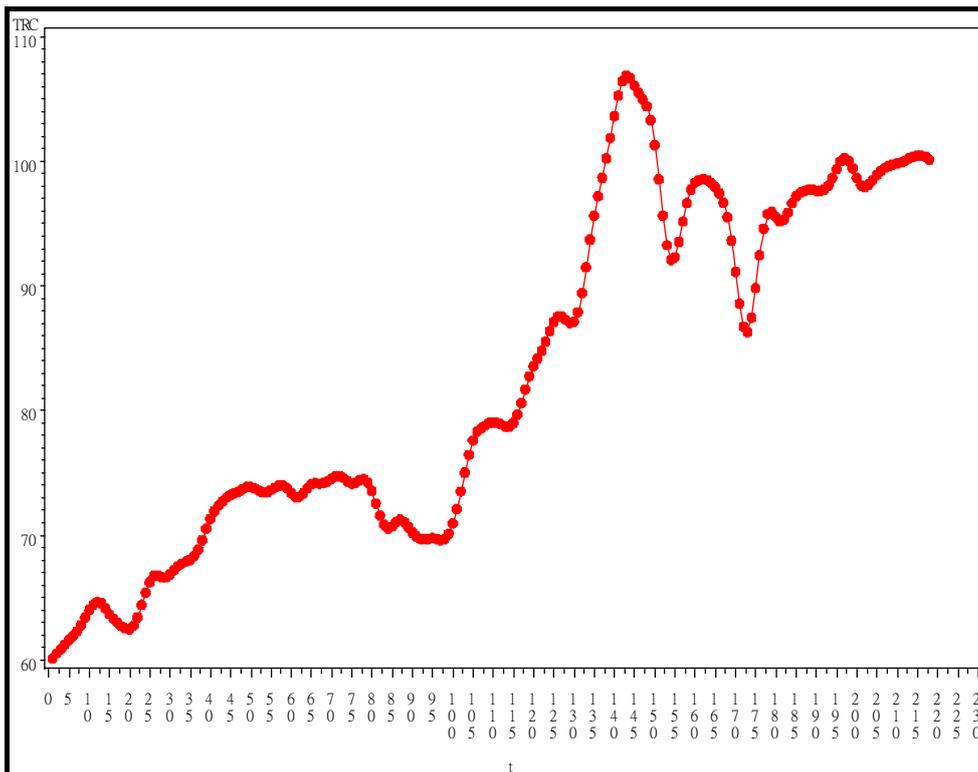


圖 二-8 分解法-趨勢循環之時間序列圖

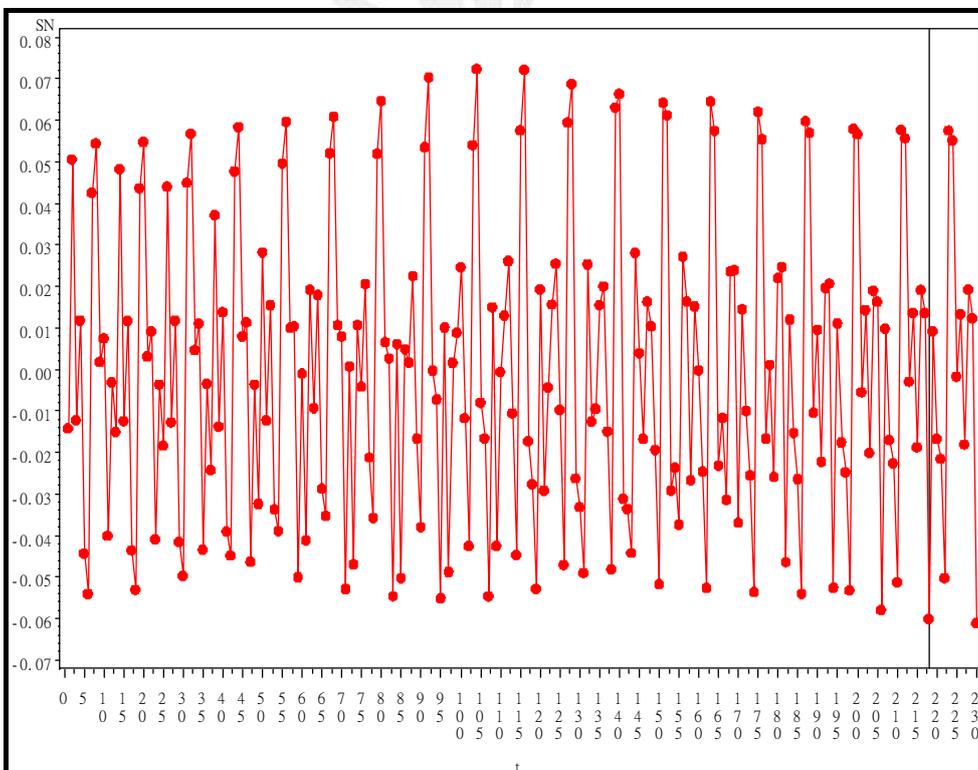


圖 二-9 分解法-季節因子之時間序列圖



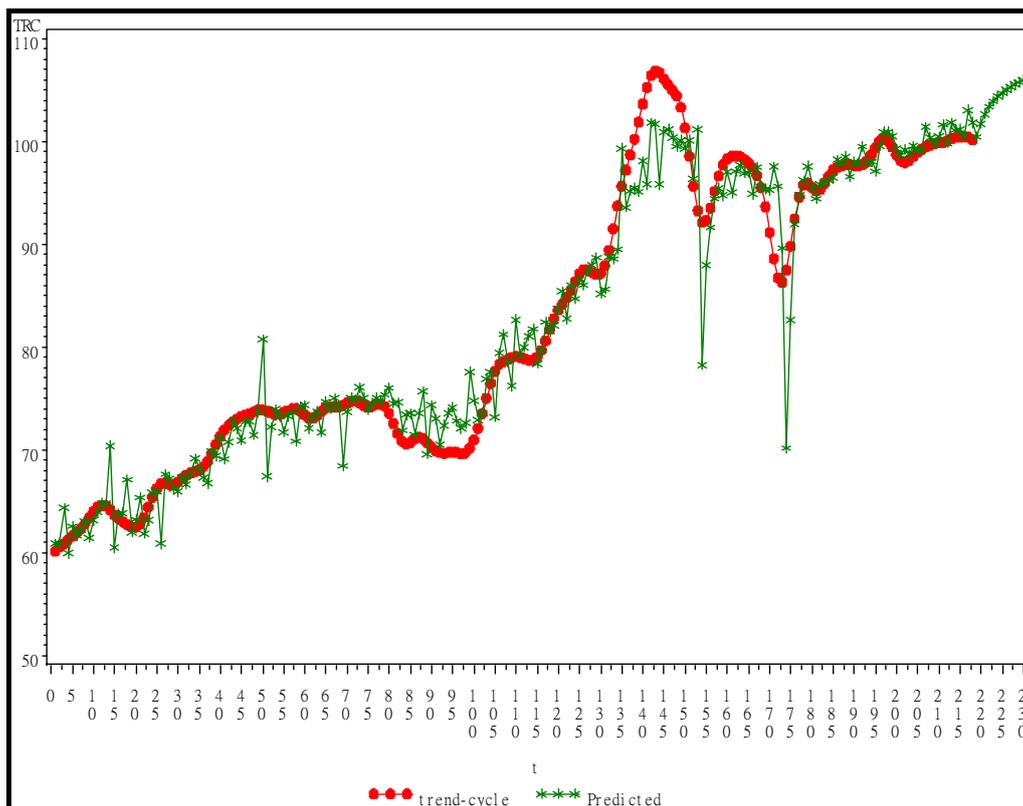


圖 二-11 分解法-加法模型的預測值描述趨勢循環圖

### 三、 診斷模型

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.6844	0.049725	-13.76

表 二-10 分解法-加法模型的自迴歸參數

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	60.674	2.0249	29.96	<.0001
t	1	0.1973	0.016	12.37	<.0001

表 二-11 分解法-加法模型參數估計表

由表二-11 可以看到，參數估計的 P 值皆小於顯著水準  $\alpha=0.05$ ，故拒絕虛無假設，也就是說參數應該留在模型裡，即此模型為合適。由表二-10 與二-11 參數估計表可以寫出下列去季節因子後的線性迴歸模型：

