

逢甲大學學生報告 ePaper

澳門輕汽油月消費量之分析與預測

Analysis and Forecast of Macao Gasoline monthly Assumption

作者：陳敏駿、 陳婉嫻

系級：統計學系三乙

學號：D9765291、D0092701

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： 100 學年度 第 2 學期

中文摘要

能源危機已經成為全球關注的重要議題之一，不僅僅是因為伊拉克戰爭、利比亞戰爭等使得國際能源價格大幅度波動，而且是因為人們日益認識到，能源的消耗會對環境產生重要影響，特別是二氧化碳的排放。而汽油作為各種能源中使用量最具有代表性的一種，與我們的生活息息相關。澳門作為一個面積很小的地區，是不是更加容易收到國際油價波動的影響呢？他們的各種節能減排活動是否成效顯著呢？作為一個和臺灣相隔不遠、關係密切的地區，他們的能源使用情況能給我們重要的參考。

我們使用澳門特別行政區政府統計暨普查局的“能源耗用量——汽油耗用量”資料，時間為西元 2001 年 1 月至 2011 年 12 月共 132 筆月資料，並保留最後 12 筆資料用於樣本外預測。我們使用了四種時間序列分析的方法：ARIMA 季節模型、指數平滑法、時間序列迴歸法、分解法來配適，並利用 MAD、MSE、MPE、MAPE 四種準則來評估樣本外預測，以確定何種方法為最佳。結果發現在這四個準則之下，ARIMA Seasonal Model 為最好的方法，因此我們借由 ARIMA Seasonal Model 的分析結果得出了若干重要結果和政策建議。

關鍵字：能源、汽油、澳門、時間序列

Abstract

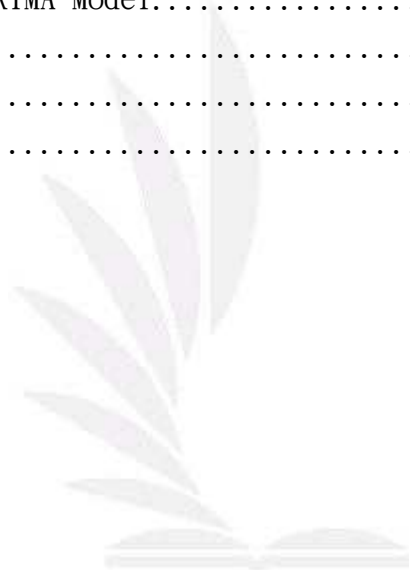
Energy crisis is becoming one of the most important issue of the world, not only for the fluctuation of energy price which result from Iraq war and Libya war, but also caused by people's concept that the assumption of fuel have a great influence on the environment, especially the discharge of carbon dioxide. As the representative of fuel, gasoline has a strong impact on our life. Considering Macao is a small place, will they be more easily influenced by the international energy price? Are their energy saving methods effective? There usage of fuel is significant conference to us.

We used the data of "Energy consumption-Gasoline consumption" from Macao's statistical bureau. The time period is from January 2011 to December 2011, total 132 observations. We save the last 12 observations for out-of-sample forecasts. We used 4 kind of time series analysis method, including ARIMA seasonal Model, exponential smoothing, time series regression and decomposition method. Besides, in order to know which model is the best one, we used the criteria of MAD、MSE、MPE、MAPE to evaluate the out-of-sample forecasts.

Keyword : Energy, Gasoline, Macao, Time series

目 次

第一章 緒論.....	6
第二章 研究方法.....	7
第二章 研究方法.....	7
第一節、資料分析方法.....	8
第二節、診斷分析方法.....	11
第三節、預測評估準則.....	12
第三章 資料分析.....	14
第一節、時間序列迴歸法.....	14
第二節、分解法.....	19
第三節、指數平滑法.....	25
第四節、Seasonal ARIMA Model.....	28
第五節、最佳模型.....	36
第四章 結論與建議.....	37
參考文獻.....	39



圖目錄

第一章 緒論.....	6
第二章 研究方法.....	7
第二章 研究方法.....	7
圖 2.1-1：研究流程圖	7
第三章 資料分析.....	14
圖 3.1-1：汽油耗用量原始時間序列圖	14
圖 3.1-2：殘差圖	18
圖 3.1-3：預測區間曲線圖	18
圖 3.2-1：原始時間序列圖	19
圖 3.2-2：去除季節因數時間序列圖	20
圖 3.2-3：趨勢循環時間序列圖	20
圖 3.2-4：季節因數時間序列圖	21
圖 3.2-5：不規則因數時間序列圖	錯誤！尚未定義書籤。 1
圖 3.2-6：預測區間曲線圖	24
圖 3.3-1：汽油耗用量原始時間序列圖	25
圖 3.3-4：預測圖	27
圖 3.4-1：汽油耗用量原始時間序列圖	28
圖 3.4-2：原始資料的 ACF 及 PACF.....	29
圖 3.4-3：一次差分後的 ACF 及 PACF.....	30
圖 3.4-4：季節差分後的 ACF 及 PACF.....	31
圖 3.4-5：殘差的 ACF 及 PACF.....	32
圖 3.4-6：White Noise Test.....	33
圖 3.4-7 預測值，預測區間及真實值曲線圖	34
第四章 結論與建議.....	37
參考文獻.....	39

表目錄

第一章 緒論.....	6
第二章 研究方法.....	7
第二章 研究方法.....	7
表 2.3-1 MAPE 的評估準則.....	13
第三章 資料分析.....	14
表 3.1-1: 誤差項的自我相關檢測表.....	15
表 3.1-2: 誤差項的自我相關檢測表.....	16
表 3.1-3: 參數估計.....	16
表 3.1-4: 實際值、預測值、95%CI 值.....	17
表 3.1-5: MSE、MAE、MPE、MAPE 值.....	18
表 3.2-1: 簡單迴歸分析.....	22
表 3.2-2: 簡單迴歸分析和一階自我相關項.....	22
表 3.2-3: 實際值、預測值、95%CI 值.....	24
表 3.2-4: MSE、MAE、MPE、MAPE 值.....	24
表 3.3-1: 參數估計.....	26
表 3.3-2 預測值及其調整.....	26
表 3.3-3 MSE、MAE、MPE、MAPE 值.....	27
表 3.4-1 Ljung-Box 檢定.....	33
表 3.4-2 預測值，預測區間及真實值.....	34
表 3.4-3 MSE、MAE、MPE、MAPE 值.....	35
表 3.5-1 四個模型四種準則的值.....	36
第四章 結論與建議.....	37
參考文獻.....	39

第一章 緒論

現代社會的發展離不開能源的消耗，其中以石油資源為主，因此國際上任何有關石油的風吹草動都能夠產生重大影響。

如今各個國家都或多或少地受到能源危機的影響，能源危機已經成為全球關注的重要議題之一，不僅僅是因為伊拉克戰爭、利比亞戰爭等使得國際能源價格大幅度波動，而且是因為人們日益認識到，能源的消耗會對環境產生重要影響，特別是二氧化碳的排放，會加劇溫室效應。如果不對能源的消耗加以限制，可以預料到，本世紀將會發生全球生態環境災難。因此才有了《京都議定書》能旨在限制全球能源使用的全球性協定。

而汽油作為各種能源中使用量最具有代表性的一種，與我們的生活息息相關。雖然新能源的研究有了飛速的發展，但是更多的處於試驗階段，很少能夠大規模應用。因此現在全球幾乎所有的車輛仍然以汽油為驅動能源。汽油的價格、消耗量能夠很好地代表整個能源家族的狀況。

關於石油的利益爭端從來沒有停止。從上世紀九十年代到今天，已經發生了海灣戰爭、科索沃戰爭、伊拉克戰事、利比亞戰爭等以石油爭端為主要目的的戰爭。這些戰爭都對國際油價產生了重大影響。例如伊拉克戰爭期間，臺灣、中國大陸的油價都有顯著上漲。

澳門作為一個面積很小的地區，是不是更加容易收到國際油價波動的影響呢？他們的各種節能減排活動是否成效顯著呢？作為一個和臺灣相隔不遠、關係密切的地區，他們的能源使用情況能給我們重要的參考。根據新聞資料，海灣戰爭、伊拉克戰爭、1999年澳門迴歸中國大陸之後，澳門的汽油消費量都因油價上漲有顯著減少。

我們使用澳門特別行政區政府統計暨普查局的“能源耗用量——汽油耗用量”資料，時間為西元2001年1月至2011年12月共132筆月資料，並保留最後12筆資料用於樣本外預測。我們使用了四種時間序列分析的方法：ARIMA季節模型、指數平滑法、時間序列迴歸法、分解法來配適，並利用MAD、MSE、MPE、MAPE四種準則來評估樣本外預測，以確定何種方法為最佳。

第二章 研究方法

本報告的研究流程圖如圖 2.1-1。

首先決定研究課題。能源危機是全球關注的重要議題之一，因為國際能源價格大幅度波動，而且能源的消耗會對環境產生重要影響，特別是二氧化碳的排放。而汽油作為各種能源中使用量最具有代表性的一種，與我們的生活息息相關。我們決定對澳門的汽油消費量進行分析和預測，從而給予臺灣一個重要的參考範本。

接下來是找尋相關資料。我們使用澳門特別行政區政府統計暨普查局的“能源耗用量——汽油耗用量”資料，時間為西元 2001 年 1 月至 2011 年 12 月共 132 筆月資料。我們將這 252 筆月資料分為兩個部分，第一部分為前 120 筆資料，即樣本內資料，主要用來配適模型。第二部分為後 12 筆資料，即樣本外資料，用於樣本外預測。

然後我們使用了四種時間序列分析的方法：ARIMA 季節模型、指數平滑法、時間序列迴歸法、分解法來配適，並對 2011 年的資料進行預測。最後利用 MAD、MSE、MPE、MAPE 四種準則來評估樣本外預測，以確定何種方法為最佳。



圖 2.1-1：研究流程圖

第一節、資料分析方法

時間序列是指把過去的觀察資料依時間順序排列而成的資料序列，時間序列分析即是應用數理統計方法對資料進行統計分析及推斷。

1. 時間序列迴歸法(Time Series Regression)：

它的基本原理：

- 假設事物發展具有連續性，從過去的資料中就能預測事物的發展趨勢。
- 時間序列迴歸分析法是僅依照變數本身過去的資料所存有的變異型態來建立模型。時間序列迴歸法是將時間序列以趨勢(Trend, TR_t)、季節(Season, SN_t)、殘差(ϵ_t)來表示。

時間序列模型： $y_t = TR_t + SN_t + \epsilon_t$

模型配適好之後我們利用診斷分析方法檢查此模型有無違反誤差項之假設： $\epsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$ 。

2. 分解法 (Decomposition Method)：

時間序列分解法是數年來一直非常有用的方法，這種方法包括譜分析、時間序列分析和傅立葉級數分析等。

它的基本原理：

- 分解法是將時間序列分解成四大因素：趨勢(TR_t)、季節(SN_t)、迴圈(CL_t)、不規則(IR_t)。
- 分解法又可分為兩種模型，若時間序列的變異數不會隨著時間的改變而改變為常數時，我們採用加法模型；若反之，時間序列的變異數不平穩時，我們則使用乘法模型。
- 再來我們利用 x11 的方法來分析我們所挑選的原始資料。

時間序列模型：
$$\begin{cases} \text{加法模型：} y_t = TR_t + SN_t + CL_t + IR_t \\ \text{乘法模型：} y_t = TR_t \times SN_t \times CL_t \times IR_t \end{cases}$$

模型配適好之後我們利用診斷分析方法檢查此模型有無違反誤差項之假設： $\epsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$ 。

