

實質 GDP 的時間序列分析：ARIMA 模型選擇與預測

Time Series Analysis of Real GDP: ARIMA Model Selection and Forecasting

學生：陳泱兌

摘要

本研究採用時間序列分析方法，針對美國實質國內生產毛額(Real Gross Domestic Product, Real GDP)進行實證研究與預測分析。研究資料取自美國聯邦儲備系統經濟資料庫(Federal Reserve Economic Data, FRED)，樣本期間為 1947 年第一季至 2024 年第二季，資料頻率為季資料。考量實質 GDP 之原始資料具有明顯趨勢與非定態特性，本研究透過取自然對數轉換與一階差分方式，將原始資料轉換為定態時間序列，並使用 Augmented Dickey-Fuller (ADF) 檢定，確認其定態性。

在模型建構方面，本研究採用自我迴歸整合移動平均模型(ARIMA)，並結合自相關函數(ACF)、偏自相關函數(PACF)、赤池與貝氏資訊評選準則(Akaike Information Criterion, AIC；Bayesian Information Criterion, BIC)進行模型期數(Order)選擇。實證結果顯示，在大樣本情況下，依據 BIC 準則，會選取之較精簡模型：ARMA(1,0)。接下來，本研究進一步將樣本劃分為訓練集與驗證集，以樣本外預測作為模型效能評估依據，並以平均絕對誤差(MAE)與均方根誤差(RMSE)衡量預測準確度。

殘差分析結果顯示，模型殘差不具顯著序列相關性，顯示 ARMA(1,0)能有效捕捉實質 GDP 成長率之主要動態結構，並且符合同質性假設，但不服從常態分配，呈現厚尾(Heavy tailed)特性。

實證結果顯示，ARMA(1,0)模型在經過訓練集擬合後，於驗證集中之預測表現有所提升，且模型之一階自我迴歸係數在統計上顯著。不過，由於模型僅考慮了一階自我迴歸項與常數項，對於波動較為劇烈之時間序列資料，其適用性可能受限，因此，未來仍有進一步改良與延伸模型之研究空間。

關鍵詞：實質國內生產毛額、時間序列分析、ARIMA 模型、定態性檢定、總體經濟預測

目錄

第一章 緒論	5
1.1 研究背景與問題	5
1.2 研究架構與流程	5
第二章 文獻回顧	6
第三章 研究方法	7
3.1 ARIMA 模型	7
3.2 AUGMENTED DICKEY-FULLER (ADF) TEST	8
3.3 LJUNG-BOX TEST	9
3.4 JARQUE-BERA TEST	10
3.5 WHITE TEST	10
3.6 透過觀察法判斷時間序列模型階數	11
3.7 透過資訊評選準則判斷時間序列模型階數	12
3.8 模型評估指標	13
第四章 實證分析與結果	13
4.1 資料準備與處理	14
4.2 敘述統計分析	17
4.3 透過觀察法與資訊評選準則判斷 ARIMA 模型期數	18

4.3.1	自相關函數(ACF)與偏自相關函數(PACF)	18
4.3.2	赤池資訊評選準則(AIC)與貝氏資訊評選準則(BIC) ..	19
4.4	設定資料訓練集與驗證集	21
4.5	模型擬合與預測分析	21
4.6	模型檢定分析與評估指標	22
第五章 結論及建議		27
第六章 參考文獻		27
一、	中文文獻	27
二、	英文文獻	27

