# 加州每月平均海平面高度時間數列分析報告

109304019 統計三周佳萱 109304029 統計三李姿穎

# 目錄

壹、研究動機	2
貳、資料背景介紹	2
參、SARIMA 模型	3
一、資料前處理	3
二、SARIMA 模型選取	5
三、模型檢定	7
四、模型參數估計	9
五、模型修正一增加非季節性的 order	11
(一)、模型選擇	11
(二)、殘差檢定	12
(三)、參數估計	14
六、模型修正—考慮 2,0,012	14
(一)、模型選擇	14
(二)、殘差檢定	15
(三)、參數估計	19
七、模型選擇結果	19
八、模型預測	20
肆、SVAR、VAR 模型	20
一、採用「溫度」動機	20
二、背景資料介紹	21
三、「溫度」和「海平面」相關性	21
四、模型選擇	22
五、模型診斷	23
(一)、殘差檢定	23
(二)、參數估計	24
六、模型預測	25
伍、SARIMA、SVAR 模型預測結果比較	27
陸、結論	28
附錄:ARCH effect	28

#### 壹、研究動機

近幾年全球氣候不穩定,出現極端的降雨或乾旱,有些地區的溫度異常升高或降低,生態環境開始改變,這些變化都跟溫室效應有關。根據 Christina Nunez 在《國家地理雜誌》的〈Sea levels are rising at an extraordinary pace.〉一文指出,溫室效應加劇使全球暖化加速進行,導致北極圈冰層大量融化,沿海低海拔地區面臨海平面上升的威脅。然而,因為洋流和海水分散的方式不同,部分地區面臨的海平面上升程度,會比全球其他區域的平均高出 20%到 70%。

而位於美國西岸加州正面臨海平面上升的威脅。根據美國地質調查局(UGSG)的研究員 Sean Vitousek 在《Projections of shoreline change for California due to 21st century sea-level rise》中的預測,如果不採取更多措施來阻止溫室氣體排放,到 2100 年,加州高達 70%的海灘可能會消失的危機。目前包括北加州的 Point Arena 和 Humboldt Bay、中加州的 Pismo Beach 和 Morro Bay、南加州的 Newport Beach 和 San Clemente 在内的沿海地區已經受到嚴重侵蝕。

在這樣的背景下,我們對研究海平面高度產生了興趣。我們希望運用時間數列分析來觀察加州每月海平面高度的變化。我們知道海平面升高是不可逆的,因此能利用過去的數據來觀察海平面高度變化,可以提前採取預防措施,減少海平面上升帶來的衝擊。例如:居住在臨海地區的居民,了解每月海平面高度可以提供搬遷的參考依據。政府在海岸工程規劃中,可以運用資料確定海岸防護設施(如:防波堤、堤防等)最適合的位置和高度,以防止海浪侵蝕造成損害。此外,加州海灘的觀光價值高達930億元,對於發展觀光業的業者,海平面資料可以作為策略經營制定的重要參考。

#### 貳、資料背景介紹

本文研究數據是美國加州每月平均海平面高度。資料來自美國國家海洋與 氣象局(NOAA)的網站,內容包含 2002 年 1 月到 2022 年 12 月在美國加州聖 地牙哥地區所量測的每月平均海平面高度。

#### 資料網址如下:

https://tidesandcurrents.noaa.gov/waterlevels.html?id=9410170&units=standard&b
date=20020102&edate=20221231&timezone=GMT&datum=MLLW&interval=m&acti
on=data

每月平均海平面高度是數值型資料,以英尺為單位。在本研究中,將 2002 年 1 月到 2021 年 12 月每個月平均海平面高度(共 240 筆)作為訓練資料, 2022 年 1 月到 12 月的資料(共 12 筆)作為測試資料。

#### 參、SARIMA 模型

#### 一、資料前處理

根據圖一,我們可以觀察到資料有隨時間上升的趨勢。然而,根據圖二的結果,其中紅色虛線為 ACF 值是 0.5 的參考線,發現 ACF 值下降速度並不緩慢,所以我們不考慮對資料進行差分。

接下來,我們選擇將資料配適一次方的迴歸模型,以處理資料隨時間上升的趨勢。在圖三中,我們展示了迴歸模型的結果,紅色線代表配適的一次迴歸線。根據表一,這個迴歸模型的方程式為 $\hat{y}_t = 2.8847 + 0.0017x_t$ ,其中  $x_t$ 表示資料從 2002 年 1 月開始的期數。之後我們計算一次迴歸後的殘差  $\mathbf{e}_t = y_t - \hat{y}_t$ ,並使用 $\mathbf{e}_t$ 來進行後續的 SARIMA 模型建構。

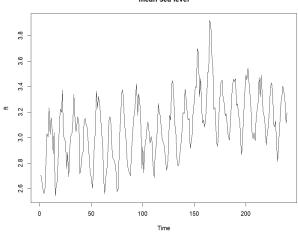
在觀察圖四時,我們可以看到資料大致在平均值為 0 的附近上下跳動,呈現出 stationary 的趨勢。為了驗證這一點,我們進行了 ADF 檢定,建立虛無假設  $H_0$ : The data is non-stationary.和對立假設  $H_1$ : The data is stationary.檢定結果顯示 p-value 小於 0.05,拒絕虛無假設  $H_0$ ,因此推論資料符合 stationary 的趨勢。

此外,由圖四我們可以發現從 2002 年 1 月到 2021 年 12 月每個月資料的 變異程度大致相同,因此我們並沒有對資料進行變數轉換。

# 表一

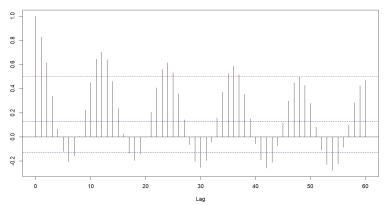
迴歸模型的係數				
	Estimate	Std.Error	t value	Pr (> t )
Intercept	2.8846	0.029	99.471	<2e-16
Time (data)	0.0017	0.000	8.266	9.79e-15