3. 指數平滑法 (Exponential Smoothing) :

假設時間序列的態勢具有穩定性或規則性，所以時間序列可以被合理的順勢推延。而指數平滑法將全期平均和移動平均的優點保留，不捨棄過去的資料，但是給予逐漸減弱的影響程度，即隨著資料資料的遠離，賦予逐漸收斂為零的權數。

指數平滑法是生產預測中常用的一種方法。當我們的資料呈現季節性且變異數平穩時，我們使用 Winters Method—Additive 做配適；若有季節性但變異數不平穩時，我們則使用 Winters Method—Multiplicative。而資料若是沒有呈現季節性，但有上升的趨勢且趨勢有減緩的現象時，我們使用 Damped Trend 做配適。

以下為所介紹之三種模型：

- Winters Method—Additive 模型：

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

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

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

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

- Winters Method—Multiplicative 模型：

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

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

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

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

- Damped Trend Exponential Smoothing 模型：

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

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

4. ARIMA 模型分析法(Autoregressive Integrated Moving Average Model，又稱 Box-Jenkins 模型)：

它的基本原理是：

- ARIMA 模型建模的基本條件是要求此時間序列滿足平穩的條件，即觀察值要圍繞時間序列均值上下波動，不能有明顯的上升或下降趨勢，如果出現上升或下降趨勢，需要對原始序列進行差分平穩化。
- 在做 ARIMA 模型分析法時，我們可分為四個步驟：
 - (a). 辨識模型：利用資料找出可適用的 ARIMA 模型。
 - (b). 估計：將資料放入模型中找出模型中的參數。
 - (c). 診斷分析：用診斷分析法瞭解此模型是否為合適的。
 - (d). 預測：找出最終模型後，即可進行時間序列預測。
- ARIMA 模型可分為 3 種：
 - (a). 自迴歸模型(簡稱 AR 模型)。
 - (b). 移動平均模型(簡稱 MA 模型)。
 - (c). 自迴歸移動平均混合模型(簡稱 ARIMA 模型)。

以下為時間序列模型：

$$\left\{ \begin{array}{l} \text{AR}(P) : Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \\ \text{MA}(q) : Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \\ \text{ARMA}(p, q) : Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \end{array} \right.$$

以上三種模型需假設白噪音來自常態分配，也就是平均數為零，變異數為常數，因此配適模式後，需要檢查殘差是否存在自我相關，當殘差彼此無自我相關，即代表配適模型是一個合適的模型。

第二節、診斷分析方法

1. 白噪音檢定(white noise test) :

噪音檢定主要用以檢測時間數列模型當中的殘差之間是否存在相關性，當殘差符合白噪音，即殘差項為零，變異數為一常數，且殘差間彼此無自我相關者，即符合白噪音當殘差符合白噪音代表模型唯一適合模型。

- 假設檢定：

H_0 ：殘差符合白噪音

H_1 ：殘差不符合白噪音

- 決策規則：

$p\text{-value} > 0.05$ ，則不拒絕 H_0 ，符合白噪音，表模型配適合適。

$p\text{-value} < 0.05$ ，則拒絕 H_0 ，不符合白噪音，表模型配適不合適。

2. 單根檢定(Unit Root Test) :

單根檢定乃是檢測時間序列是否為平穩，平穩即代表平均數不隨時間改變而改變，亦表示殘差的期望值等於零。在 ARIMA 的分析過程中，序列是否平穩是首要工作，當序列平穩後才可再做進一步的分析，因此，單根檢定的目的在於決定要以原始之資料或差分調整後之資料進行分析研究。

- 假設檢定：

H_0 ：有單根（時間序列不平穩）

H_1 ：沒有單根（時間序列平穩）

- 決策規則：

$p\text{-value} > 0.05$ ，則不拒絕 H_0 ，表示有單根，模型配適不合適。

$p\text{-value} < 0.05$ ，則拒絕 H_0 ，表示沒有單根，模型配適合適。

3. DW 檢定法：

主要用以檢定迴歸分析中的殘差是否存在自我相關，在一個合適的模型中，我們假設殘差彼此無相關，表示一個合適的模型中，殘差彼此的相關性為零。

- 檢定規則：

當 $Pr < DW$ 的值小於顯著水準 $\alpha = 0.05$ 時，表示顯著存在正自我相關。

當 $Pr > DW$ 的值小於顯著水準 $\alpha = 0.05$ 時，表示顯著存在負自我相關。

第三節、預測評估準則

預測能力的好壞，也就是預測精確度，可以用指標來做衡量，我們將時間序列資料分為兩部分，第一部分為前 120 筆資料，即樣本內資料，主要用來分析與預測未來的走勢；第二部分為最後的 12 筆資料，即樣本外資料，主要與所預測出的結果做對照以瞭解模型的預測程度是否在一定的水準上。以下是我們用來衡量不同預測方法的評估準則。假設 n 為總樣本數， y_t 表示第 t 個時間點的實際值， \hat{y}_t 表示第 t 個時間點的預測值，其評估準則如下：

1. 平均絕對離差 (Mean Absolute Deviation, MAD)：
用以評估評價模型與資料之誤差程度，越接近零表其誤差程度越小。

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

2. 均方差 (Mean Squared Error, MSE)：
計算實際值與預測值之預測誤差取平方加總後並加以平均，通常數字太大則會將之取根號，值越大表示預測越不精準。

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

3. 平均誤差率 (Mean Percentage Error, MPE)：
將誤差除以實際值，得到誤差率並加以平均，值越小代表誤差率越小。

$$MPE = \left[\frac{1}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t} \right] \times 100\%$$

4. 平均絕對誤差百分比 (Mean Absolute Percentage Error, MAPE)：
將誤差絕對值除以實際值，得到誤差率絕對值並加以平均，用以作為預測模式好壞之評估指標。

$$MAPE = \left[\frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \right] \times 100\%$$

MAPE	評估
<5	高準確的預測
6 - 10	尚可的預測
>10	不準確的預測

表 2.3-1 MAPE 的評估準則



第三章 資料分析

第一節、時間序列迴歸法

A. 時間序列圖

首先，我們對原始數據進行月份調整，即各個月份的數據乘以係數 α ， $\alpha=30/\text{當月天數}$ 。由圖 3.1-1 為澳洲汽油月消費量原始時間序列圖，坐標 X 軸為日期，坐標 Y 軸為澳洲汽油月消費量，以千公升為單位，由圖可以發現，汽油耗用量隨著時間有逐年增加的趨勢，且有明顯的季節變化，這與人口增加有關，主要原因是移名大量湧入，造成車子需求量增加，汽油消費量也就增加。因為季節變異波動沒有明顯變大趨勢，所以不需要進行差分平穩化。

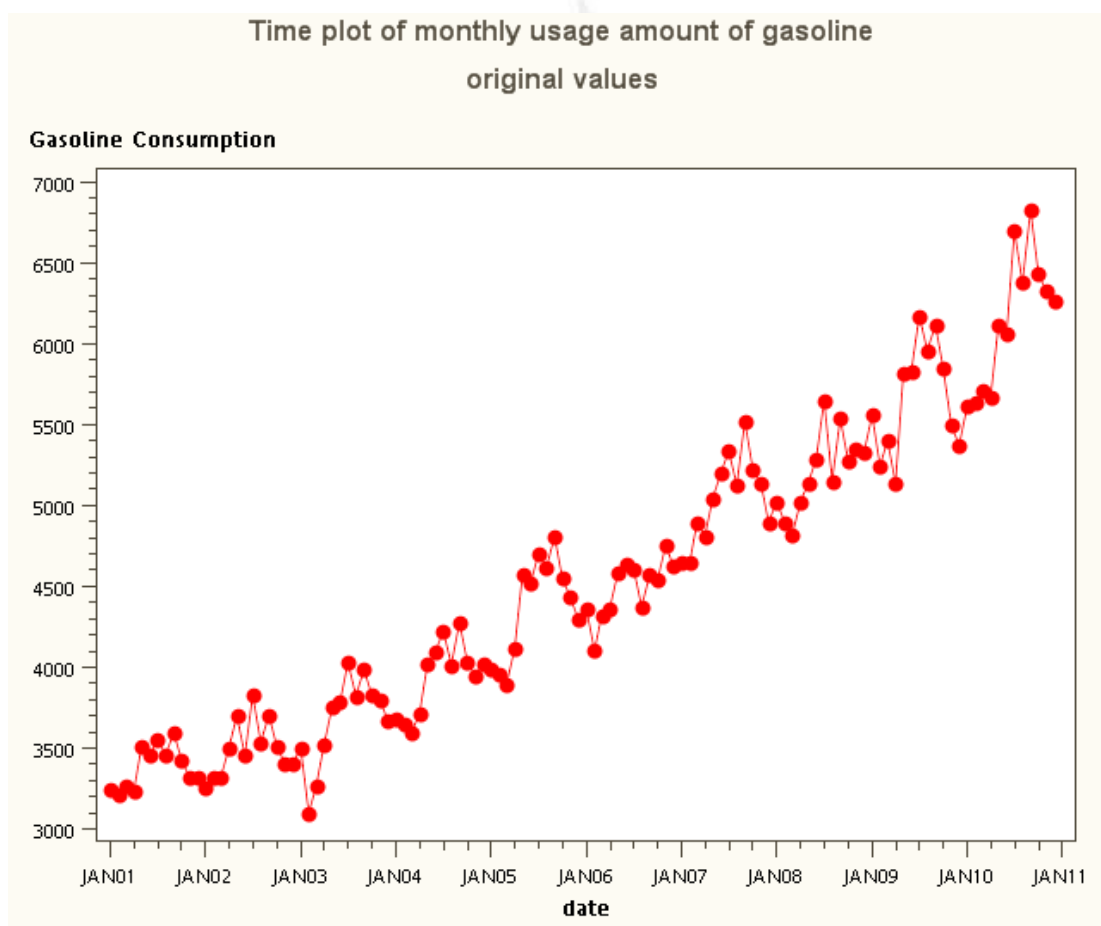


圖 3.1-1：汽油耗用量原始時間序列圖

B. 配適迴歸模型

假設模型如下： $y_t = TR_t + SN_t + \varepsilon_t$; y_t : 在 t 時間下的實際觀察值

TR_t : 在 t 時間下的趨勢因子

SN_t : 在 t 時間下的季節因子

ε_t : 在 t 時間下的誤差項，且 $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$

Ordinary Least Squares Estimates			
SSE	4099205.34	DFE	107
MSE	38310	Root MSE	195.73025
SBC	1655.44007	AIC	1619.20268
MAE	143.867153	AICC	1622.63664
MAPE	3.29892155	HQC	1633.91895
Durbin-Watson	0.5281	Regress R-Square	0.9610
		Total R-Square	0.9610
Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	0.5281	<.0001	1.0000

表 3.1-1: 誤差項的自我相關檢測表

- 由表 3.1-1 誤差項的自我相關檢測，得知此迴歸模型 R-Square 值為 0.9610，也就是汽油耗用量的總變異有 96.1% 可被此模型所解釋。但這不代表此模型已經是適當的了，還要進一步檢定誤差項是否有自我相關，若有自我相關，則違反一開始的迴歸假設，於是要重新建立模型。
- Durbin-Waston 檢定:
 當 $Pr < DW$ ，其值小於顯著水準 $\alpha = 0.05$ 時，誤差項間有正自我相關。
 當 $Pr > DW$ ，其值小於顯著水準 $\alpha = 0.05$ 時，誤差項間有負自我相關。
 由表一可得知 DW 值為 0.5281， $Pr < DW$ 的值為 < 0.0001 ，所以得知此模型誤差項間有正自我相關，須重新建立模型，為更正其模型，將誤差項配適為 AR(1) 模型，也就是一階自我相關模型。