$$d_t = 60.674 + 0.1973t$$

$$\varepsilon_t = 0.6844\varepsilon_{t-1} + a_t$$

#### 四、 預測

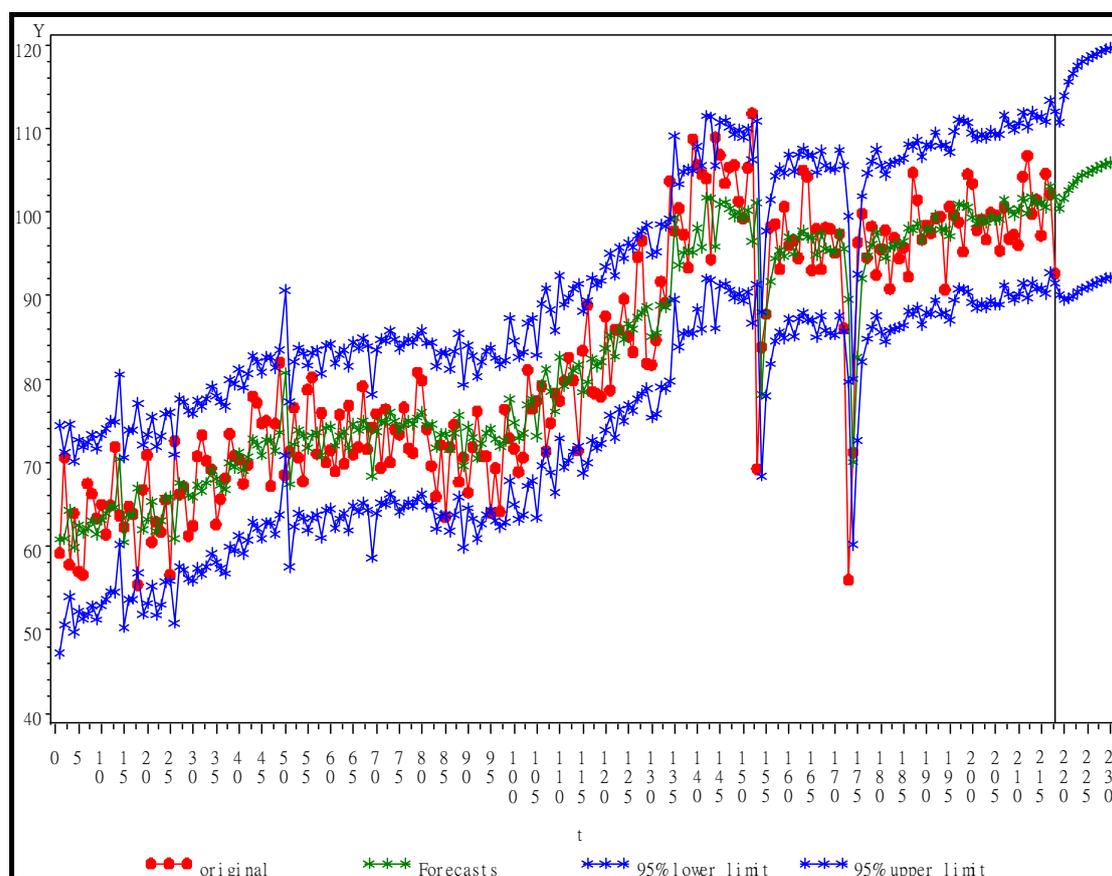


圖 二-12 分解法-95%信賴水準預測圖

日期	實際值	預測值	95%信賴區間下限	95%信賴區間上限	實際值與預測值之差
07年03月	104	100.471	90.1682	110.773	3.529
07年04月	97.67	101.723	89.5324	113.913	-4.053
07年05月	98.08	102.655	89.67	115.641	-4.575
07年06月	96.36	103.33	89.9845	116.676	-6.97
07年07月	107.94	103.982	90.4663	117.498	3.958
07年08月	99.51	104.414	90.8151	118.013	-4.904
07年09月	99.9	104.717	91.0745	118.359	-4.817
07年10月	97.42	105.04	91.3732	118.708	-7.62
07年11月	97.03	105.282	91.5991	118.966	-8.252
07年12月	103.33	105.569	91.8736	119.264	-2.239
08年01月	100.45	105.795	92.0893	119.501	-5.345
08年02月	93.24	105.943	92.2282	119.658	-12.703

表 二-12 分解法-調整後實際值與預測值

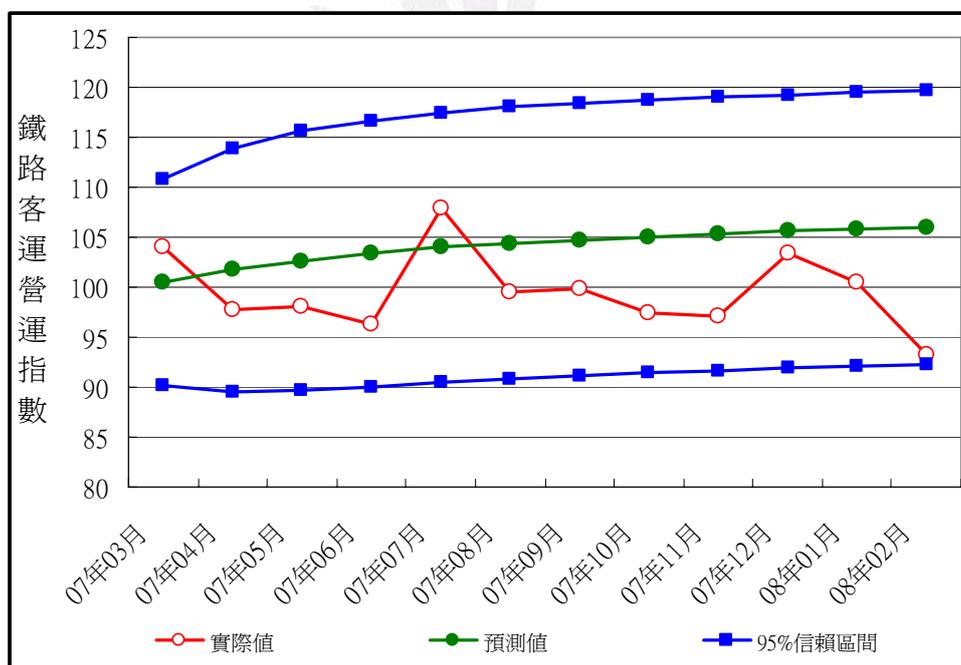


圖 二-13 分解法-調整後 95%的預測區間

從圖二-13 來看，預測值和實際值大致上都在 95%的預測區間，以信賴區間的角度來看此模型的預測能力還不錯。可是在後半段部分預測值和實際值比較沒

有前半部來的靠近，所以，在前面幾筆預測來的比後面佳。

## 第五節 ARIMA 之介入分析

移動平均法(moving average)係以過去資料為依據，將最近 N 期資料之算術平均數值，做為下一期之預測值，在部分自我相關函數(Partial Autocorrelation Function, PACF)中，時間序列為不平穩，即下降緩慢時所配適模型。自我迴歸(AutoRegressive)在自我相關函數(Autocorrelation Function, ACF)中，時間序列為不平穩，即下降緩慢時所配適模型。在 ACF 與 PACF 時間序列皆不平時，則可配適 ARMA(AutoRegressive And Moving Average)模型。當時間序列平均數不平穩時，則可使用一階差分(Integrated)使平均數達到平穩；但是當平均數呈季節變動不平穩時，則可使用季節差分。所以 ARIMA 為 AR 模式、差分與 MA 模式的組合模式。而介入分析是用來分析某一件事或政策(稱之介入因子)對於時間數列模型之影響。

介入分析的步驟：

1. 對介入時間點前之數列，進行 ARMA 模型配適。
2. 將介入因子的時間，以虛擬變數(dummy variable)方式建立新變數(介入變數)。
3. 重新進行模型配適及參數估計。
4. 解釋介入因子的係數對數列的影響。