#### mean sea level

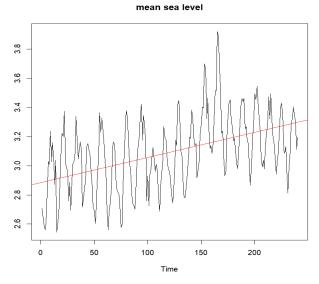


圖一、海平面原始資料的時間序列圖

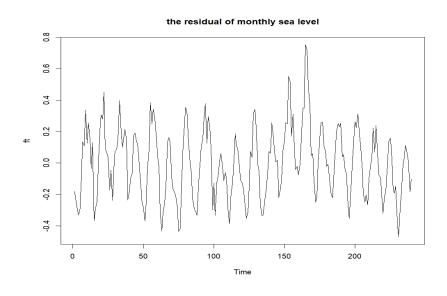
#### ACF of Mean Sea Level



圖二、原始資料的 ACF 圖



圖三、原始資料的時間序列圖加上資料配適一次方迴歸線



圖四、資料配適一次方的迴歸模型後殘差(et)的時間序列圖

## 二、SARIMA 模型選取

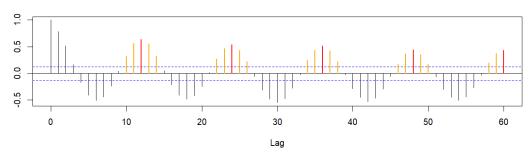
我們將一次迴歸後的殘差 $(e_t)$ 分別繪製 ACF 圖和 PACF 圖,從 ACF 圖(圖五)中的紅色線,可以發現 ACF 在  $12 \times 24 \times 36 \times 48$  等 12 的倍數上呈現比較明顯的突出現象,所以推論資料存在季節性,且週期為 12。

接下來進行模型選擇的部分,我們觀察 ACF 圖(圖五)、PACF(圖六)來判 斷模型的季節性: 1、ACF 和 PACF 都有 tails off 的現象,所以推論季節性模型可能為(1,0,1)<sub>12</sub>。
2、PACF 截斷在 12,ACF 有 tails off 的現象。從圖六可以發現 PACF 其實在 12 沒有特別突出,而是在周圍期數 10、11、14 時 PACF 較為顯著,推測可能是受到一些 noise 的影響,導致在 12 期的 PACF 並不明顯。儘管如此,考慮 PACF 截斷在 12 的可能性,我們仍然推論季節性模型可能為(1,0,0)<sub>12</sub>。

再來,我們由ACF圖(圖五)、PACF(圖六)來觀察模型的非季節性:

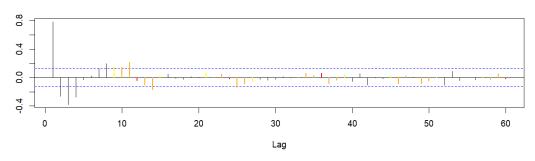
- 1、根據圖五橘色的部分,可以觀察到 ACF 在季節性期數的前後 2 期截斷,PACF 是呈現 tails off 的現象,所以推測模型非季節的部分可能為(0,0,2)。
- 2、根據圖六橘色和黃色的部分,可以觀察到 PACF 在季節性期數的前後 2、3 期截斷,ACF 是 tails off 的現象,所以推測模型非季節的部分可能為(2,0,0)、(3,0,0)。
- 3、ACF 和 PACF 都呈現 tails off,所以模型非季節的部分可能為(1,0,1)、(1,0,2)。

#### ACF of Mean Sea Level



圖五、e<sub>t</sub>的 ACF 圖

#### **PACF of Mean Sea Level**



圖六、e<sub>t</sub>的 PACF 圖

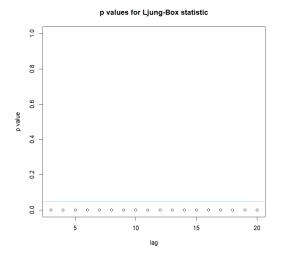
綜上所述,以下為我們推論出 10 種模型的組合:

- 1.  $SARIMA(0,0,2) \times (1,0,0)_{12}$
- 2. SARIMA(2,0,0)  $\times$  (1,0,0)<sub>12</sub>
- 3. SARIMA(3,0,0)  $\times$  (1,0,0)<sub>12</sub>
- 4. SARIMA(1,0,1) ×  $(1,0,0)_{12}$
- 5. SARIMA $(1,0,2) \times (1,0,0)_{12}$
- 6. SARIMA $(0,0,2) \times (1,0,1)_{12}$
- 7. SARIMA(2,0,0) ×  $(1,0,1)_{12}$
- 8. SARIMA(3,0,0)  $\times$  (1,0,1)<sub>12</sub>
- 9. SARIMA $(1,0,1) \times (1,0,1)_{12}$
- 10. SARIMA $(1,0,2) \times (1,0,1)_{12}$

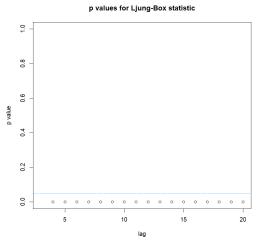
# 三、模型檢定

我們使用 Ljung-box test 對上述十個模型進行檢定,建立虛無假設  $H_0$ : The fitted model is adequate.和對立假設  $H_1$ : The fitted model is not adequate.。由圖九至圖十四和圖十七的結果,我們觀察到 p-value 小於 0.05,所以拒絕虛無假設  $H_0$ ,表示所選擇的模型不合適。因此,我們可以排除以下模型:

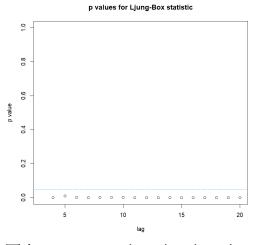
SARIMA(0,0,2) ×  $(1,0,0)_{12}$  \ SARIMA(2,0,0) ×  $(1,0,0)_{12}$  \ SARIMA(3,0,0) ×  $(1,0,0)_{12}$  \ SARIMA(1,0,1) ×  $(1,0,0)_{12}$  \ SARIMA(1,0,2) ×  $(1,0,0)_{12}$  \ SARIMA(0,0,2) ×  $(1,0,1)_{12}$  \ SARIMA(1,0,1) ×  $(1,0,1)_{12}$  \ SARIMA(0,0,2) ×  $(1,0,1)_{12}$  \ SARIMA(1,0,1) ×  $(1,0,1)_{12}$  \



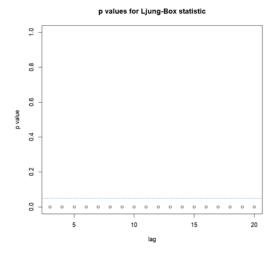
圖七、SARIMA(0,0,2)  $\times$  (1,0,0) $_{12}$ 的 Ljung-box test 圖