C. 最終模型

假設模型如下： $y_t = TR_t + SN_t + \varepsilon_t$ ；其中 $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \alpha_t$

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.705875	0.068800	-10.26
Yule-Walker Estimates			
SSE	1930466.99	DFE	106
MSE	18212	Root MSE	134.95167
SBC	1570.5535	AIC	1531.52861
MAE	97.793657	AICC	1535.52861
MAPE	2.20118009	HQC	1547.3768
Durbin-Watson	2.2238	Regress R-Square	0.8636
		Total R-Square	0.9816
Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	2.2238	0.8771	0.1229

表 3.1-2: 誤差項的自我相關檢測表

- 由表 3.1-2 重新配適模型後誤差項的自我相關檢測，得知此迴歸模型 R-Square 值下降為 0.8636，也就是汽油耗用量的總變異有 86.36% 可被此模型所解釋；但其 DW 值變更為 2.2238，Pr < DW 的值為 0.8771，Pr > DW 的值為 0.1229，得知此模型誤差間沒有自我相關，符合一開始的迴歸假設。

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	2837	92.6156	30.64	<.0001
t	1	25.8301	1.1435	22.59	<.0001
d1	1	53.7106	47.8751	1.12	0.2644
d2	1	-60.8469	60.7888	-1.00	0.3191
d3	1	-48.5684	67.9777	-0.71	0.4785
d4	1	-16.4207	72.1087	-0.23	0.8203
d5	1	271.4425	74.2396	3.66	0.0004
d6	1	251.6018	74.8311	3.36	0.0011
d7	1	474.3758	74.0092	6.41	<.0001
d8	1	212.0013	71.6240	2.96	0.0038
d9	1	439.6430	67.1832	6.54	<.0001
d10	1	189.2283	59.5695	3.18	0.0020
d11	1	97.3040	45.9156	2.12	0.0364

表 3.1-3: 參數估計

- 由表 3.1-3 參數估計中，可以發現 d1、d2、d3、d4 的 P-value 大於 0.05 為不顯著，照理說應將之刪除，但由於在季節因數裡面的虛擬變數是月分，所以所有的月份必須同進同出，所以將之留在模型中，不予以刪除。

- 將參數配入模型，可得預測模型如下：

$$y_t = 2837 + 25.8301t + 53.7106M_1 - 60.8469M_2 - 48.5684M_3 - 16.4207M_4 + 271.4425M_5 + 251.6018M_6 + 474.3758M_7 + 212.0013M_8 + 439.6430M_9 + 189.2288M_{10} + 97.3040M_{11} + \varepsilon_t$$

$$\varepsilon_t = 0.705875 \varepsilon_{t-1} + \alpha_t$$

$$\begin{matrix} \text{iid} \\ \alpha_t \sim N(0, \sigma^2) \end{matrix}$$

$$M_1 \begin{cases} 1 & \text{when Jan.} \\ 0 & \text{Otherwise} \end{cases} \text{ 以此類推}$$

D. 評估模型

為了瞭解此模型的預測好壞，在一開始我們保留了 12 筆資料的真實值，將之與預測資料做比較。在表 3.1-4 中，我們對預測值及預測區間進行了月份調整，即除以係數 α ， $\alpha = 30 / \text{當月天數}$ 。計算出 95% 信賴水準的上下區間，並將之畫成圖形，詳細資料請看表 3.1-4，圖 3.1-2 為殘差圖，圖 3.1-3 為預測區間曲線圖。

日期	觀察值	預測值	L95%	U95%
JAN11	6476	6457.17	6162.39	6751.95
FEB11	5763	5685.70	5294.59	6076.80
MAR11	6393	6284.42	5887.68	6681.16
APR11	6318	6105.63	5677.13	6534.12
MAY11	6953	6608.47	6185.24	7031.70
JUN11	6864	6384.32	5955.91	6812.72
JUL11	7566	6841.64	6409.46	7273.82
AUG11	7354	6588.48	6146.81	7030.15
SEP11	7078	6623.46	6201.50	7045.41
OCT11	7210	6607.81	6166.43	7049.20
NOV11	6800	6325.59	5901.36	6749.82
DEC11	6907	6460.42	6018.43	6902.42

表 3.1-4：實際值、預測值、95%CI 值

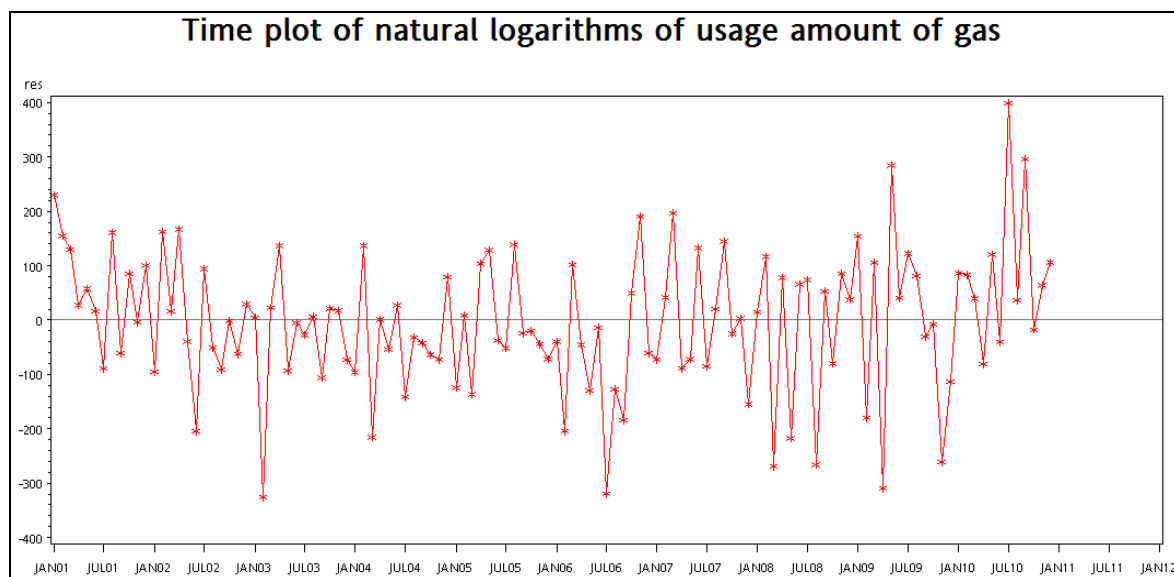


圖 3.1-2：殘差圖

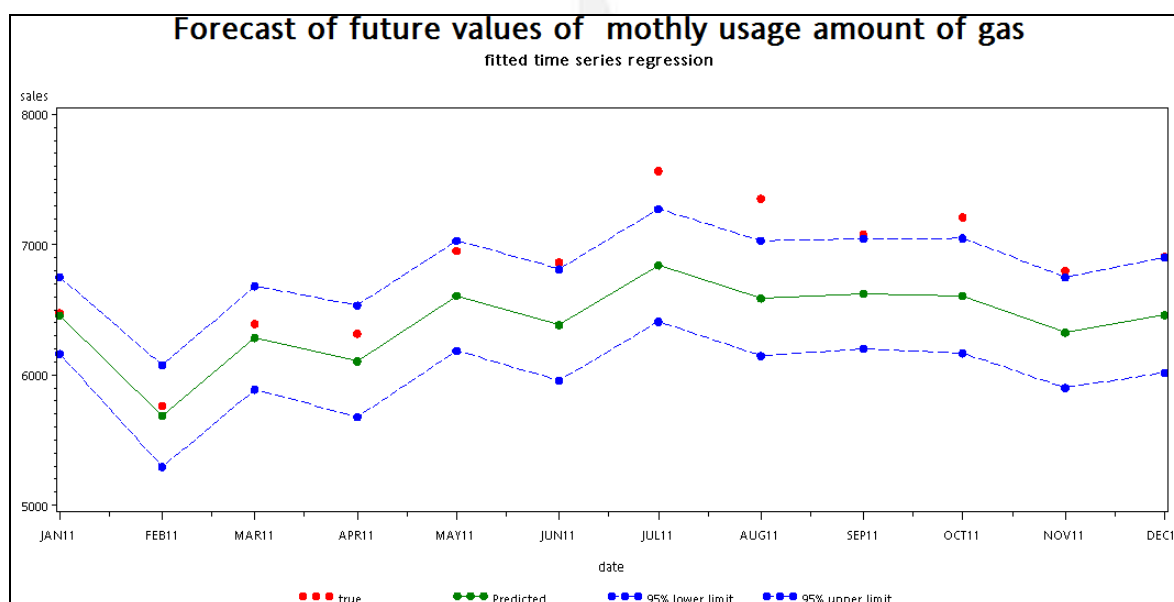


圖 3.1-3：預測區間曲線圖

由圖 3.1-3 可看出在 2011 年的六月開始，預測值都超出 95% 的信賴區間，且整體預測直接大於實際值非常多，可見，此時間序列迴歸法的預測能力上有不足的地方。為了瞭解此分析法的好壞，於是算出此模型之 MSE、MAE、MPE、MAPE，詳細資料請看表 3.1-5：

MSE	MAE	MPE	MAPE
209705.71	392.409	5.56960	5.56960

表 3.1-5：MSE、MAE、MPE、MAPE 值

第二節、分解法

分解法有兩種模型，情況如下：

- 當變異數為常數時，為加法模型： $y_t = TR_t + SN_t + CL_t + IR_t$
- 當變異數不為常數時，為乘法模型： $y_t = TR_t \times SN_t \times CL_t \times IR_t$