圖目錄

圖 1、美國實質 GDP 原始資料之時間序列圖.....	14
圖 2、針對美國實質 GDP 原始資料(非定態時間序列)進行 ADF test	16
圖 3、差分後美國實質 GDP 資料之時間序列圖	16
圖 4、針對差分後美國實質 GDP 資料(定態時間序列)進行 ADF test	17
圖 5、美國實質 GDP 原始資料與差分後美國實質 GDP 資料之敘述統計 量.....	17
圖 6、差分後美國實質 GDP 資料之 ACF(上)和 PACF(下)圖.....	18
圖 7、美國實質 GDP 原始資料之 ACF(上)和 PACF(下)圖	19
圖 8、計算不同 ARMA 模型的 AIC 與 BIC.....	20
圖 9、不同模型下之 AIC 與 BIC 值最小的模型	20
圖 10、透過 auto.arima()自動選擇最適模型	21
圖 11、ARMA(1,0)模型於驗證集資料之預測值與實際值時間序列圖.	22
圖 12、Ljung-Box test 檢測模型殘差是否具序列相關性之結果	23
圖 13、Jarque-Bera test 檢測模型殘差是否服從常態分配之結果..	23
圖 14、針對模型殘差之 Normal Q-Q plot	24
圖 15、模型殘差時間序列圖	25
圖 16 使用 Python 進行異質性檢定之結果	26
圖 17、模型評估指標 MAE, RMSE 結果	26

第一章 緒論

1.1 研究背景與問題

國內生產毛額(Gross Domestic Product, GDP)討論的是一國家或地區境內，於特定期間生產之最終商品勞務總市場價值，GDP 作為在總體經濟中針對國民經濟核心最重要的指標之一，與國家經濟生產行為息息相關，涵蓋民間消費、政府投資與消費、進出口等資訊。而考慮到通貨膨脹和物價變動等因素，我們為了比較在不同時間下的經濟活動水準，故將固定年份之價格水準作為基期，來衡量商品生產量的變化，從而消除通貨膨脹的影響，即為實質國內生產毛額(Real GDP)，以下以實質 GDP 表示。

實質 GDP 為衡量國家景氣循環與生活水準的總體指標，若我們能夠有效地正確預測其長期趨勢，將能夠對於政府決策、經濟決策、企業投資、金融市場等面向，提供潛在機會與風險資訊。然而，現實世界中的經濟體系是十分複雜的系統，因此，要能準確預測實質 GDP 的趨勢與變化並非易事。在本研究中，我將使用 ARIMA 時間序列模型，針對美國實質 GDP 資料進行分析。

1.2 研究架構與流程

在本研究中，我使用 R 與 Python 程式語言進行資料準備、資料處理、檢驗時間序列資料定態性、敘述統計分析、ARIMA 模型期數判斷、設定資料訓練集與驗證集、模型擬合與預測分析，並使用 Ljung-Box test, Jarque-Bera test, White test 檢驗模型殘差之序列相關性、同質性與常態分配假設。

首先，我透過美國聯邦儲備系統經濟資料庫(Federal Reserve Economic Data, FRED)取得美國實質 GDP 資料，時間範圍為 1947 年 1 月 1 日至 2024 年 7 月 1 日，並該資料為季資料。

再來，針對原始資料進行資料視覺化，進行初步瞭解，透過圖像分析判斷在原始資料中是否存在明顯固定趨勢或者季節性，若時間序列資料為非定態，則需要透過取自然對數與一階差分方式轉換為定態時間序列，並

透過 Augmented Dickey-Fuller (ADF) test 來檢定時間序列資料是否具有單根 (unit root)，若拒絕虛無假設，則代表被解釋變數 y_t 不具單根，此時為定態時間序列。

然後，我針對原始資料(非定態時間序列)與一階差分轉換後的資料(定態時間序列)分別計算出其敘述統計量，包含平均數(Mean)、標準差(Standard deviation)、偏態係數(Skewness)與峰態係數(Kurtosis)，並針對資料動差進行分析。

接著，我透過觀察法與資訊評選準則來判斷 ARIMA 模型的 p, q 階數 (order)，其中，觀察法即透過觀察自相關函數(Autocorrelation Function, ACF)與偏自相關函數(Partial Autocorrelation Function, PACF)之圖形呈現進而做判斷；資訊評選準則(Information criteria)即透過赤池資訊評選準則(Akaike information criteria, AIC)與貝式資訊評選準則(Bayesian information criteria, BIC)之數值呈現進而做判斷，當資訊評選準則越小，則該模型越佳。