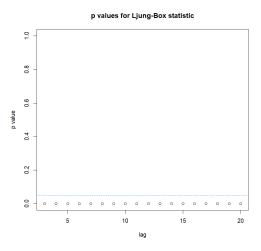
圖九、SARIMA(3,0,0)  $\times$  (1,0,0) $_{12}$ 的 Ljung-box test 圖



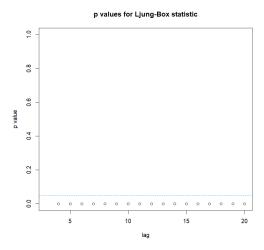
圖十一、SARIMA(1,0,2) ×  $(1,0,0)_{12}$  的 Ljung-box test 圖



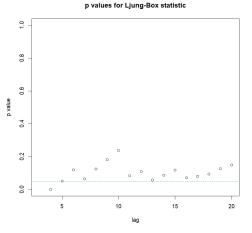
圖八、SARIMA(2,0,0)  $\times$  (1,0,0) $_{12}$ 的 Ljung-box test 圖



圖十、SARIMA(1,0,1)  $\times$  (1,0,0) $_{12}$ 的 Ljung-box test 圖



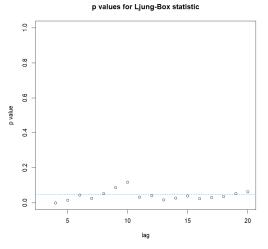
圖十二、SARIMA(0,0,2)  $\times$  (1,0,1) $_{12}$  的 Ljung-box test 圖

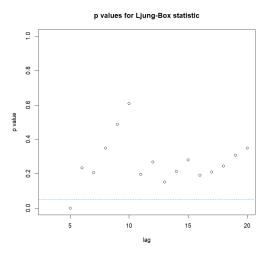


p values for Ljung-Box statistic

圖十三、SARIMA(2,0,0)  $\times$  (1,0,1) $_{12}$ 的 Ljung-box test 圖

圖十四、SARIMA(3,0,0) ×  $(1,0,1)_{12}$  的 Ljung-box test 圖





圖十五、SARIMA(1,0,1) ×  $(1,0,1)_{12}$ 的 Ljung-box test 圖

圖十六、SARIMA(1,0,2) ×  $(1,0,1)_{12}$  的 Ljung-box test 圖

#### 四、模型參數估計

觀察表二SARIMA(3,0,0) ×  $(1,0,1)_{12}$ 的參數估計值,發現非季節AR(3)的係數和 0 之間沒有顯著差異,再加上SARIMA(2,0,0) ×  $(1,0,1)_{12}$  ,就能通過 Ljung box-test 檢定,考量模型簡約性,我們決定不考慮SARIMA(3,0,0) ×  $(1,0,1)_{12}$ 這個模型。

藉由表三、表四我們可以列出SARIMA模型方程式為:

1 · SARIMA(2,0,0) ×  $(1,0,1)_{12}$ :

 $(1 - 0.5569B - 0.2121B^2)(1 - 0.9999B^{12})e_t = (1 - 0.9826B^{12})Z_t + 0.0009$ 

# 2 \ SARIMA(1,0,2) \times (1,0,1)<sub>12</sub>:

 $(1 - 0.7443B)(1 - 0.9999B^{12})e_t$ 

 $=(1-0.1898B+0.1952B^2)(1-0.9786B^{12})Z_t-0.0174$ 但觀察模型季節性的參數,我們發現SARIMA(2,0,0)× $(1,0,1)_{12}$ 左邊的  $(1-0.9999B^{12})$ 和右邊的 $(1-0.9826B^{12})$ 可以相互抵消,SARIMA(1,0,2)× $(1,0,1)_{12}$ 左邊的 $(1-0.9999B^{12})$ 和右邊的 $(1-0.9786B^{12})$ 也大致可以抵消,代表在季節性ARIMA $(1,0,1)_{12}$ 的組合中,季節性的係數對於模型的解釋能力沒有顯著影響,所以不考慮SARIMA(2,0,0)× $(1,0,1)_{12}$ 和SARIMA(1,0,2)× $(1,0,1)_{12}$ 這兩個模型。

表二

SARIMA(3,0,0) × (1,0,1) <sub>12</sub> 的參數估計							
	ar1	ar2	ar3	sar1	sma1	intercept	
	0.5720	0.2393	-0.0553	0.9999	-0.9812	0.0010	
s.e.	0.0666	0.0729	0.0675	0.0005	0.0288	0.1289	

Log likelihood=198.88, aic=-385.76

表三

SARIMA(2,0,0) × (1,0,1) <sub>12</sub> 的參數估計						
	ar1	ar2	sar1	sma1	iintercept	
	0.5569	0.2121	0.9999	-0.9826	0.0009	
s.e.	0.0641	0.0649	0.0003	0.0255	0.1429	

Log likelihood=198.56, aic=-387.11

#### 表四

SARIMA(1	$1,0,2) \times (1,0)$	,1) <sub>12</sub> 的參數	估計			
	ar1	ma1	ma2	sar1	sma1	intercept
	0.7443	-0.1898	0.1952	0.9998	-0.9786	-0.0174
s.e.	0.0667	0.0873	0.0761	0.0007	0.0327	0.1186

#### Log likelihood=200.23, aic=-388.47

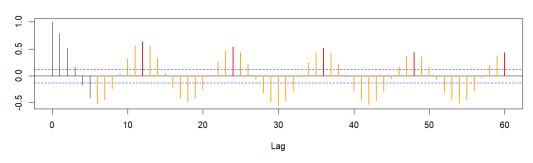
#### 五、模型修正一增加非季節性的 order

前面十個模型有的因為不符合殘差檢定,有的因為 parameter redundancy 而被刪除。又因為在 $(1,0,1)_{12}$ 和其他非季節性的搭配,會出現季節性的兩邊係數抵消的現象,接下來我們只考慮季節性 $(1,0,0)_{12}$ ,並搭配更大的非季節性的 order。