y_t : 在 t 時間下的實際觀察值

TR_t : 在 t 時間下的趨勢因數

SN_t : 在 t 時間下的季節因數

CL_t : 在 t 時間下的迴圈因數

IR_t : 在 t 時間下的不規則因數

A. 時間序列圖

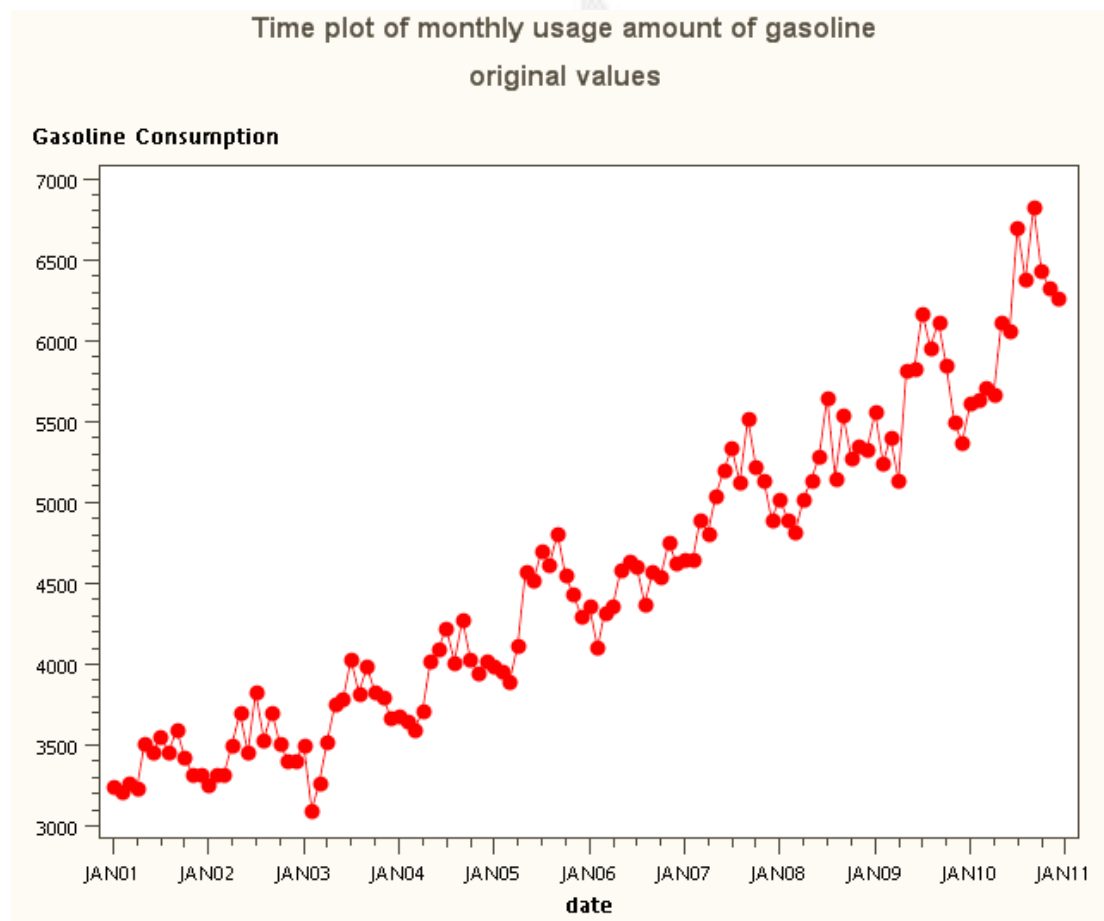


圖 3.2-1：原始時間序列圖

首先，我們對原始數據進行月份調整，即各個月份的數據乘以係數 α ， $\alpha = 30 /$ 當月天數。由圖 3.2-1 可得知，此資料變異數為常數，於是選擇用加法模型。

B. X11 分解法

採用美國統計局X11分解法對變數時間序列之四個特性：趨勢、季節、循環與不規則進行分析。

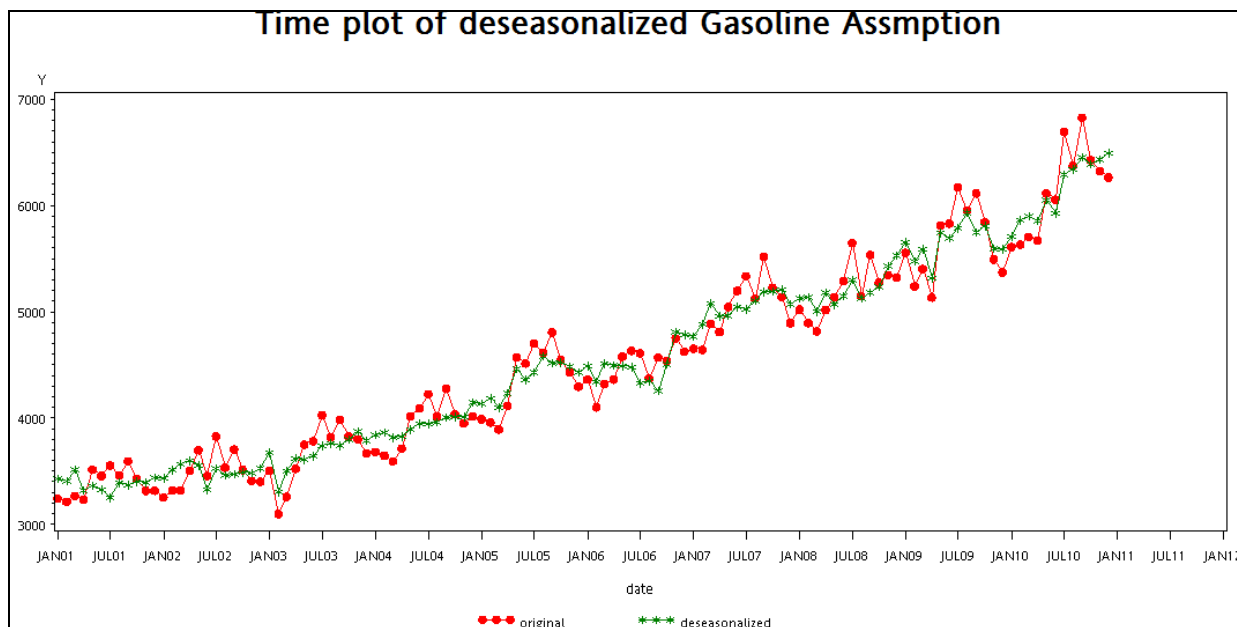


圖 3.2-2：去除季節因數時間序列圖

由圖 3.2-2 可看出在去除掉季節因數後，每個月的汽油耗用量隨著時間而上升，可得知即使沒有季節因數的影響，澳門汽油消費量也會隨著時間增加的。

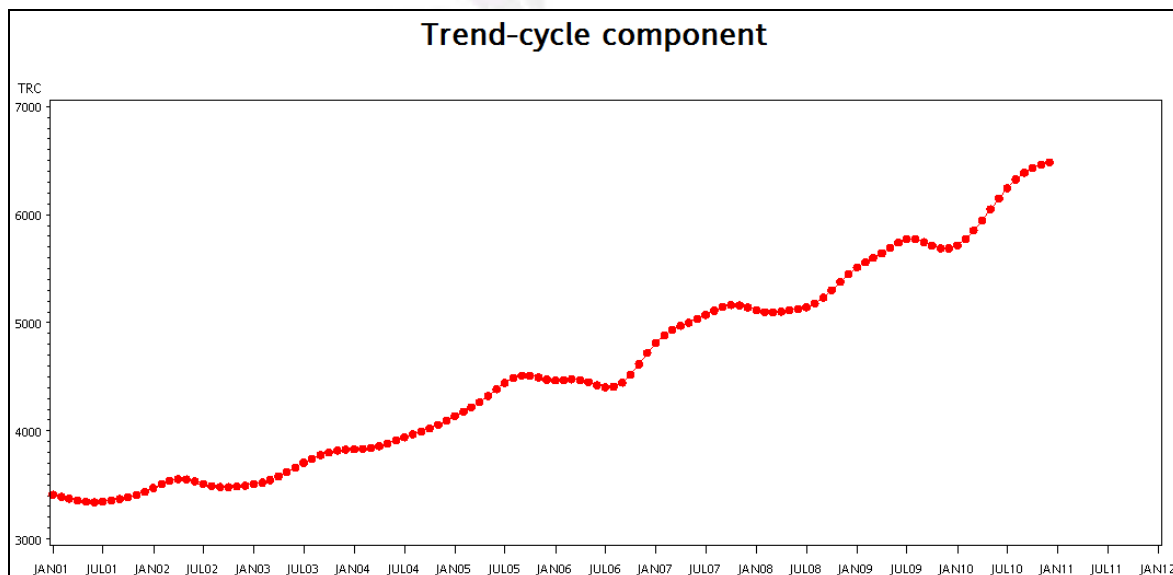


圖 3.2-3：趨勢循環時間序列圖

由圖 3.2-3 可看出在趨勢循環時間序列圖裡，有明顯具有往上的趨勢，這一點在去季節因數時間序列圖可以得到同樣結論。

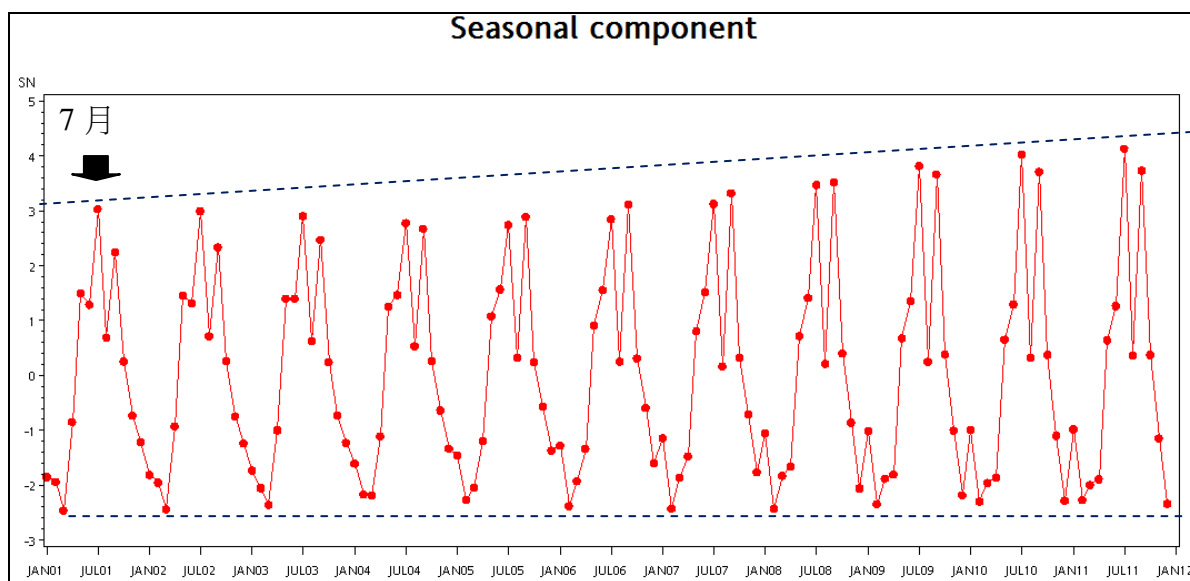


圖 3.2-4：季節因數時間序列圖

由圖 3.2-4 可看出季節因數時間序列圖裡，有明顯的季節變化，尤其是在每年七、九月時澳門汽油消費量最高；其季節變異有越來越大的趨勢，也就是澳門汽油在每年消費量越高的月份與消費量越低的月份，之間差異越來越大。

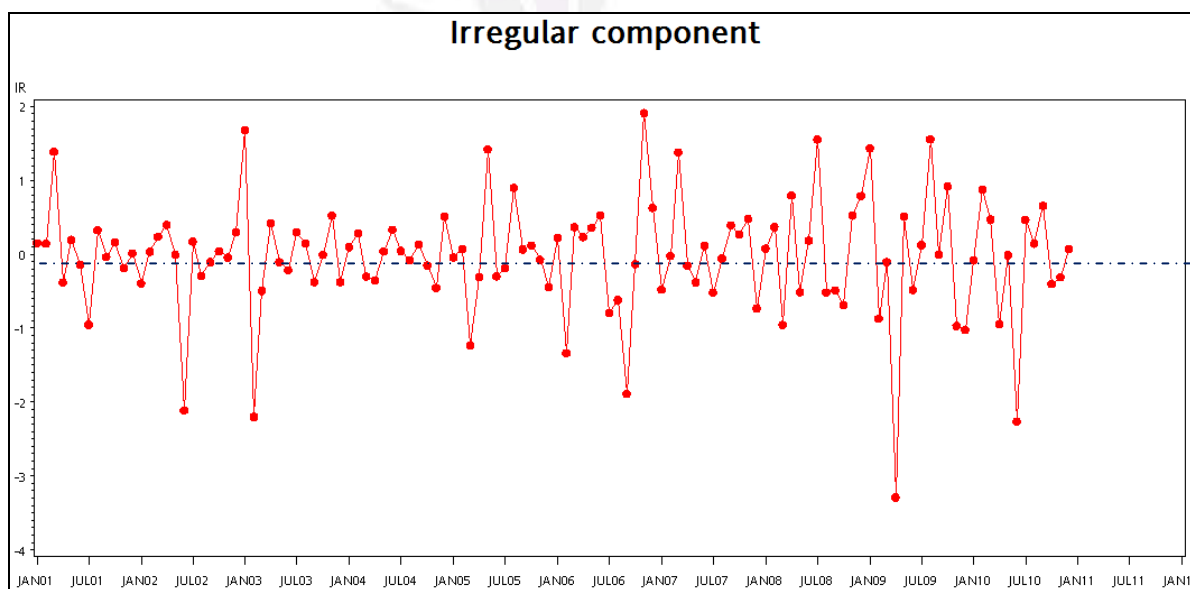


圖 3.2-5：不規則因數時間序列圖

由圖 3.2-5 可看出，在不規則因數時間序列圖裡，有明顯的平均數平穩。

C. 簡單迴歸分析

Ordinary Least Squares Estimates			
SSE	3942036.11	DFE	118
MSE	33407	Root MSE	182.77605
SBC	1598.08617	AIC	1592.51119
MAE	143.127525	AICC	1592.61375
MAPE	3.28651219	HQC	1594.77522
Durbin-Watson	0.4779	Regress R-Square	0.9606
		Total R-Square	0.9606