時間序列經常受到特殊的外部事件影響，譬如主要公司；自然災害；政治或經濟政策主動性或變動；技術變動；停工鬥爭；推銷活動；做廣告等等。這些外部事件共同地為人所知作為干預，所謂的干預影響變數預測，這樣干預為人所知，我們也許或希望評估這些外部事件的作用或合併干預我們的時間數列模型可能改進參量估計或展望。介入分析又稱干預分析(或衝擊分析)。使用特殊類型的虛擬變數，所謂的階梯函數和衝擊函數，建立干預模式。Box 和 Tiao(1975 年)提供了一個程序，分析時間序列的存在已知的外部事件。這一程序已成為被稱為干預(或影響)分析。在他們的方法，時間序列為代表的是兩個截然不同的部分組成：一個潛在的干擾來說，和一套干預對系列。

在唯一干預情況下，干預模型的形式是

$$Y_t = C + \frac{\omega(B)}{\delta(B)} I_t + N_t$$

$I_t$  是一個二元向量指標(即是一種向量假設值為 0 或 1)定義時期的干預。

$\frac{\omega(B)}{\delta(B)}$  是一個表徵的影響的干預

$N_t$  是所謂的干擾  $N_t = Y_t - C + \frac{\omega(B)}{\delta(B)} I_t$

我們假設  $N_t$  可能仿照作為 ARIMA 過程中所界定的。在案件有沒有外源的事件，

那麼，模型  $Y_t$  減少到 ARIMA 模型討論了以前。

假設時間序列的過程中可以說是由以下過程：

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t$$

$$\phi(B) = (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q)$$

d : 有多少差分生成一個固定的數據。

假設差分 d=0

$$y_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

上述模型沒有考慮到的數據要進行干預。

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b I_t + \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

**階梯函數(Step Function)**

$$S_t^{(T)} = \begin{cases} 0 & \text{if } t < T \\ 1 & \text{if } t \geq T \end{cases}$$

時間序列受到外在干預，但是干預一時間為衝擊而後是

持續的漸增影響或漸減影響，或是干預是漸漸增加持續影響或是像階梯狀有上下起伏。

**衝擊函數(Pulse Function)**

$$P_t^{(T)} = \begin{cases} 0 & \text{if } t \neq T \\ 1 & \text{if } t = T \end{cases}$$

時間序列受到外在干預，但是但是干預一時間為衝擊而後下

降。  $P_t^{(T)} = (1-B)S_t^{(T)}$

## 一、 分析原始資料

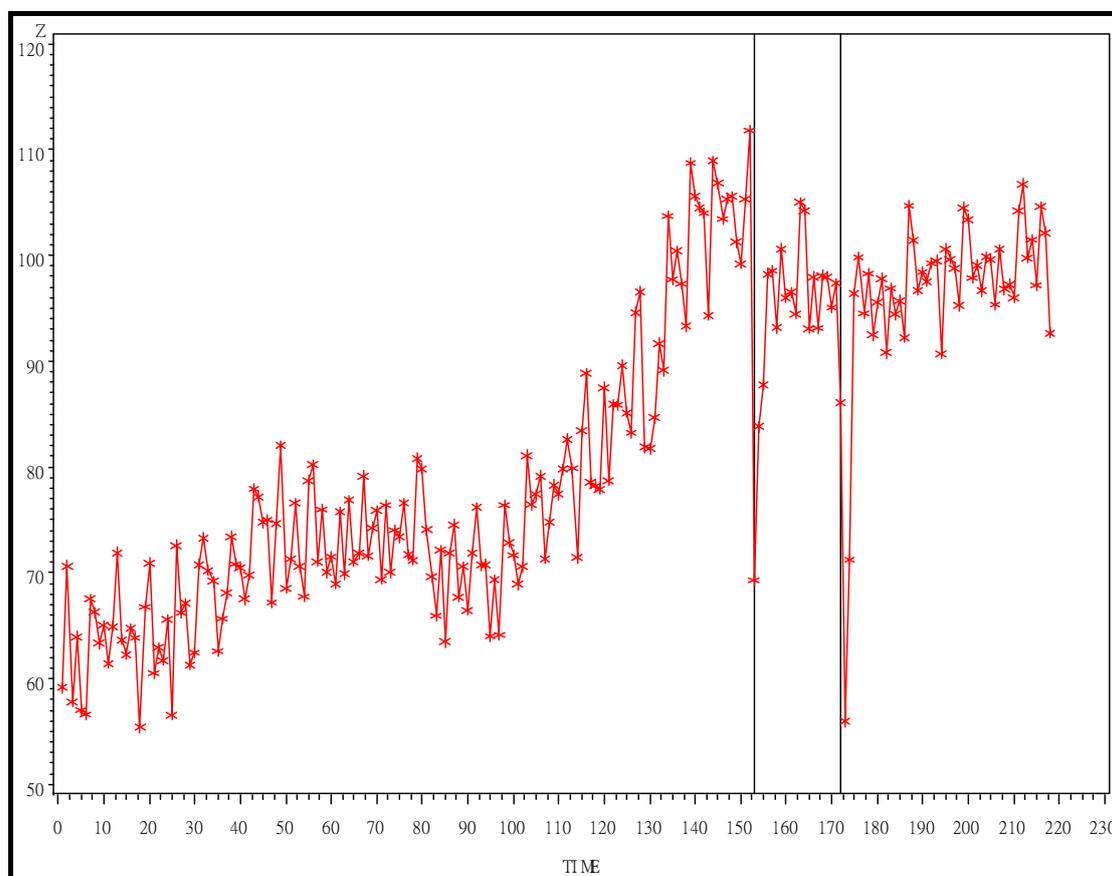


圖 二-14 ARIMA 之介入分析-原始時間序列圖

上圖二-14 為原始時間序列圖，在畫線部分分別為第 153 筆及第 172 筆資料，經我們調查，在這兩個時間點分別為 2001 年 9 月受到納莉颱風侵台影響，以及 2003 年 4 月到 7 月受到 SARS 疫情影響，所以才會導致搭乘火車的人數驟減，故將這兩個影響點加入模型。