#### (一)、模型選擇

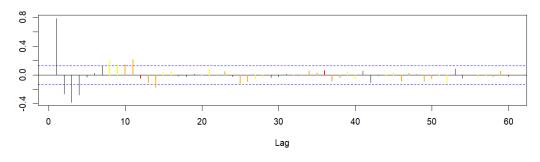
- 1、觀察圖十七橘色的部分,ACF 在季節性期數的前後  $4 \times 5 \times 6$  期截斷,PACF 是呈現 tails off 的現象,所以考慮增加非季節的部分到(0,0,4)、(0,0,5)、(0,0,6)。
- 2、觀察圖十八橘色和黃色的部分,PACF 在季節性期數的前後 4 期截斷,ACF 是 tails off 的現象,所以考慮增加非季節的部分到(4,0,0)。
- 3、ACF和 PACF都有 tails off 的現象,所以模型非季節的部分可能為(3,0,4)、(3,0,5)、(3,0,6)、(4,0,4)、(4,0,5)、(4,0,6)。

#### ACF of Mean Sea Level



圖十七、 $e_t$ 的 ACF 圖

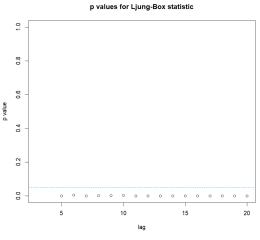
#### PACF of Mean Sea Level



圖十八、 $e_t$ 的 PACF 圖

# (二)、殘差檢定

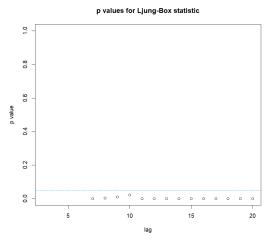
接下來,我們使用 Ljung-box test 進行檢定,建立虛無假設  $H_0$ : The fitted model is adequate.和對立假設  $H_1$ : The fitted model is not adequate.。由圖十九至二十八的結果,除了圖二十三以外,所有的 p-value 小於 0.05,拒絕虛無假設  $H_0$ ,表示模型較不合適。

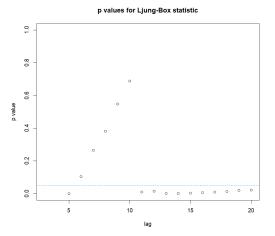


p values for Ljung-Box statistic

圖十九、SARIMA(0,0,4) ×  $(1,0,0)_{12}$ 的 Ljung-box test 圖

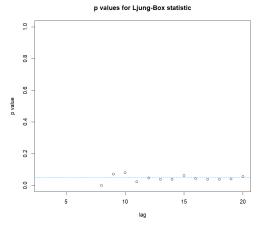
圖二十、SARIMA(0,0,5) ×  $(1,0,0)_{12}$ 的 Ljung-box test 圖



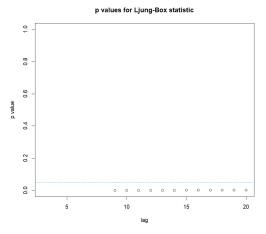


圖二十一、 $SARIMA(0,0,6) \times (1,0,0)_{12}$ 的 Ljung-box test 圖

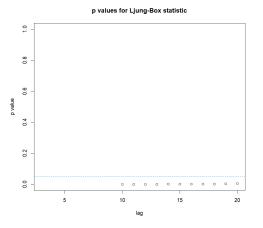
圖二十二、 $SARIMA(4,0,0) \times (1,0,0)_{12}$ 的 Ljung-box test 圖



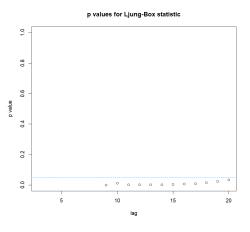
圖二十三、SARIMA(3,0,4) ×  $(1,0,0)_{12}$ 的 Ljung-box test 圖



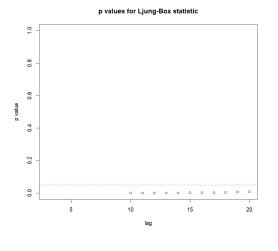
圖二十四、 $SARIMA(3,0,5) \times (1,0,0)_{12}$ 的 Ljung-box test 圖



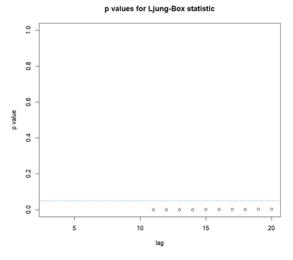
圖二十五、SARIMA(3,0,6) ×  $(1,0,0)_{12}$  的 Ljung-box test 圖



圖二十六、 $SARIMA(4,0,4) \times (1,0,0)_{12}$ 的 Ljung-box test 圖



圖二十七、SARIMA(4,0,5) ×  $(1,0,0)_{12}$  的 Ljung-box test 圖



圖二十八、 $SARIMA(4,0,6) \times (1,0,0)_{12}$ 的 Ljung-box test 圖

#### (三)、參數估計

進一步觀察 $SARIMA(3,0,4) \times (1,0,0)_{12}$ 的參數估計,根據表五,我們可以發現在季節性AR(1)的係數和 0 之間沒有顯著差異,因此,模型較不合適,不考慮 $SARIMA(3,0,4) \times (1,0,0)_{12}$ 的模型。

在嘗試非季節性的ARMA(p,q)模型搭配季節性AR(1) $_{12}$ 時,我們注意到隨著 p、q增加,季節性AR(1)的係數和 0 之間會沒有顯著差異。另外,嘗試非季節的AR(p)或 MA(q)搭配季節性AR(1) $_{12}$ 時,Ljung-box test 的 p-value 小於 0.05,模型較不合適,因此我們決定不再增加非季節性的次數,而是選擇增加季節性至AR(2) $_{12}$ ,並嘗試AR(2) $_{12}$ 與其他非季節性模型的搭配。

表五 SARIMA(3,0,4) × (1,0,0)<sub>12</sub>的參數估計圖

	ar1	ar2	ar3	ma1	ma2
	2.4463	-2.2372	0.7144	-1.9298	1.4706
s.e.	0.0499	0.0864	0.0499	NAN	NAN
	ma3	ma4	Sar1	intercept	
	-0.4033	0.1138	0.1047	-0.0021	
s.e.	0.0807	0.0555	0.0677	0.0242	

Log likelihood=198.85, aic=-377.71

六、模型修正一考慮(2,0,0)12

#### (一)、模型選擇

綜合前文第二節曾提過的非季節性模型搭配,我們列出接下來要嘗試的模型組合:

- 1. SARIMA $(0,0,2) \times (2,0,0)_{12}$
- 2. SARIMA(2,0,0)  $\times$  (2,0,0)<sub>12</sub>