在設定好 ARIMA 模型的階數後，我將轉換過的定態時間序列資料分成訓練集(train set)與驗證集(validation set)，其中，訓練集資料範圍為 1947 年 1 月 1 日至 2022 年 10 月 1 日；驗證集資料範圍為 2023 年 1 月 1 日至 2024 年 7 月 1 日。

然後，我使用設定好的 ARIMA 模型，首先針對訓練集做擬合(Fitting)，再使用模型對驗證集進行預測，並且，畫出時間序列圖以便於辨識模型在驗證集中的實際值與預測值狀況，同時，我透過平均絕對誤差(Mean Absolute Error, MAE)與均方根誤差(Root mean Squared Error, RMSE)作為模型評估指標。

最後，我透過 Ljung-Box test 檢驗模型的殘差是否具序列相關，並透過 Jarque-Bera test 檢驗模型的殘差是否服從常態分配，再透過 White test 檢驗模型的殘差是否符合同質性假設。

第二章 文獻回顧

自 Box and Jenkins(1976)提出系統化的時間序列模型建立流程以來，ARIMA 模型即為時間序列預測中最具代表性的單變量方法之一。該方法強

調模型識別(Identification)、參數估計(Estimation)與分析(Diagnostic checking)三個步驟，透過自我迴歸(AR)、移動平均(MA)與差分(I)成分，捕捉時間序列資料中的動態結構。

大多數的總體經濟之時間序列資料會具有趨勢性與持續性，若我們直接於非定態資料上進行迴歸分析，可能導致虛假迴歸(Spurious Regression)問題，其結果是不可信的。Dickey and Fuller(1981)提出之單根檢定(ADF test)，成為檢定時間序列資料之定態性的標準方法，應用於時間序列資料分析中。

Nelson and Plosser(1982)針對美國多項總體經濟變數進行實證研究，發現多數變數呈現單根(Unit root)，顯示具有隨機趨勢(Stochastic trend)。因此，將非定態時間序列資料進行自然對數轉換與差分處理，成為實證研究中常見的資料預處理方式，以確保模型估計與預測結果的有效性。

此外，殘差分析在實證研究中也是重要的一環。Ljung and Box(1978)提出的檢定方法可用以檢定模型殘差是否仍存在序列相關性，若殘差近似白噪音，則表示模型能有效捕捉資料中的系統性結構。

第三章 研究方法

在第一章節中，我們說明了實質 GDP 對於國家總體經濟面之重要性，並在此研究中，將嘗試使用時間序列模型分析美國實質 GDP 資料，因此，在此章節中，我將分別說明在此研究中會使用的時間序列模型，即 ARIMA 模型，以及定態時間序列資料檢定方式與時間序列模型檢定方式，即 ADF test, Ljung-Box test, Jarque-Bera test, White test，以及判斷時間序列模型階數(order)之觀察法與資訊評選準則，即自相關函數(ACF)、偏自相關函數(PACF)與赤池資訊評選準則(AIC)、貝式資訊評選準則(BIC)。此外，也會說明作為模型評估指標之平均絕對誤差(MAE)與均方根誤差(RMSE)。

3.1 ARIMA 模型

ARIMA 模型，即自我迴歸整合移動平均模型(Autoregressive Integrated Moving Average, ARIMA)，可表示如下：

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

其中， Y_t 為解釋變數在 t 時間下的值， α 為常數項， $\phi_1, \phi_2, \dots, \phi_p$ 為自我迴歸參數之階數 p ， $\theta_1, \theta_2, \dots, \theta_q$ 為移動平均參數之階數 q ， ϵ_t 為誤差項在 t 時間下的值。

ARIMA 模型由自我迴歸模型(Autoregressive model, AR model)與移動平均模型(Moving Average model, MA model)以及整合(Integrated, I)元素組成。