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	0.4779	<.0001	1.0000

表 3.2-1：簡單迴歸分析

接著我們特別對趨勢因數的時間序列資料做簡單迴歸分析如表 3.2-1，發現 R-Square 值為 0.9606，也就是此迴歸模型的解釋能力為 96.06%，但其 DW 值為 0.4779，Pr<DW 值<0.0001，由 DW 檢定法可知此迴歸模型殘差間有正自我相關，違反了迴歸分析的假設，須另外配適新模型，於是我們將殘差配適一階自我相關模型，結果如下圖。

Estimates of Autoregressive Parameters			
Lag	Coefficient	Standard Error	t Value
1	-0.715799	0.064558	-11.09

Yule-Walker Estimates			
SSE	1736448.62	DFE	117
MSE	14841	Root MSE	121.82545
SBC	1505.20922	AIC	1496.84674
MAE	91.4464119	AICC	1497.05364
MAPE	2.03733105	HQC	1500.24278
Durbin-Watson	2.1912	Regress R-Square	0.8348
		Total R-Square	0.9826

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	2.1912	0.8311	0.1689

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	2997	74.8252	40.06	<.0001
t	1	25.8279	1.0621	24.32	<.0001

表 3.2-2：簡單迴歸分析和一階自我相關項

由表 3.2-2 重新配適模型後誤差項的自我相關檢測，得知此迴歸模型 R-Square 值下降為 0.8348，也就是汽油耗用量的總變異有 83.48%可被此模型所解釋；但其 DW 值變更為 2.1912， $Pr < DW$ 的值為 0.8311， $Pr > DW$ 的值為 0.1689，得知此模型誤差間沒有自我相關，符合一開始的迴歸假設。

再由參數估計可得知此模型如下：

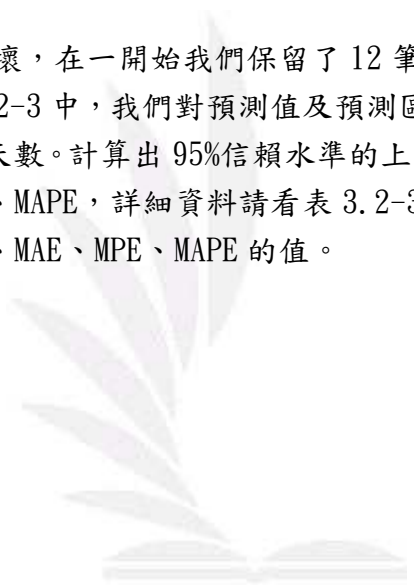
$$dy_t = 2997 + 25.8279t + \varepsilon_t$$

$$\varepsilon_t = 0.715799 \varepsilon_{t-1} + \alpha_t$$

$$\alpha_t \sim N(0, \sigma^2)$$

預測值模型則為： $y_t = TR_t + SN_t$

為了瞭解此模型的預測好壞，在一開始我們保留了 12 筆資料的真實值，將之與預測資料做比較。在表 3.2-3 中，我們對預測值及預測區間進行了月份調整，即除以係數 α ， $\alpha = 30 / \text{當月天數}$ 。計算出 95%信賴水準的上下區間，將之畫成圖形，也計算出 MSE、MAE、MPE、MAPE，詳細資料請看表 3.2-3，圖 3.2-6 為預測區間曲線圖，表 3.2-4 為 MSE、MAE、MPE、MAPE 的值。



D. 評估模型

日期	觀察值	預測值	L95%	U95%
JAN11	6476	6624.11	6370.79	6877.42
FEB11	5763	5929.26	5581.71	6276.80
MAR11	6393	6530.85	6183.56	6878.13
APR11	6318	6303.92	5919.17	6688.67
MAY11	6953	6512.23	6137.46	6887.01
JUN11	6864	6306.91	5921.50	6692.31
JUL11	7566	6531.08	6145.39	6916.77
AUG11	7354	6542.61	6144.97	6940.24
SEP11	7078	6352.60	5974.03	6731.17
OCT11	7210	6581.72	6184.07	6979.37
NOV11	6800	6389.66	6007.77	6771.54
DEC11	6907	6625.08	6226.72	7023.44

表 3.2-3：實際值、預測值、95%CI 值

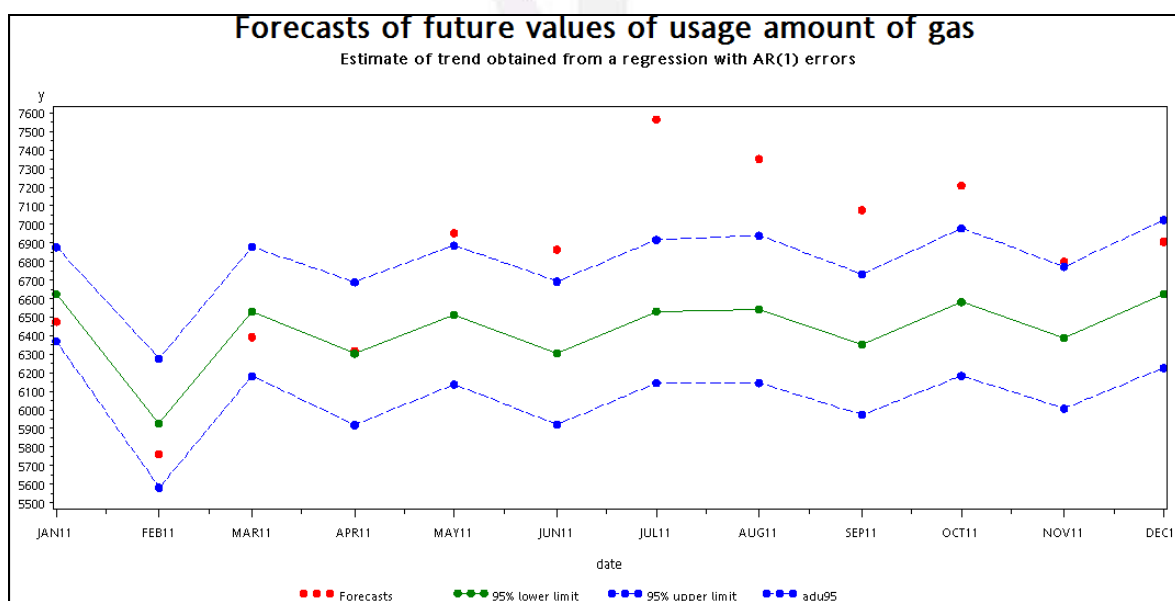


圖 3.2-6: 預測區間曲線圖

由圖 3.2-6 可以看到，在 2011 年的六月開始，預測值都超出 95%的信賴區間，且整體預測直接大於實際值非常多，可見，此時間序列迴歸法的預測能力上有不足的地方。為了瞭解此分析法的好壞，於是算出此模型的 MSE、MAE、MPE、MAPE，詳細資料請看表 3.2-4：

MSE	MAE	MPE	MAPE
289302.45	446.367	5.09507	6.31643

表 3.2-4：MSE、MAE、MPE、MAPE 值

第三節、指數平滑法

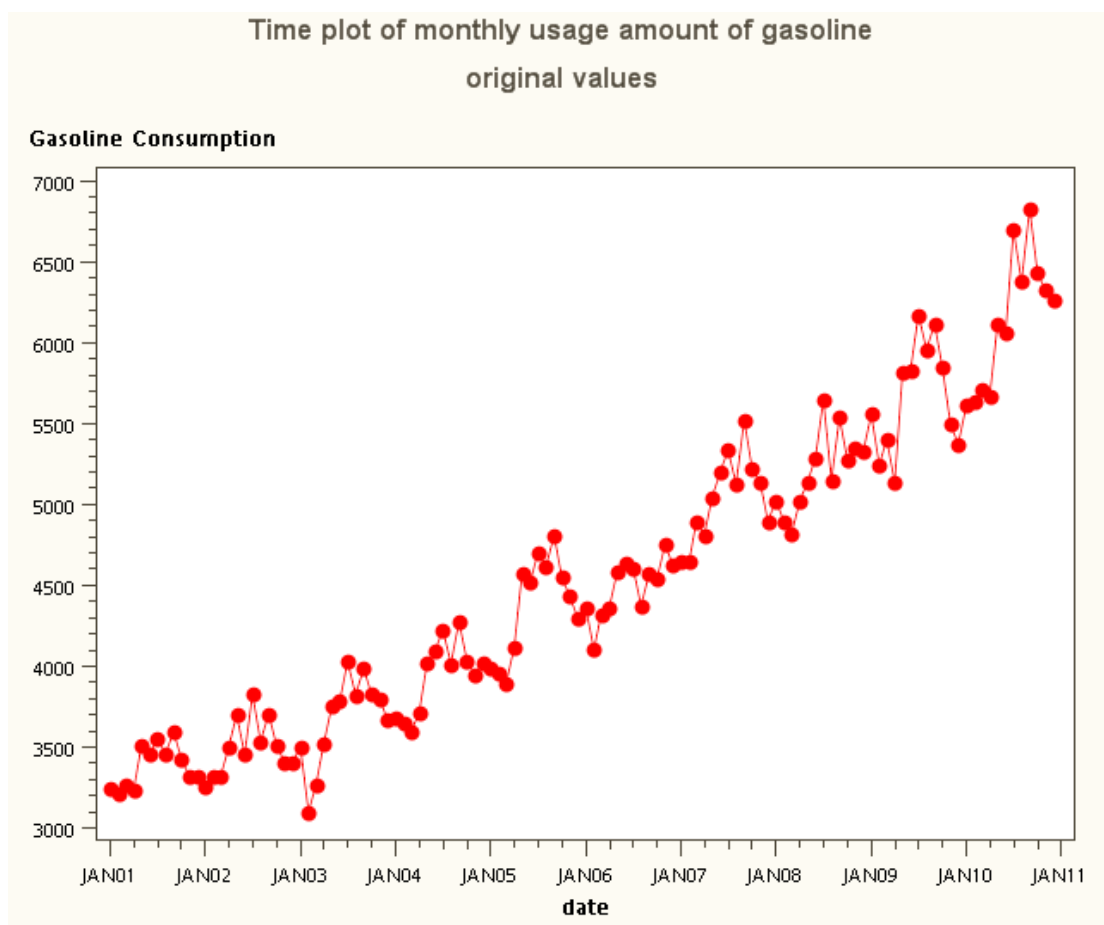


圖 3.3-1：汽油耗用量原始時間序列圖

首先，我們對原始數據進行月份調整，即各個月份的數據乘以係數 α ， $\alpha=30/\text{當月天數}$ 。從圖 3.3-1 我們可以看到，原始時間序列季節變異波動沒有明顯變大趨勢，表示變異為常數，及變異數平穩，因此我們選用 Winters Method-Additive 模型作配適。

Model Parameter	Estimate	Std. Error	T	Prob> T
LEVEL Smoothing Weight	0.62567	0.0611	10.2476	<.0001
TREND Smoothing Weight	0.00100	0.0087	0.1153	0.9084
SEASONAL Smoothing Weight	0.00100	0.0811	0.0123	0.9902
Residual Variance (sigma squared)	14931	.	.	.
Smoothed Level	6431	.	.	.
Smoothed Trend	25.63394	.	.	.
Smoothed Seasonal Factor 1	-110.20044	.	.	.
Smoothed Seasonal Factor 2	-248.51144	.	.	.
Smoothed Seasonal Factor 3	-201.09741	.	.	.
Smoothed Seasonal Factor 4	-165.88901	.	.	.
Smoothed Seasonal Factor 5	123.84214	.	.	.
Smoothed Seasonal Factor 6	104.90176	.	.	.
Smoothed Seasonal Factor 7	327.71578	.	.	.
Smoothed Seasonal Factor 8	64.52435	.	.	.
Smoothed Seasonal Factor 9	290.38943	.	.	.
Smoothed Seasonal Factor 10	37.01961	.	.	.
Smoothed Seasonal Factor 11	-59.39655	.	.	.
Smoothed Seasonal Factor 12	-163.26656	.	.	.