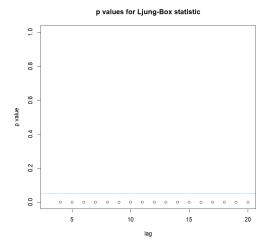
- 3. SARIMA(3,0,0)  $\times$  (2,0,0)<sub>12</sub>
- 4. SARIMA(1,0,1)  $\times$  (2,0,0)<sub>12</sub>
- 5. SARIMA(1,0,2)  $\times$  (2,0,0)<sub>12</sub>
- 6. SARIMA(2,0,1)  $\times$  (2,0,0)<sub>12</sub>
- 7. SARIMA(2,0,2)  $\times$  (2,0,0)<sub>12</sub>

#### (二)、殘差檢定

我們使用 Ljung-Box test 對模型進行檢定,建立虛無假設  $H_0$ : The fitted model is adequate.和對立假設  $H_1$ : The fitted model is not adequate.。由圖二十九 到圖三十四的結果,發現 p-value 小於 0.05,所以拒絕虛無假設  $H_0$ ,表示模型較不合適。只有圖三十五  $SARIMA(2,0,2) \times (2,0,0)_{12}$ 的 p-value 大致大於 0.05,不拒絕虛無假設  $H_0$ ,因此我們決定再增加非季節性的 order 來找尋其他可能更配適的模型。

我們考慮SARIMA(3,0,2) ×  $(2,0,0)_{12}$  、SARIMA(4,0,2) ×  $(2,0,0)_{12}$  、SARIMA(2,0,3) ×  $(2,0,0)_{12}$  、SARIMA(2,0,4) ×  $(2,0,0)_{12}$  這幾個模型,由圖三十六 到圖三十九發現只有SARIMA(4,0,2) ×  $(2,0,0)_{12}$ 的 p-value 大於 0.05,不拒絕虛無假設  $H_0$ ,模型較合適,所以接下來對SARIMA(2,0,2) ×  $(2,0,0)_{12}$ 和 SARIMA(4,0,2) ×  $(2,0,0)_{12}$ 兩個模型進行殘差檢定。

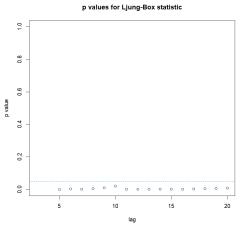
根據圖四十、圖四十一的結果,我們發現殘差在平均值為 0 的附近上下跳動,呈現出 stationary 的趨勢。觀察殘差的 ACF 圖,可以發現資料只有在一個 lag 有特別的突出,將他視為資料的 noise 所造成的影響,這表示殘差大致沒有相關性。接著,觀察殘差的 Q-Q plot,殘差大致符合常態分佈。最後,觀察 Ljung box test,發現 p-value>0.05,不拒絕虛無假設 Ho,推論出兩者模型均合適。

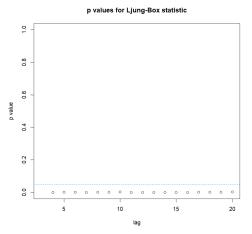


p values for Ljung-Box statistic

圖二十九、SARIMA(0,0,2) ×  $(2,0,0)_{12}$  的 Ljung-box test 圖

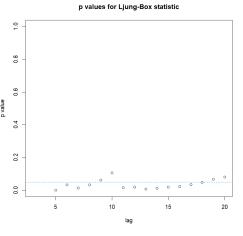
圖三十、SARIMA(2,0,0) ×  $(2,0,0)_{12}$ 的 Ljung-box test 圖

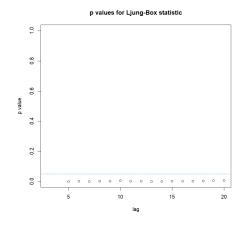




圖三十一、SARIMA(3,0,0) ×  $(2,0,0)_{12}$  的 Ljung-box test 圖

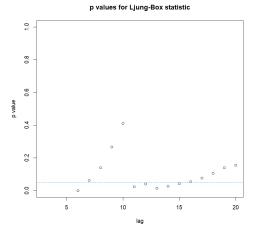
圖三十二、 $SARIMA(1,0,1) \times (2,0,0)_{12}$ 的 Ljung-box test 圖



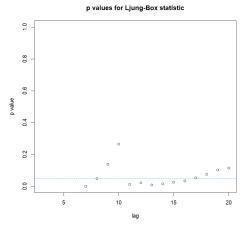


圖三十三、SARIMA(1,0,2) ×  $(2,0,0)_{12}$  的 Ljung-box test 圖

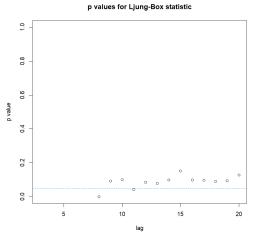
圖三十四、SARIMA(2,0,1)  $\times$  (2,0,0) $_{12}$ 的 Ljung-box test 圖



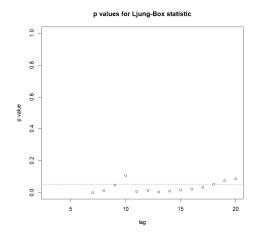
圖三十五、SARIMA(2,0,2) ×  $(2,0,0)_{12}$  的 Ljung-box test 圖



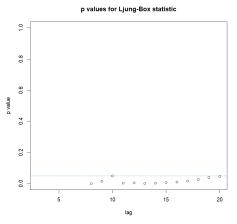
圖三十六、 $SARIMA(3,0,2) \times (2,0,0)_{12}$ 的 Ljung-box test 圖



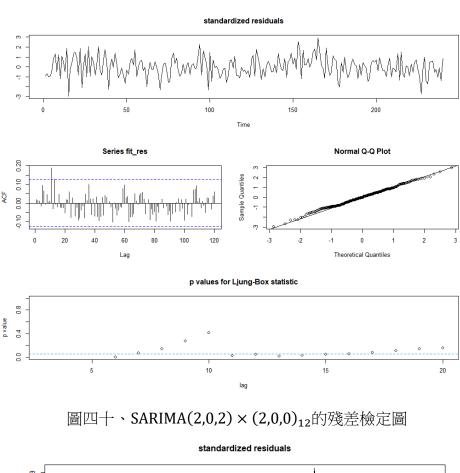
圖三十七、SARIMA(4,0,2) ×  $(2,0,0)_{12}$  的 Ljung-box test 圖