其中，AR(p) model 使用變數當期與變數的前 p 期之歷史資料，用以解釋或預測未來之變數值，而 p 階數代表模型中前 p 期資料的意思，若 p 為 1，則代表模型使用前一期歷史資料來解釋當期資料，也就是 AR(1)模型；MA(q) model 代表使用當期白噪音(White Noise)與前 q 期之白噪音資料的線性組合，以作為當期被解釋變數 y_t 的解釋變數，若 q 為 1，則代表模型使用前一期白噪音資料作為解釋變數，以解釋當期被解釋變數，意即 MA(1)模型；I 代表差分(difference)次數的意思，目的是讓時間序列資料能為定態序列，關於 ARIMA 模型，研究者會通常會以 ARIMA(p, d, q)形式呈現，而若設定參數 $d = 1$ ，則代表進行一階差分(first - order differencing)。

3.2 Augmented Dickey-Fuller (ADF) test

ADF test 的目的是用來檢定時間序列資料是否具單根(unit root)，由於大部分時間序列資料呈現隨機趨勢(stochastic trend)，若隨機趨勢發生，則會對時間序列產生影響，因此，在進行模型階數決策與預測分析之前，我們需要透過 ADF test 檢驗時間序列資料是否具有單根，若資料存在單根，則表示該資料為非定態(non-stationary)資料，此時，我們應對資料進行差分處理，使資料呈現穩定狀態。另外，若我們想針對非定態時間序列資料進行迴歸分析的話，則因為在序列中使用的變數具有單根，故會產生虛假迴歸(spurious regression)問題，其結果是不可信的。

以下針對 Dicky&Fuller(1981)提出的 ADF test 進行說明，其迴歸模型如下：

$$\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{i=1}^p \gamma_i \Delta Y_{t-i} + \epsilon_t$$

其中， Y_t 為時間序列， α 為常數項， δ 為自我相關係數， $\sum_{i=1}^p \Delta Y_{t-i}$ 為時

間序列資料的落後項，而 ϵ_t 為白噪音。

ADF test 的虛無假設與對立假設為：

$H_0: \delta = 0$ ，時間序列具單根(非定態)；

$H_1: \delta \neq 0$ ，時間序列不具單根(定態)。

若拒絕虛無假設，代表 y_t 不具單根，為定態時間序列；若無法拒絕虛無假設，代表 y_t 具單根，為非定態時間序列。

3.3 Ljung-Box test

Ljung-Box test 的目的是檢定時間序列資料是否存在序列相關(serial correlation)問題。

假設一時間序列資料 $\{(X_t, Y_t)\}_{t=1}^T$ ，並令母體迴歸模型為：

$$Y_t = \alpha + \beta X_t + u_t$$

若不同期間的誤差項存在相關性，則此時誤差項間具序列相關，即：

$$\text{Cov}(u_t, u_{t-s}) \neq 0, \forall s > 0$$

對於時間序列模型而言，理想上，模型殘差應不存在趨勢，並且應為白噪音，代表模型能夠合理捕捉時間序列中的趨勢。

在 Ljung-Box test 中，其檢定方式是利用模型中的殘差項所計算出的 k 階樣本自我相關係數 $\widehat{\tau}_k^2$ 進行檢定，其中，檢定統計量為：

$$Q = T(T+2) \sum_k^m \frac{\widehat{\tau}_k^2}{T-k} \sim \chi^2(m)$$

而其虛無假設與對立假設為：

$H_0: \tau_k = 0, k = 1, 2, \dots, m$ ，時間序列的殘差不具序列相關性；

$H_1: \tau_k \neq 0, k = 1, 2, \dots, m$ ，時間序列的殘差具序列相關性。

3.4 Jarque-Bera test

Jarque-Bera test 的目的是檢定資料是否服從常態分配，檢定統計量為：

$$n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right] \xrightarrow{d} \chi^2(2)$$