表 3.3-1：參數估計

由表 3.3-1 可獲得各個參數的估計值，可知趨勢項和季節項都不顯著，代表這兩個變數對此模型的影響不大。指數平滑法在統計上並不是一個完善的方法，因此即使趨勢項和季節項都不顯著，我們仍然把它們保留在模型內。然後我們將各個參數估計值帶入，就得到了指數平滑法預測方程式：

$$Level : L_t = 0.62567(Y_t - S_{t-s}) + (1 - 0.62567)(L_{t-1} + b_{t-1})$$

$$Trend : b_t = 0.00100(L_t - L_{t-1}) + (1 - 0.00100)b_{t-1}$$

$$Seasonal : S_t = 0.00100(Y_t - L_t) = (1 - 0.00100)S_{t-s}$$

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

Lt: the level bt:the growth rate st: the seasonal factor of the time series in period t

預測日期	真實值	預測值	調整預測值	U95	L95	調整 U95	調整 L95
Jan-12	6476	6346.75687	6558.31543	6586.24655	6107.26719	6805.7881	6310.843
Feb-12	5763	6234.07981	5818.47449	6516.66185	5951.49777	6082.21773	5554.731
Mar-12	6393	6307.12778	6517.36538	6627.11979	5987.13578	6848.02379	6186.707
Apr-12	6318	6367.97013	6367.97013	6721.49831	6014.44194	6721.49831	6014.442
May-12	6953	6683.33522	6906.11306	7067.54159	6299.12886	7303.12631	6509.1
Jun-12	6864	6690.02878	6690.02878	7102.69328	6277.36428	7102.69328	6277.364
Jul-12	7566	6938.47675	7169.75931	7377.81102	6499.14248	7623.73806	6715.781
Aug-12	7354	6700.91927	6924.28324	7165.44312	6236.39541	7404.29122	6444.275
Sep-12	7078	6952.41829	6952.41829	7440.88063	6463.95594	7440.88063	6463.956
Oct-12	7210	6724.68241	6948.83849	7236.00794	6213.35688	7477.2082	6420.469
Nov-12	6800	6653.9002	6653.9002	7187.15198	6120.64841	7187.15198	6120.648
Dec-12	6907	6575.66413	6794.85293	7130.01646	6021.3118	7367.68368	6222.022

表 3.3-2 預測值及其調整

在表 3.3-2 中，我們對預測值及預測區間進行了月份調整，即除以係數 α ， $\alpha=30/\text{當月天數}$ 。然後將真實值、預測值、調整預測值、上下區間、調整后的上下區間列成表格。

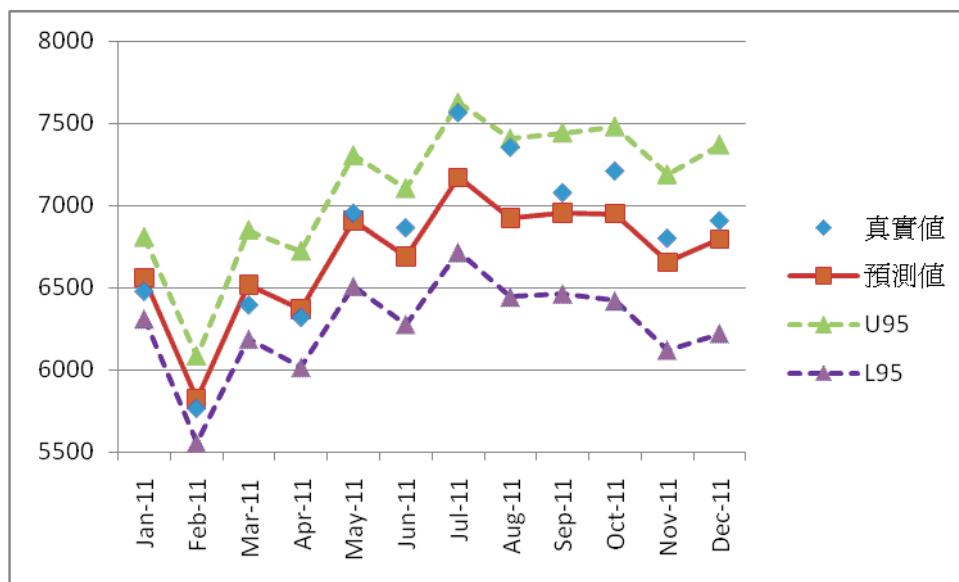


圖 3.3-4：預測圖

在圖 3.3-4 中，所有的預測值皆與真實值相去不遠，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，僅有 1 筆觀測值是落在 95%信賴區間的邊界附近，為 2011 年 8 月。若從整體看來，指數平滑法的預測表現，較時間序列迴歸及分解法的預測表現為佳，表示該模式的預測能力還不錯

MSE	43320.2314
MAE	166.9943
MPE	1.5407
MAPE	2.2641

表 3.3-3 MSE、MAE、MPE、MAPE 值

在表 3.3-3 中，我們可以看到 MAPE 為 2.26，小於 5，根據前述預測評估準則，這是一個非常小的誤差，因此預測效果上佳。

第四節、Seasonal ARIMA Model

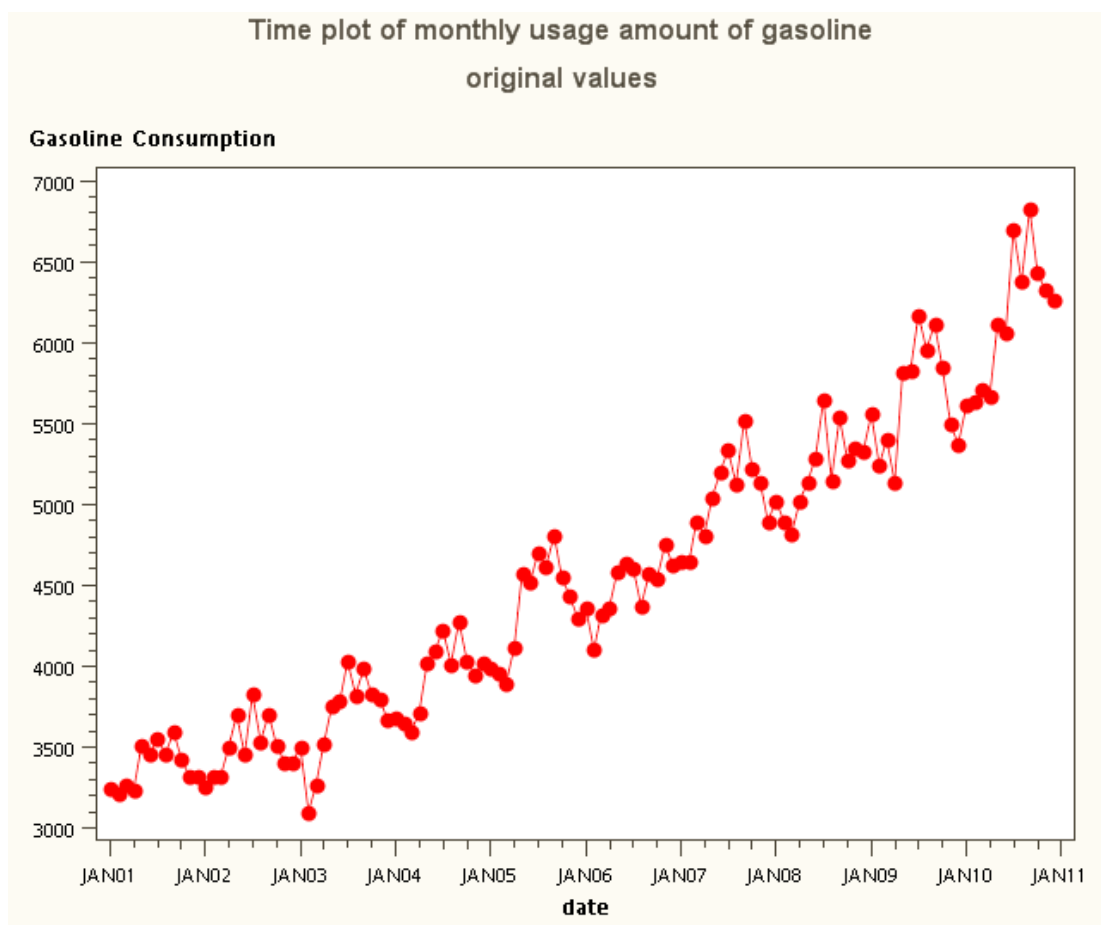


圖 3.4-1：汽油耗用量原始時間序列圖

首先，我們對原始數據進行月份調整，即各個月份的數據乘以係數 α ， $\alpha = 30 /$ 當月天數。從圖 3.4-1 我們可以看到，原始時間序列季節變異波動沒有明顯變大趨勢，表示變異為常數，即變異數平穩，因此無需做資料轉換。