## 二、 判斷是否需要做差分

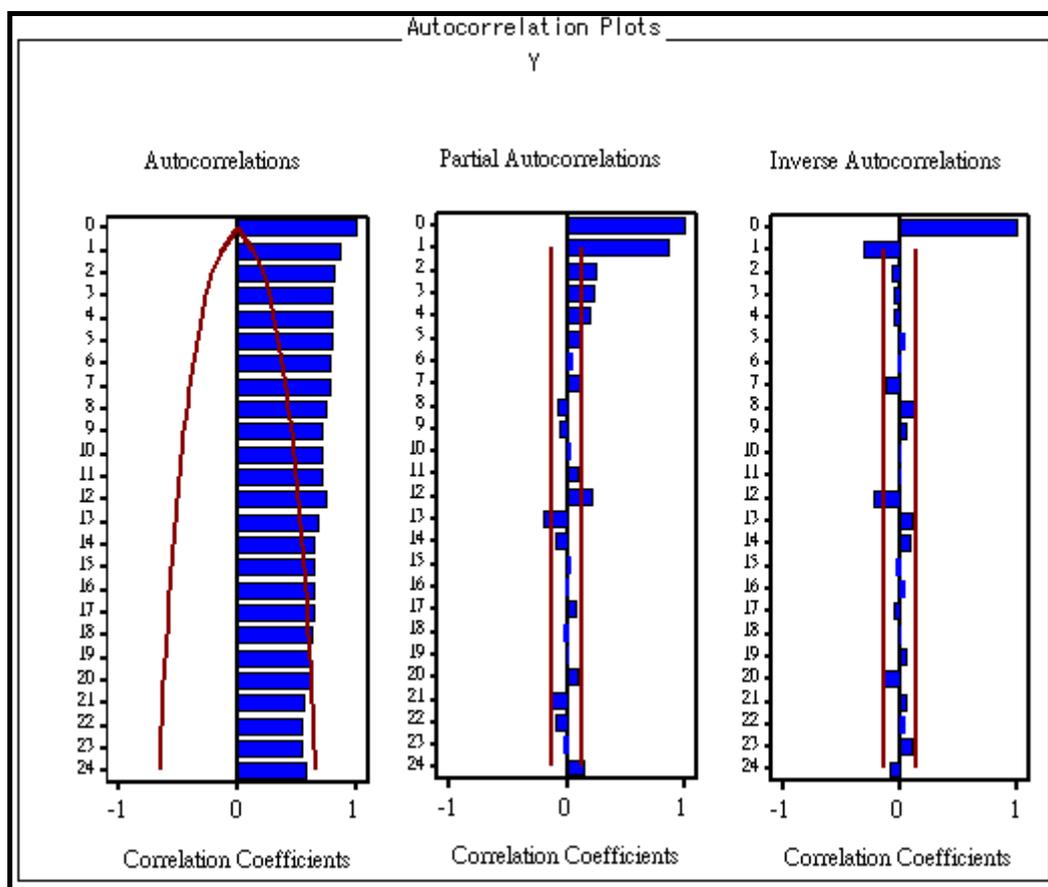


圖 二-15 ARIMA 之介入分析-原始 ACF、PACF 圖

由圖二-15 之 ACF 圖可以看出，相關係數的數值是下降的速度非常的緩慢。因此我們可以知道此資料之平均數呈現一個不平穩的序列。由於 ACF 圖其震幅下降的速度非常緩慢，所以判斷資料需要做差分轉換，進一步的從 ACF 圖可以看到，殘差在 Lag12、Lag24、Lag36...的部分皆有著特別突出的現象，明顯的此筆資料隨著時間的進行有擺盪的現象，所以判斷此筆資料需要做季節差分；我們嘗試對資料做一階差分與季節差分，並且互相比較，發現做季節差分比做一階差分收斂較快；但同時做一階差分與季節差分，卻又過度差分，所以，綜合以上判斷結果認為對資料做季節差分最為適當。