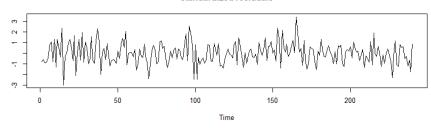


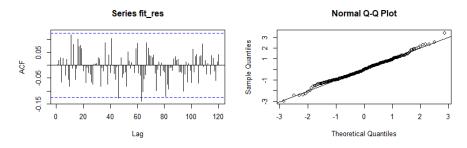
圖三十八、SARIMA(2,0,3) ×  $(2,0,0)_{12}$ 的 Ljung-box test 圖

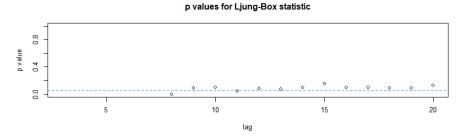


圖三十九、SARIMA(2,0,4) × (2,0,0) $_{12}$  的 Ljung-box test 圖









圖四十一、SARIMA(4,0,2)×(2,0,0)<sub>12</sub>的殘差檢定圖

#### (三)、參數估計

我們觀察SARIMA(4,0,2) ×  $(2,0,0)_{12}$ 的參數估計,由表六,我們可以發現季節性AR(1)、AR(2)的係數和 0 之間沒有顯著差異,模型季節性的係數相對不顯著,因此不考慮SARIMA(4,0,2) ×  $(2,0,0)_{12}$ 的模型,所以我們選擇SARIMA(2,0,2) ×  $(1,0,0)_{12}$  做為最終模型。

表六

SARIMA(2,0,2) × (2,0,0)<sub>12</sub>的參數估計圖

	ar1	ar2	ma1	ma2
	1.5328	-0.7515	-0.9224	0.4534
s.e.	0.0943	0.0904	0.1112	0.0753
	Sar1	Sar2	intercept	
	0.2160	0.2683	-0.0072	
s.e.	0.0709	0.0765	0.0316	

Log likelihood=185.77, aic=-357.553

表七

SARIMA(4,0,2)	×	(2,0,	$0)_{1}$	2的參	數估計	·圖
---------------	---	-------	----------	-----	-----	----

	ar1	ar2	ar3	ar4	ma1
		arz	ais	art	
	2.2560	-1.7132	0.1873	0.1945	-1.7227
s.e.	0.0648	0.1622	0.1625	0.0652	0.0198
	ma2	Sar1	Sar2	intercept	
	0.9991	0.0767	0.1343	-0.0057	
s.e.	0.0234	0.0662	0.0700	0.0299	

Log likelihood=200.65, aic=-383.3

## 七、模型選擇結果

根據迴歸分析,我們得到了以下的迴歸式: $\hat{y_t} = 2.8847 + 0.0017x_t$ ,用來

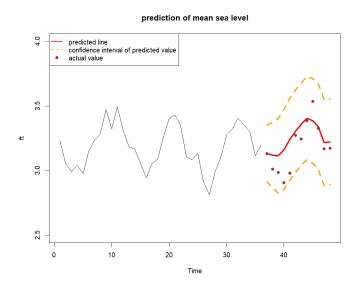
配適資料,之後我們對殘差配適SARIMA(2,0,2) ×  $(2,0,0)_{12}$ ,方程式如下: $(1-1.5328B+0.7515B^2)(1-0.2160B^{12}-0.2683B^{24})e_t=$  $(1-0.9224B+0.4534B^2)Z_t-0.0072 \circ$ 

最終我們完整模型為 $y_t = \hat{y_t} + e_t$ ,其中包含資料的迴歸式和殘差配適的 SARIMA(2,0,2) × (2,0,0)<sub>12</sub>模型。

# 八、模型預測

圖四十二為資料做一步到十二步預測的結果,預測值的線以紅色來表示, 預測的信賴區間以橘色虛線表示,而實際的每月平均海平面高度則以咖啡色表 示,我們可以發現實際資料都在資料預測的信賴區間以內。

然而,我們觀察 2022 年 2 月到 5 月、11、12 月實際的每月海平面高度資料都位於預測線之下,推測可能跟 2021 年、2022 年冬季全球出現反聖嬰現象有關。



圖四十二、SARIMA(2,0,2)×(2,0,0)<sub>12</sub>的殘差檢定圖

#### 肆、SVAR、VAR 模型

## 一、採用「溫度」動機

根據美國《國家科學院學報》的文章指出影響全球海平面上升的四個主要 因素是:冰河融化、格陵蘭冰蓋融化、南極冰蓋融化以及增溫造成的海水擴張。

( Peter. U Clark , 2012 ) 研究指出:「12 萬年前, 氣溫比現在高 1-2 度,當時的海平面大概比現在高 5-9 公尺,而在數千年的長時間尺度下,模擬結果預測出溫度每上升 1 度海平面將上升 2.3 公尺。結合現況,如果接下來兩年溫度上升了 1 度,海平面或許不會立即上升那麼多,因為地球必須變暖且隨著時間增加持續暖化,但我們排放到大氣的二氧化碳會留在大氣中數千年,所以即便我們降低排放,全球暖化導致海平面上升仍會是非常顯著的。」透過這份研究,我們認為溫度與海平面有一定程度的正相關,因此我們透過結合溫度以及海平面的歷史資料,使用 VAR 和 SVAR 模型來預測未來海平面的上升高度。

#### 二、背景資料介紹

溫度採用的數據是美國國家海洋與氣象局(NOAA)量測的美國加州聖地 牙哥每月平均氣溫(°F),資料網址:

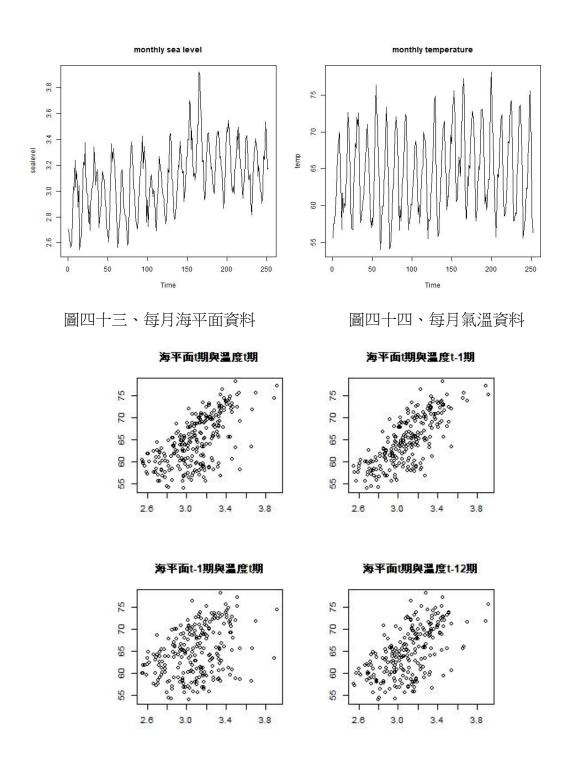
https://tidesandcurrents.noaa.gov/waterlevels.html?id=8724580&units=standard&bdate=19 711011&edate=20191211&timezone=GMT&datum=MLLW&interval=m&action=

此筆資料為數值型資料,時間從 2002 年 1 月至 2022 年 12 月,將 2002 年 到 2021 年的逐月共 240 筆數據切分為訓練資料,2022 年共 12 筆資料逐月資料則當作測試資料。

## 三、「溫度」和「海平面」相關性

從資料方面(圖四十三、四十四)來看,海平面在大約十年前有顯著的增長,而氣溫的增加幅度雖然沒有那麼多,但經由計算前十年平均氣溫:63.635°F、後十年平均氣溫:65.725°F,可知兩者大致上呈現上升趨勢。

根據圖四十五「海平面 t 期與溫度 t 期」、「海平面 t 期與溫度 t-1 期」、「海平面 t 期與溫度 t 期」、「海平面 t 期與溫度 t-12 期」的散布圖,可證實溫度與海平面之間確實存在相關性。



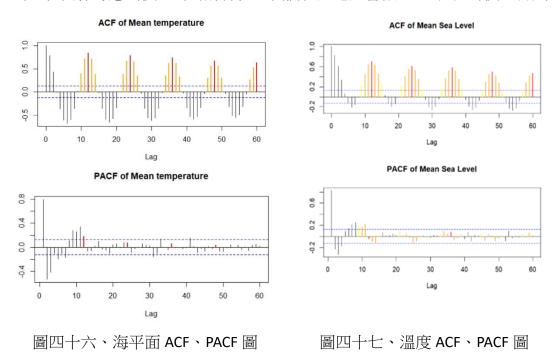
圖四十五、海平面、溫度散布圖

#### 四、模型選擇

從圖四十六、四十七,結合海平面和溫度的 ACF、PACF 圖,可發現兩個變數的 ACF 都有明顯的 tail off,考量到模型的簡約性以及可解釋性,我們只考慮 SVAR、VAR 兩種情況。SVAR 模型方面,可觀察到 ACF 間隔 12 期會特別顯著,

因此推斷季節性是以 12 期為單位,而 PACF 在溫度和海平面的 12 期附近都算顯著,所以猜測模型是 SVAR (1)。

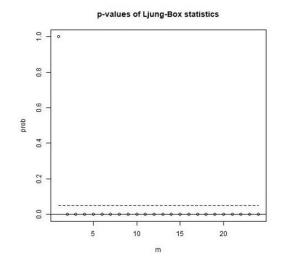
在 VAR 模型的選擇上,雖然可發現在 PACF 圖上一直到 12 期幾乎都是顯著的,但同樣考慮到模型的可解釋性,我們在這邊只會做 VAR (1)的模型嘗試。

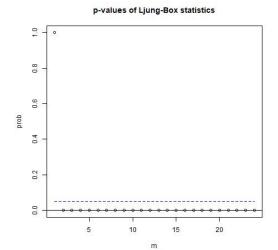


#### 五、模型診斷

#### (一)、殘差檢定

根據圖四十八、四十九,p-value 大多小於 0.05,可知這兩個模型都是不合適的,但此模型的係數及預測結果有其參考價值,所以仍會進行後續的診斷、預測。





圖四十八、SVAR(1)的 Ljung-box test 圖四十九、VAR(1)的 Ljung-box test 圖

## (二)、參數估計

#### 1、SVAR(1)模型

由表八可知, $\phi_{1,21}$ 也就是 t-12 期海平面高度對 t 期溫度的係數,其 p-value 約等於 0.385,遠大於 0.05,因此判斷此項不顯著,模型由式(一)變為式(二)。

其中  $x_{t,1}$  是海平面時間序列資料; $x_{t,2}$  是溫度時間序列資料; $a_{t,1}$ 、 $a_{t,2}$  服從白噪音

表八

1007			
SVAR(1)係	數		
	Estimate	Std. Error	Pr(> t )
$\phi_{_{f 01}}$	0.614	0.142	0.0000152
$\phi_{_{02}}$	7.648	2.190	0.000479
$\phi_{_{1,11}}$	0.663	0.054	< 2e-16
$\phi_{_{1,12}}$	0.007	0.003	0.006625
$\phi$	-0.729	0.838	0.384795
$\overset{\prime}{\psi}_{1,22}^{1,21}$	0.919	0.039	< 2e-16

#### 2、VAR(1)模型

由表九可知, $\phi_{01}$ (海平面高度的常數項)是 0.175,標準差是 0.104,可發現在 95%的信賴區間內包含 0,因此判斷此項不顯著,模型由式(三)變為式(四)。

$$\begin{pmatrix} x_{t,1} \\ x_{t,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 19.466 \end{pmatrix} + \begin{pmatrix} 0.599 & 0.017 \\ -5.543 & 0.964 \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \end{pmatrix} + \begin{pmatrix} a_{t,1} \\ a_{t,2} \end{pmatrix}$$

其中  $x_{t,1}$  是海平面時間序列資料;  $x_{t,2}$  是溫度時間序列資料;  $a_{t,1}$ 、  $a_{t,2}$  服從白噪音

表九

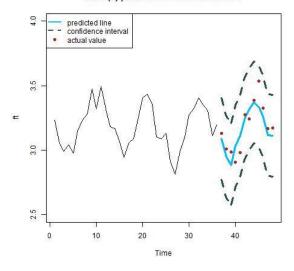
VAR(1)係數			
	Estimate	Std. Error	
$\phi_{_{01}}$	0.175	0.104	
$\phi_{_{02}}$	19.466	2.700	
$\phi_{_{1,11}}^{_{02}}$	0.599	0.040	
$\phi_{_{1,12}}^{^{1,11}}$	0.017	0.002	
$\phi$	-5.543	1.044	
$\phi_{1,22}$	0.964	0.048	
1,22			