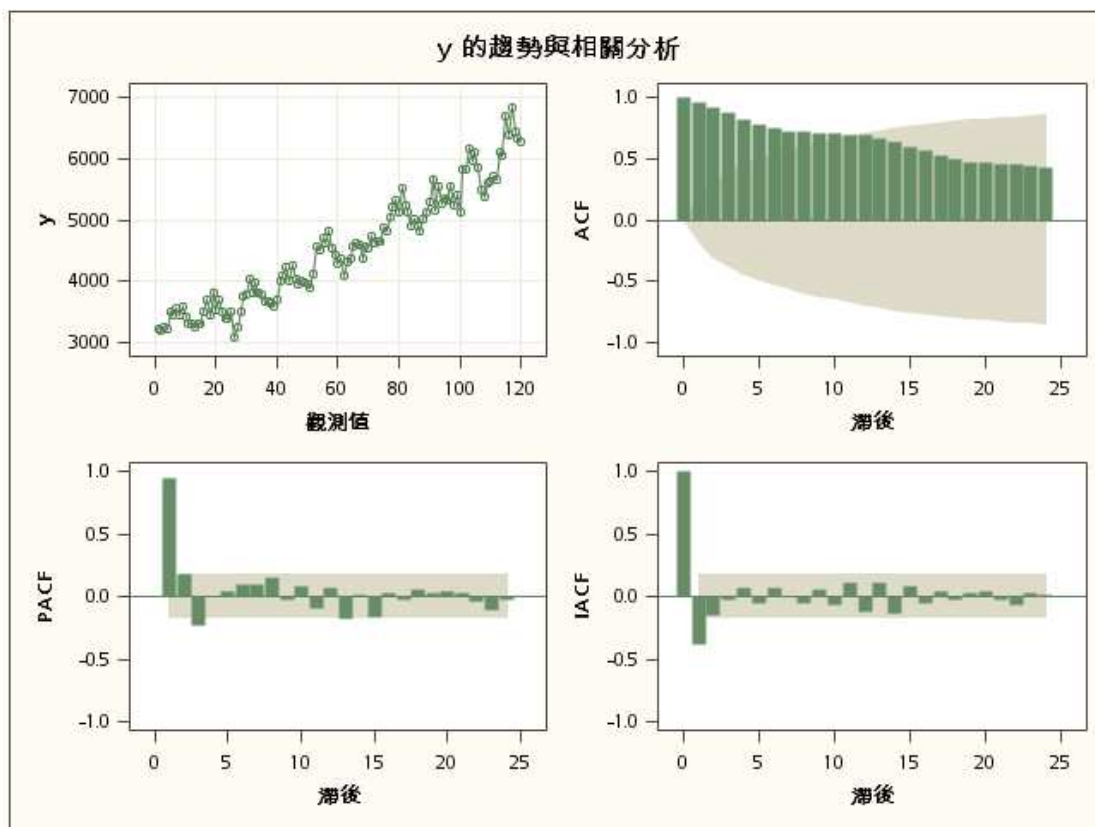


圖 3.4-2：原始資料的 ACF 及 PACF

從圖 3.4-2 可以看到，ACF dies down slowly 且 PACF cuts off after lag 2，表示平均數不平穩。當平均數不平穩時須對資料做差分，差分的方式有簡單差分、季節差分與同時做簡單差分及季節差分三種。我們先做一次差分。令 $z_t = y_t - y_{t-1}$

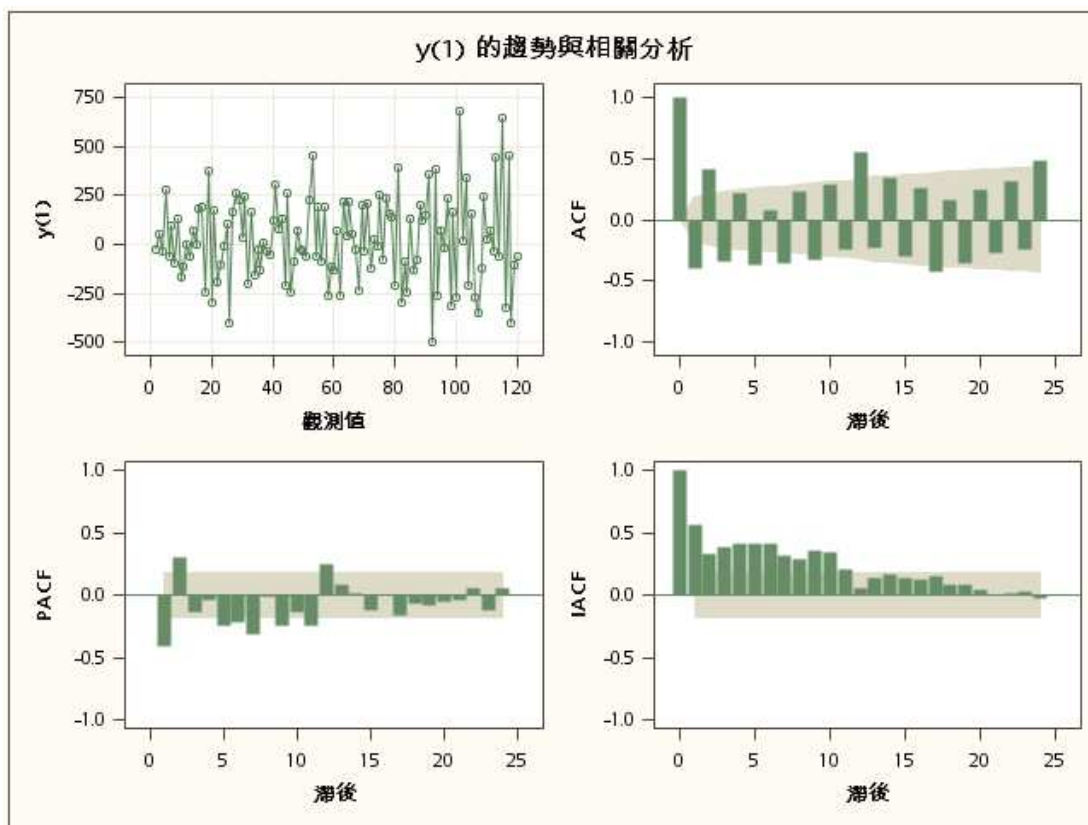


圖 3.4-3：一次差分後的 ACF 及 PACF

從圖 3.4-3 一次差分後的 ACF 可以看到，Lag12 和 Lag24 超出了兩倍標準差，所以該資料有明顯的季節循環，因此我們需要做季節差分，令 $z_t = y_t - y_{t-12}$ 。

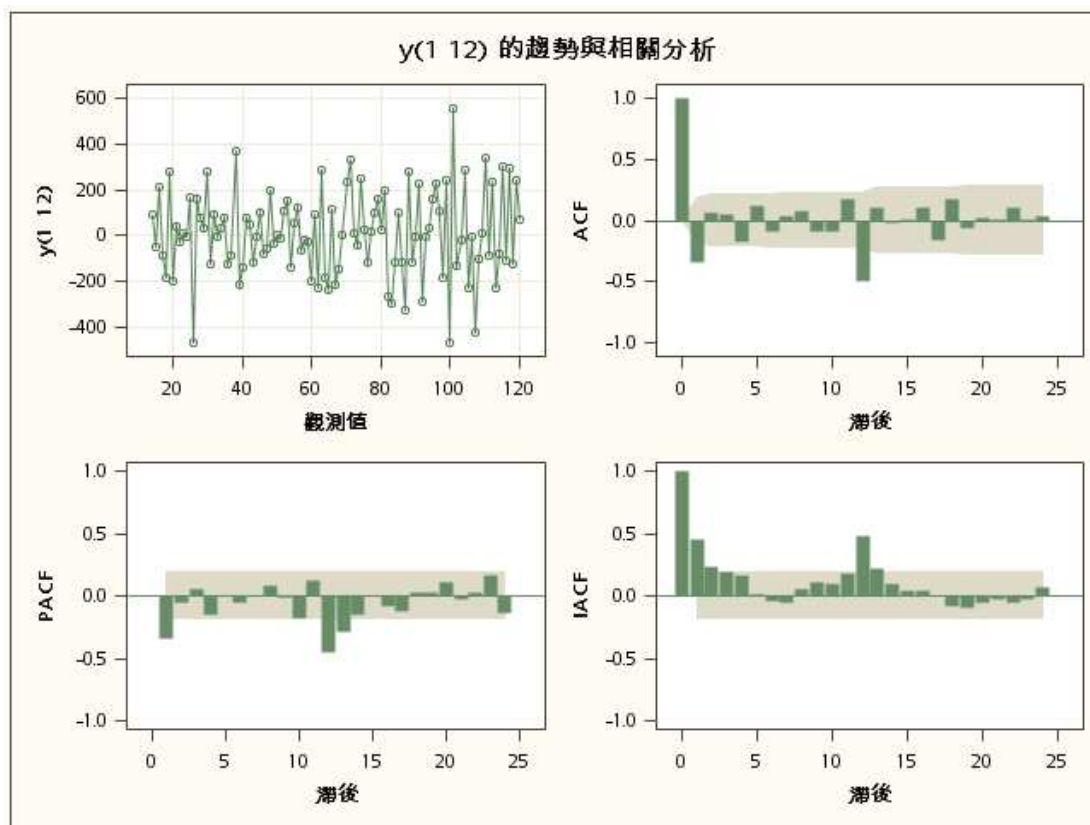


圖 3.4-4：季節差分後的 ACF 及 PACF

從圖 3.4-4 季節差分後的 ACF 及 PACF 可以看到，ACF 圖有明顯的 cut off，且 PACF 圖 dies down quickly。所以 $\{z_t\}$ 現在已經平穩了。

因為 Autocorrelation 在 Lag 12 處為負值且超出了兩倍標準差，我們配適模型 $MA(1,1)*SMA(1)$ ，即 Airline Model。我們選擇最大似估計法。

最大似估計					
參數	估計值	標準誤差	t 值	Approx Pr > t	滯後
MU	4.18975	3.02766	1.38	0.1664	0
MA1,1	0.39762	0.08663	4.59	<.0001	1
MA2,1	0.74912	0.12011	6.24	<.0001	12

常數 估計值	4.189753
變異數 估計值	18985.78
標準誤差 估計值	137.7889
AIC	1370.731
SBC	1378.75
殘差數目	107

常數項估計值為 4.18975，P-value 為 0.1664 > 0.05，不是顯著的，所以我們舍去模型中的常數項。

最大概度估計					
參數	估計值	標準誤差	t 值	Approx Pr > t	滯後
MA1,1	0.37110	0.08718	4.26	<.0001	1
MA2,1	0.72931	0.11754	6.20	<.0001	12

變異數 估計值	19242.07
標準誤差 估計值	138.7158
AIC	1370.407
SBC	1375.753
殘差數目	107

捨去常數項之後，所有參數的 P-value < 0.05，都是顯著的。將各參數估計值帶入模型中，得到我們的預測方程式如下：

$$(1-B)(1-B^{12})Z_t = (1-\theta B)(1-\Theta B^{12})a_t$$

$\hat{\theta}_1 = 0.37110$, P-value of $\hat{\theta}_1 < 0.0001$.
 $\hat{\Theta} = 0.72931$, P-value of $\hat{\Theta} < 0.0001$.
 $\hat{\sigma}^2 = 19242.07$