### 三、 季節差分

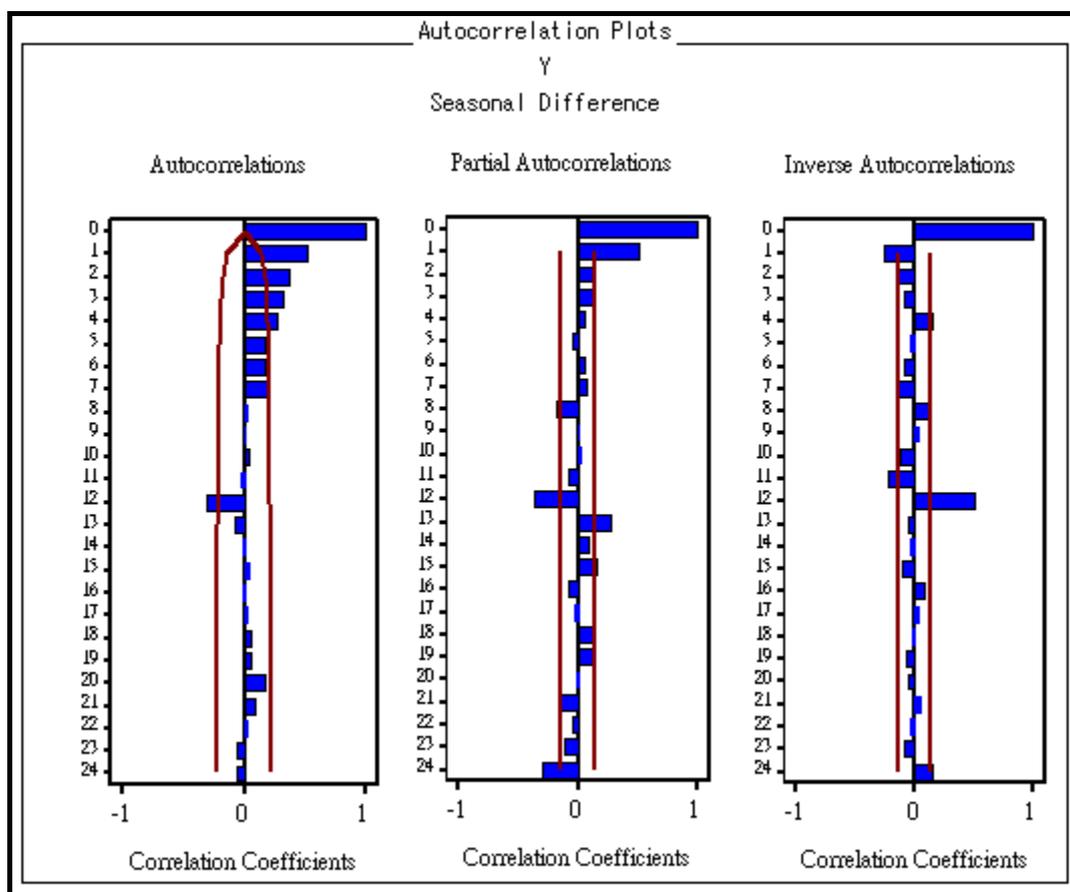


圖 二-16 ARIMA 之介入分析-季節差分後之 ACF、PACF 圖

從圖二-16 可以看出，經過季節差分之後，ACF 圖除了 Lag1 到 Lag4、Lag12 有很明顯的超出兩倍標準差以外，其餘殘差的值都有在兩倍標準差以內，而 PACF 圖較 ACF 圖有明顯截斷且在 Lag1、Lag8、Lag12、Lag13、Lag24 處有很明顯的超出兩倍標準差以外，其餘殘差的值也都有在兩倍標準差以內，因此利用這些特徵來選擇配適模型的參數；在配適模型的過程中，不能將全部的參數都配適到模型中，一定要符合節約參數原則，用最少的參數來配適模型，才能配適出最佳模型。

### 四、 配適模型

經過模型配適的相互比較之後，選取 ARIMA p=(1) d=(12) q=(1,12)再加上介入分析做為此法的最終模型：

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

$$(1 - \phi_1 B)z_t = (1 - \theta_1 B - \theta_{12} B^{12})a_t$$

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

$$P_{1t} = \begin{cases} 1 & \text{if } t = 153 \text{ (September 2001)} \\ 0 & \text{otherwise} \end{cases}$$

$$P_{2t} = \begin{cases} 1 & \text{if } t = 172 \text{ (April 2003)} \\ 0 & \text{otherwise} \end{cases}$$

$$z_t = \delta_0 + C_1 z_t(P_1) + C_2 z_t(P_2) + \varepsilon_t$$

$$z_t = (1 - B^{12})y_t \quad z_t(P_1) = (1 - B^{12})P_{1t} \quad z_t(P_2) = (1 - B^{12})P_{2t}$$