#### 六、模型預測

#### 1、SVAR(1)模型

由圖五十可知,模型預測結果不錯,真實數據皆有包含在信賴區間內,且 真實數據大致上符合預測的結果,MSE= 0.0089。

#### SVAR(1) prediction of mean sea level

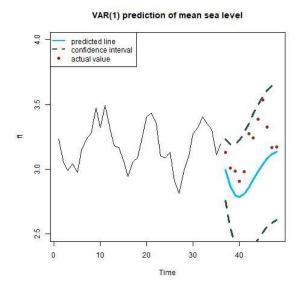


圖五十、SVAR(1)預測結果

棕色點是 2022 年真實數據;藍色實線是模型預測結果;深綠色虛線是 95%信賴區間

# 2、VAR(1)模型

由圖五十一可知,模型預測結果不是很好,真實數據雖有包含在信賴區間內,但原因可能是因為標準差大造成信賴區間範圍太大,而真實數據的週期變化也不符合預測的結果,MSE=0.0727。



圖五十一、VAR(1)預測結果

棕色點是 2022 年真實數據;藍色實線是模型預測結果;深綠色虛線是 95%信賴區間

結合以上診斷及預測結果,在多變量自迴歸模型中,雖根據 Ljung-box test 結果,我們所做的兩個模型都不適合,但兩者相比,SVAR(1)仍是較好的模型, 原因如下:MSE 值較小,預測結果有大致掌握資料的週期變化

#### 伍、SARIMA、SVAR 模型預測結果比較

SVAR 模型的 MSE 是 0.0089, 配適迴歸線的 SARIMA 模型則是 0.0170, 單比較 MSE, SVAR 是表現比較好的,但兩個模型都有各自的問題等待被解決、優化,從圖五十二可以發現,配適迴歸線的 SARIMA 模型預測結果較大,而 SVAR 則較小,從模型角度來說,我們推測原因如下:

配適迴歸線的 SARIMA 高估原因是由於海平面大幅增長的時間主要集中在 十年前,而非穩定成長至今,因此我們配適一次方的迴歸模型容易高估預測結果。SVAR 模型低估原因可能是氣溫相較海平面較無明顯增長,將較平穩的氣溫 資料納入模型可能導致低估海平面的上升情況。

# with regression SARIMA predicted line SVAR predicted line actual value 98 00 10 20 30 40 Time

圖五十二、SARIMA、SVAR 模型預測結果 棕色點是 2022 年真實數據;藍色實線是 SVAR 模型預測結果; 紅色實線是 SARIMA 模型預測結果

#### 陸、結論

上一小節我們從模型角度解釋了可能造成誤差的原因,但或許這是無法從任何模型預測得到的結果,而是必須從自然科學的角度做判讀。

我們預測的資料是 2022 年度的海平面資料,但在 2021、2022 年冬季,全球連續 2 年出現反聖嬰現象。量測地點加州位在美國西部,靠近太平洋東部,加上當地大部分降雨發生在冬季,因此反聖嬰會造成加州在冬天和春天兩個季節的降雨明顯減少,這可能是 SARIMA 模型每月平均海平面的預測較實際值高估的原因。反聖嬰現象也會造成加州冬季較暖,夏季更為炎熱,整年溫度變化幅度不大,造成每月平均海平面的預測較實際值低估的原因。

氣象專家指出,反聖嬰現象是自然現象,但這一波反聖嬰已經持續 2 年多, 預估還會持續到年底,反聖嬰現象連續 3 年發生的情況相當罕見,雖然可暫時 減緩全球氣溫上升,但氣候變遷已經不可逆,甚至會和反聖嬰和聖嬰現象產生 交錯影響,南威爾斯大學氣候變遷研究副教授塔斯切圖表示,「因為全球暖化, 極端聖嬰和極端反聖嬰將會在未來愈發頻繁。」

這種情況表明,在海平面預測方面,使用過去的海平面高度,或者使用過去的溫度資料可能無法完全捕捉到這種非週期的海流現象和氣候變遷的影響。 或許我們需要考慮更多的變數,像是溫室氣體排放量、海流循環等影響,來提高海平面高度預測的準確性。

#### 附錄:ARCH effect

為了觀察資料是否存在 ARCH effect,將資料配適完迴歸模型和 SARIMA 後, 繪製模型殘差平方的 PACF 圖。根據圖五十三的結果,發現 PACF 在期數為 12 和 20 時有突出的現象,因此我們覺得資料可能受到 ARCH effect 的影響。

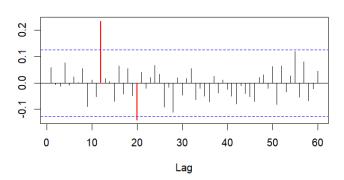
因此,我們進行 ARCH effect 的相關檢定,首先,我們使用 Portmanteau-Q test 來檢定殘差是否具有波動性,建立虛無假設  $H_0$ : Residual is white noise.和對立假設  $H_1$ : Residual is not white noise.。根據圖五十四,檢定結果顯示 p-value

大於 0.05,不拒絕虛無假設 Ho,推論殘差是 white noise,沒有明顯的波動性。

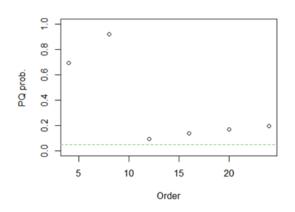
再來,我們使用 Lagrange Multiplier test,檢定 ARCH model 的係數是否顯著,建立虛無假設 $H_0:\alpha_0=\alpha_1...=\alpha_n=0$ ,和對立假設 $H_1:\alpha_i$  is not equal to zero.。根據圖五十五的結果,檢定結果顯示 p-value 大於 0.05,不拒絕虛無假設  $H_0$ ,推論 ARCH model 的係數並不顯著。

綜上所述,資料整體來看沒有明顯的 ARCH effect,所以不再對他進行 ARCH 模型的配適。

#### PACF of residuals square



圖五十三、SARIMA(2,0,2) × (2,0,0)<sub>12</sub>的殘差平方的 PACF 圖



圖五十四、Portmanteau-Q test 檢定結果

圖五十五、Lagrange multiplier test 檢定結果