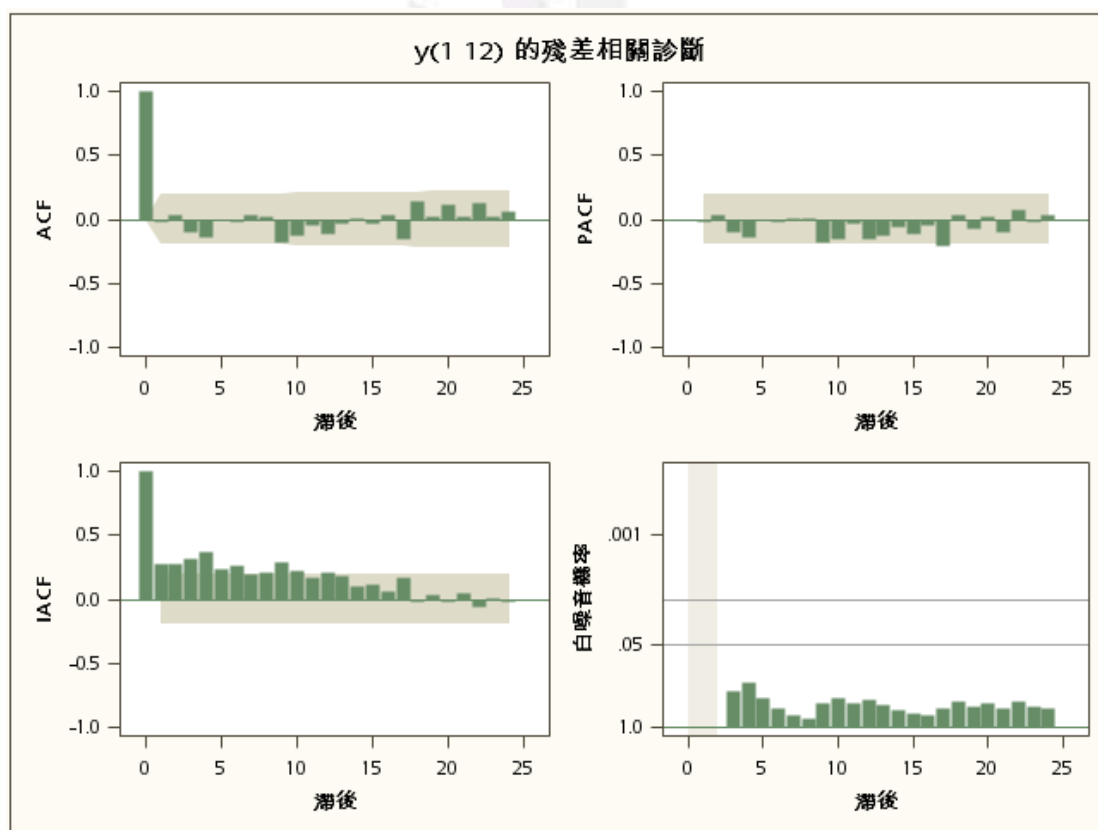
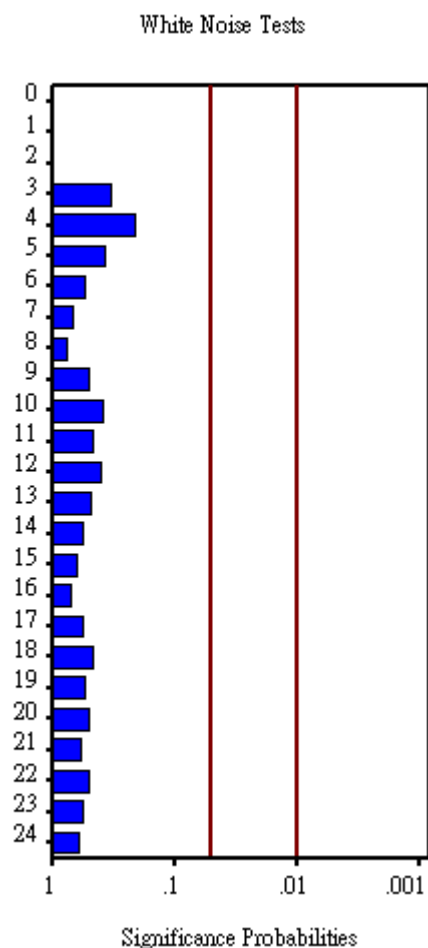


圖 3.4-5：殘差的 ACF 及 PACF

在圖 3.4-5 中，根據殘差的 ACF 圖和 PACF 圖，所有的 Lag 都在兩倍標準差之內，除了 PACF 的 Lag17 處，因此這個模型是合適的。接下來，為了檢驗模型是否合適，我們進行白噪音檢定和 Ljung-Box 檢定。一個好的配飾模型殘差平均數為 0，變異數為常數且殘差彼此不相關，即為白噪音。



在圖 3.4-6 中，我們做白噪音檢定。

$H_0: \{z_t\}$ 是白噪音

$H_1: \{z_t\}$ 不是白噪音

因為對於每一個 Lag，P-values 都大於 0.1，因此我們不能拒絕 H_0 。因此殘差是白噪音，這個模型是合適的。

圖 3.4-6: White Noise Test

殘差的自相關檢查									
至滯後	卡方	DF	Pr > ChiSq	自相關					
6	2.66	4	0.6155	-0.002	0.049	-0.082	-0.119	0.012	-0.010
12	9.30	10	0.5042	0.050	0.035	-0.160	-0.115	-0.037	-0.108
18	14.94	16	0.5291	-0.018	0.021	-0.021	0.056	-0.133	0.146
24	20.34	22	0.5619	0.027	0.117	0.026	0.140	0.030	0.064

表 3.4-1 Ljung-Box 檢定

在表 3.4-1 中，我們做 Ljung-Box 檢定。

$H_0: \rho_1 = \rho_2 = \dots = \rho_h = 0$

H_1 : 至少有一個 $\rho_i \neq 0$

從 Ljung-Box 檢定的結果我們知道，對於每一個 Lag，P-values 都大於 0.1，因此我們不能拒絕 H_0 。因此殘差是白噪音，這個模型是合適的。

預測日期	真實值	預測值	調整預測值	U95	L95	調整 U95	調整 L95
Jan-12	6476	6394.91952	6608.0835	6123.04157	6666.79747	6879.96645	6336.20056
Feb-12	5763	6281.34516	5862.58882	5960.17095	6602.51937	6183.76893	5541.40871
Mar-12	6393	6367.77913	6580.03844	6003.92735	6731.63092	6943.89691	6216.17996
Apr-12	6318	6360.69879	6360.69879	5958.67469	6762.72289	6762.73027	5958.6673
May-12	6953	6738.69849	6963.32177	6301.82479	7175.57219	7400.2035	6526.44004
Jun-12	6864	6761.14526	6761.14526	6292.00361	7230.2869	7230.29552	6291.99499
Jul-12	7566	7107.52242	7344.43983	6608.19373	7606.8511	7843.77769	6845.10197
Aug-12	7354	6830.08338	7057.75283	6302.29139	7357.87537	7585.55452	6529.95114
Sep-12	7078	7137.00446	7137.00446	6582.20752	7691.80139	7691.81159	6582.19733
Oct-12	7210	6860.71163	7089.40201	6280.16457	7441.25869	7669.95974	6508.84429
Nov-12	6800	6761.90374	6761.90374	6156.70118	7367.10631	7367.11743	6156.69006
Dec-12	6907	6672.36092	6894.77296	6043.46873	7301.25312	7523.67671	6265.8692

表 3.4-2 預測值，預測區間及真實值

在表 3.4-2 中，我們對預測值及預測區間進行了月份調整，即除以係數 α ， $\alpha=30/\text{當月天數}$ 。然後將真實值、預測值、調整預測值、上下區間、調整后的上下區間列成表格。

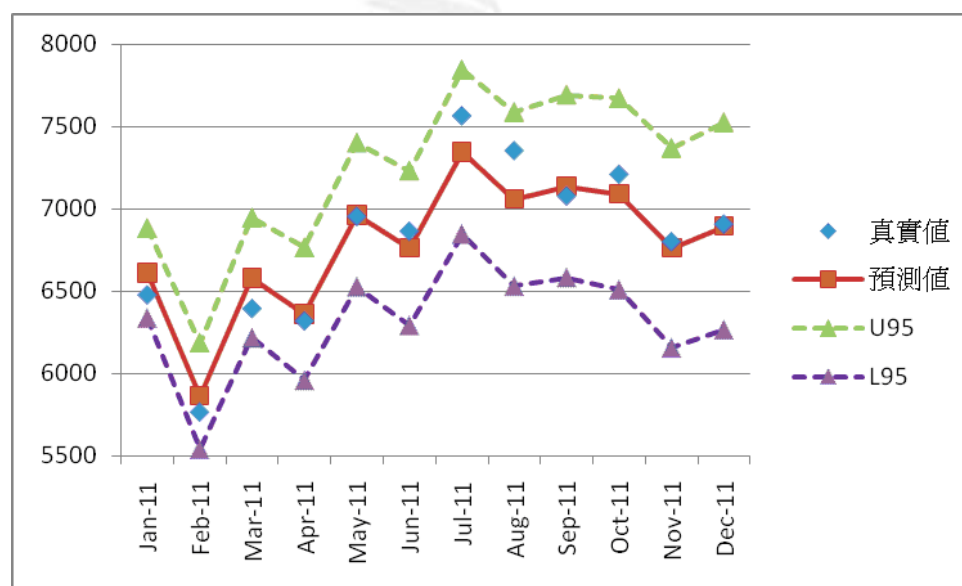


圖 3.4-7 預測值，預測區間及真實值曲線圖

在圖 3.4-7 中，我們可以看到，所有的預測值皆與真實值相去不遠，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。若從整體看來，ARIMA Seasonal Model 的預測表現，較指數平滑法、時間序列迴歸及分解法的預測表現為佳，表示該模式的預測能力還不錯。

MSE	19277.8090
MAE	110.1933
MPE	0.2095
MAPE	1.4331

表 3.4-3 MSE、MAE、MPE、MAPE 值

在表 3.3-3 中，我們可以看到 MAPE 為 1.4331 小於 5，根據前述預測評估準則，這是一個非常小的誤差，因此預測效果非常好。



第五節、最佳模型

利用在第二章第三節介紹的評估準則，以及本章使用了時間序列迴歸法、分解法、指數平滑法、ARIMA 模式，分析澳洲汽油消費量，並且將實際值與各種方法的預測值做比較，所計算出來的各個模型的 MAD、MSE、MPE、MAPE 的值，我們繪製出了表 3.5-1。

	時間序列 迴歸	分解法	指數平滑法	Seasonal ARIMA Model
ROOT MSE	457.936	537.868	208.135	138.844
MAD	392.409	446.367	166.994	110.193
MPE	5.570	5.095	1.540	0.209
MAPE	5.570	6.316	2.264	1.433

表 3.5-1 四個模型四種準則的值

從表 3.5-1 的結果可以看出，除了分解法之外，另外三種模型的 MAPE 值都小於 5，除開模型合適之外，我們可以說原始數據也是比較優質的數據。對於四種標準，Seasonal ARIMA Model 都有最小值，因此 Seasonal ARIMA Model 是最適合這一份資料的模型。

第四章 結論與建議

在本次報告中，我們共使用了時間序列迴歸法、分解法、指數平滑發和 ARIMA Seasonal Model 四種方法來分析澳門的汽油消費情況，并用樣本外預測的方法來將 2011 年的實際值和預測值作比較，結果發現在 MSE、MAD、MPE、MAPE 四個準則之下，ARIMA Seasonal Model 為最好的方法，因此我們借由 ARIMA Seasonal Model 的分析結果來得出結論和提供建議。

本次分析的結論和建議有以下幾條：

1. 澳門地區的汽油消費量存在明顯的月份波動，2 月份消費最少，然後消費量隨月份持續增加，到 7 月達到峰值，然後持續下降，直到次年 2 月。
2. 澳門地區的汽油消費量隨年份有明顯的上升趨勢，從 2001 年 3000 千公升數量級到 2011 年的 7000 千公升數量級。年增長率約為 7.8%。
3. 季節波動不會隨著年份的遞增而更大。可以預測將來澳門的汽油消費量也不會有波動的增加。

4. 汽油消費量持續平穩增長，沒有因為 2001 年的阿富汗戰爭和 2011 年的利比亞戰爭導致的國際油價變動而顯著變動，這說明汽油價格也沒有受到這兩次戰爭的衝擊，應該歸功於澳門政府的有效調控，值得台灣政府學習。



參考文獻

1. 陳婉淑 (2011), 《統計預測方法》講義
2. Bowerman, B., O'Connell, R., and Koehler, A. (2005) 《Forecasting, Time Series, and Regression》, 4th edition, Duxbury Press.
3. 澳門特別行政區政府統計暨普查局數據庫：
<http://www.dsec.gov.mo/default.aspx>
4. 維基百科
<http://www.wikipedia.com/>
5. 陳宜莉, 徐若芳, 劉彥淳, 吳憶澄, 郭根連, (2011)《台灣鮮乳月產量之分析》
<http://dspace.lib.fcu.edu.tw/handle/2377/30849>
6. 吳坤益, 王俊勛, 呂冠霖, 林欣霓, 林強相, (2011)《預測國內能源電力消費》
<http://dspace.lib.fcu.edu.tw/handle/2377/30873>