$$(1 - \phi_1 B)\varepsilon_t = (1 - \theta_1 B - \theta_{12} B^{12})a_t$$

### 五、 診斷模型

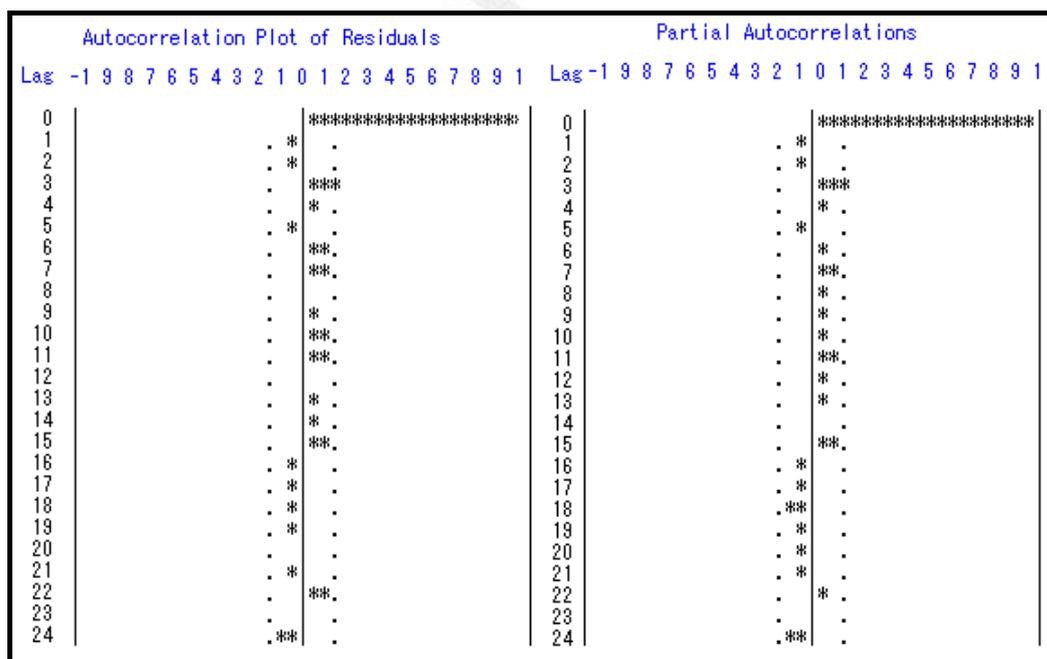


圖 二-17 ARIMA 之介入分析-配適最終模型之樣本 ACF 圖與 PACF 圖

從圖二-17 配適最終模型的樣本 ACF 圖和 PACF 圖可以看出，殘差皆在兩倍標準差內，表示配適模型後的殘差皆無自我相關，故此模型為適當模型。

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.27	3	0.0407	-0.058	-0.058	0.144	0.051	-0.062	0.078
12	13.81	9	0.1292	0.104	0.014	0.034	0.076	0.086	0.021
18	16.77	15	0.3325	0.045	0.026	0.084	-0.041	-0.025	-0.038
24	20.25	21	0.5054	-0.029	-0.019	-0.028	0.081	-0.013	-0.08
30	22.15	27	0.7299	0.051	-0.014	-0.054	0.005	0.008	-0.047
36	25.75	33	0.8114	-0.048	0.032	0.04	-0.029	0.045	0.083

表 二-13 ARIMA 之介入分析-Q 統計量

根據表二-13 的Q統計量可以得知，P值皆大於顯著水準  $\alpha=0.05$ ，所以拒絕虛無假設，在Lag6 的虛無假設為  $\rho_1=\rho_2=\dots=\rho_6$ 、Lag12 的虛無假設為  $\rho_1=\rho_2=\dots=\rho_{12}$ 、...以下虛無假設類推，經以上檢定此模型為適當模型。

## 六、 檢測模型中的參數是否適合

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag	Variable	Shift
MU	2.26311	0.6449	3.51	0.0004	0	Y	0
MA1,1	0.6937	0.05631	12.32	<.0001	1	Y	0
MA2,1	0.8847	0.0723	12.24	<.0001	12	Y	0
AR1,1	0.97921	0.01922	50.94	<.0001	1	Y	0
NUM1	-32.4602	3.73941	-8.68	<.0001	0	p1	0
DEN1,1	0.5372	0.08717	6.16	<.0001	1	p1	0
NUM2	-24.6234	3.44796	-7.14	<.0001	0	P2	0
DEN1,1	0.79356	0.06214	12.77	<.0001	1	P2	0

表 二-14 ARIMA 之介入分析-參數估計表

從表二-14 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。綜合以上的結果可推知，我們所配適的模式ARIMA(1,0,1)(0,1,1)<sub>s</sub>再加入介入分析為一個合適的模型，而參數估計值如下：

$$\begin{aligned} \text{where } \hat{\delta}_0 &= 2.26311 & \hat{\phi}_1 &= 0.97921 & \hat{\theta}_1 &= 0.6937 & \hat{\theta}_{12} &= 0.8847 \\ \hat{C}_1 &= -32.4602 & \hat{C}_2 &= -24.6234 & \hat{\delta}_1 &= 0.5372 & \hat{\delta}_2 &= 0.79356 \end{aligned}$$

由參數估計得知，在第一個影響點受到納莉颱風侵台影響，使整個鐵路客運營運指數下降 32.46，而在第二個影響點受到 SARS 疫情影響，亦使整個鐵路客運營運指數下降 24.62，所以可以知道這兩個點具有影響力，若不將這兩個影響點加入模型中，模型的預測能力可能會下降。

## 七、 預測

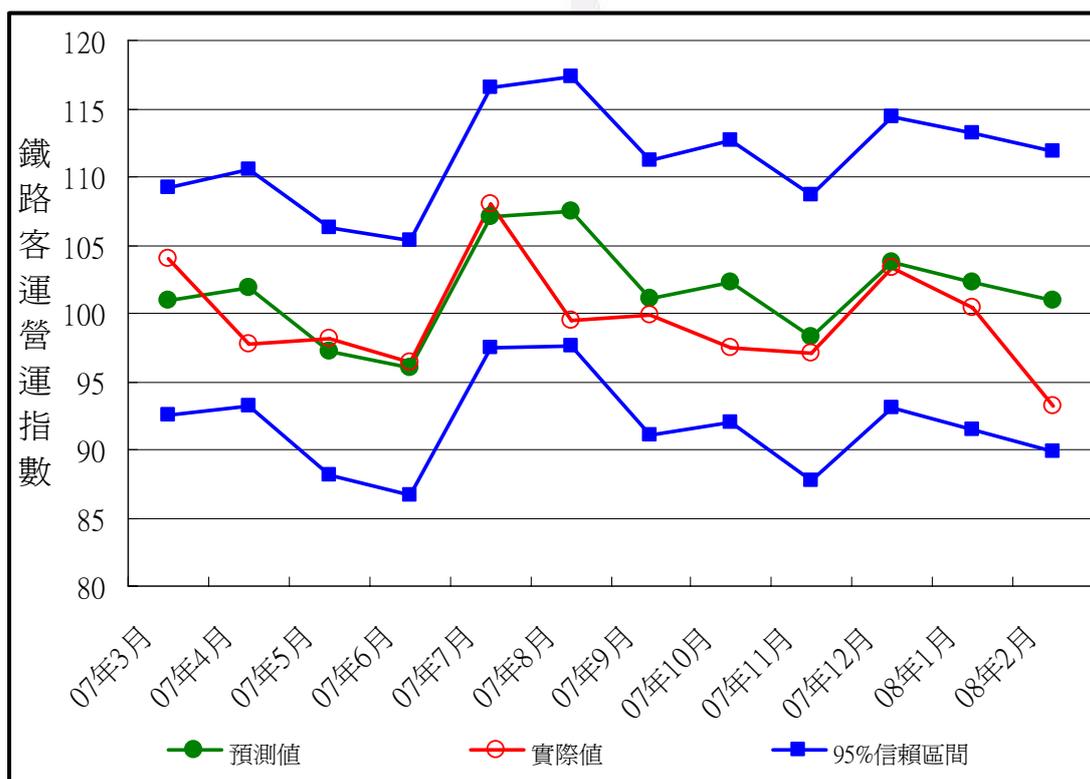


圖 二-18 ARIMA 之介入分析-95%信賴水準預測圖

日期	實際值	預測值	95%信賴區 間下限	95%信賴區 間上限	實際值與預 測值之差
07年03月	104.000	100.888	92.5101	109.265	3.112
07年04月	97.670	101.851	93.139	110.564	-4.181
07年05月	98.080	97.216	88.1944	106.238	0.864
07年06月	96.360	96.022	86.7129	105.33	0.338
07年07月	107.940	107.007	97.4318	116.583	0.933
07年08月	99.510	107.442	97.6169	117.267	-7.932
07年09月	99.900	101.102	91.0442	111.16	-1.202
07年10月	97.420	102.333	92.0558	112.609	-4.913
07年11月	97.030	98.208	87.7255	108.69	-1.178
07年12月	103.330	103.690	93.0151	114.366	-0.36
08年01月	100.450	102.330	91.4732	113.188	-1.88
08年02月	93.240	100.889	89.8595	111.918	-7.649

表 二-15 ARIMA 之介入分析-實際值與預測值的差

由圖二-18 和表二-15 我們可以看出，實際值和預測值的差距皆非常的小；因此，我們可以說我們所選定的模型其預測能力相當的不錯。

## 第五節 最佳模型

分析方法	MSE	MAE	MPE	MAPE
時間序列迴歸法	44.30018	5.58137	-5.69559	5.69559
指數平滑法	27.13798	3.32669	-0.34800	4.24202
分解法	31.88446	4.88958	-4.50729	4.94590
ARIMA介入分析	15.0787	2.87851	-2.08724	2.93535
在各準則下的最佳模型	ARIMA	ARIMA	指數平滑法	ARIMA

以 MSE、MAE 及 MPE 準則而言，其值愈小表現愈良好，反之則否；而 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。綜合上表之比較，我們可以發現 ARIMA 之介入分析法在所有準則之表現都是良好的，因此選擇 ARIMA 之介入分析配適模型為我們的最佳模式。

## 第三章 結論與建議

### 第一節 結論

根據四種方法分析結果顯示，ARIMA 之介入分析的配適模型，在 MSE、MAE、MPE 及 MAPE 四個準則呈現效果大致良好，於是選擇 ARIMA 之介入分析的配適模型為最佳模型。分析 ARIMA 之介入分析模式其預測結果，可看出幾點方向可供討論：

1. 圖 0-18 的預測圖中，雖然有些點估計與實際值有差距，如 2007 年 4、8、10 月及 2008 年 2 月，但整體的趨勢皆涵蓋在 95% 的信賴區間內，因此，此種誤差是屬於可以容忍的範圍內。

2. 提供台鐵公司在未來一年內營運的載客需求預測，並提供圖 0-18 的信賴區間，有 95% 的信心可以掌握客運指數的範圍，唯有 5% 的可能會偏離此區間。

3. 可從圖 0-18 的信賴區間圖看出，在 2007 年的 5-6 月間呈現客運指數下降的趨勢，而 7-9 月間有明顯的上升現象，可供台鐵公司作為營運調度的參考。

從上述結果得知，從一般的角度看來，台灣鐵路客運營運不但受到航空客運威脅，還面臨公路客運的搶佔，加上高鐵的完工通車，及都會捷運系統的競爭，導致其營運指數時而成長時而停滯；但從預測分析結果看來，天然災害如颱風或流行病學如 SARS，都可能使台鐵營運陷入低潮，而如何從低潮中走出來，就是台鐵當局應該思考的問題。

### 第二節 建議

台鐵是歷史悠久的交通運輸，也曾經風光一時，面對新式交通建設的挑戰，光是保守舊有經營觀念是無法並駕齊驅的，尤其是近年來事故頻傳，負面新聞太多，服務品質又差，誤點率又高，就算沒有高鐵帶來的影響，台鐵本身就已是千瘡百孔，高鐵的出現，只是加速問題的呈現而已。

台鐵應從本身的問題先解決，否則內憂外患下，勢必不能達到理想效果。政府應監督並促使台鐵進行改革，政策方面除了可修法開放鐵路局經營腹地場站開發外，並設法解決台鐵長期的退休金問題，加速公司化的腳步，定期招考新進人員，全面提昇服務品質，做好短途運輸的角色，並對於高鐵無法到達的地區，提供更快速、便捷的運輸。

## 參考文獻

O'Connell • Koehler Thomson Learning Forecasting, Time Series, And Regression

陳婉淑 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](http://cache.moe.edu.tw/aremos_ly/search.html)

交通部台灣鐵路管理局：<http://www.railway.gov.tw/index/index.aspx>