其中，S 為樣本偏態係數(Skewness)；K 為樣本峰態係數(Kurtosis)。

而其虛無假設與對立假設為：

H_0 ：資料母體服從常態分配；

H_1 ：資料母體不服從常態分配。

3.5 White test

White test 的目的是檢測迴歸模型中的殘差項是否存在異質性(heteroskedasticity)。

假設一樣本多元迴歸模型：

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki}$$

則模型殘差為：

$$\hat{u}_i = Y_i - \hat{Y}_i$$

在 White 檢定中，我們利用模型殘差平方 \hat{u}_i^2 對所有解釋變數的一次項、平方項與交乘項進行輔助迴歸，並得到判定係數 R^2 ，其檢定統計量即為資料樣本數 n 乘上判定係數 R^2 ，且檢定統計量服從卡方分配，其自由度為我們在輔助迴歸中估計的 k^* 個斜率項參數，意即：

$$\varphi = nR^2 \sim \chi^2(K^*)$$

而其虛無假設與對立假設為：

H_0 ：模型殘差具同質性(homoskedasticity)，即 $Var(u | X) = \sigma^2$ ；

H_1 ：模型殘差具異質性(heteroskedasticity)，即 $Var(u | X) \neq \sigma^2$ 。

另外，由於此研究主要使用 R 語言進行，而在 R 語言中，若要使用 White test，則可以透過 lmtest 套件提供的 bptest() 函式實現，並將函式中的變數 studentize 設定為 FALSE，就能使用 White test，若不這樣做的話，程式預設會使用 Breusch-Pagan test，其缺點是在輔助迴歸中，僅檢定模型殘差平方與解釋變數，而 White test 則加入了解釋變數平方項與交乘項。問題在於 R 語言中的 bptest() 函式，因為它的預設會需要輸入一迴歸模型(原理即上述說明的針對模型殘差與解釋變數建構之輔助迴歸)，才能得到統計結果，因此，我後來針對模型殘差同質性檢驗問題，改成使用 Python 語言中 statsmodels 套件提供的 summary() 函式來檢驗，其能夠自動針對擬合後的模型殘差與解釋變數進行運算，並提供報表結果，若使用時間序列模型，則會包含模型參數之係數與標準誤、Ljung-Box test, Jarque-Bera test, Heteroskedasticity test，甚至會自動計算該 ARIMA 模型階數之 AIC 與 BIC 值。相對來說，若使用 R 語言的話，則需要分別透過不同函式來針對模型殘差是否具序列相關、同質性與常態分配進行檢驗。不過，根據後續透過使用 R 語言與 Python 產出之結果進行交叉比對，我發現在使用不同程式語言下，其透過不同套件與函式之計算與統計假設檢定結果皆相同。

3.6 透過觀察法判斷時間序列模型階數

在時間序列分析中，我們可先透過自相關函數(ACF)與偏自相關函數(PACF)來大略判斷時間序列模型之階數，以下說明 ACF 與 PACF 與判斷方式。

自相關函數(Autocorrelation Function, ACF)，為描述解釋變數 Y_t 與其落後期(lag)解釋變數在時間序列中的相關程度，我們可透過其判斷 MA model 之期數 q ，若我們發現時間序列資料的 ACF 圖呈現在第 q 期出現截斷(cut off)現象，則我們可大略判斷其應為 MA(q)model。

偏自相關函數(Partial Autocorrelation Function, PACF)，為描述解釋變數 Y_t 與在排除特定落後期之解釋變數之後的相關程度，即在控制了特定落後期數之下的解釋變數之間相關程度。假設在控制落後期數為 3 期下，則代表我們想觀察的是 Y_t 與 Y_{t-3} 之間的相關程度。我們可透過其判斷 AR model

之期數 p ，若我們發現時間序列資料的 PACF 圖呈現在第 p 期出現截斷(cut off)現象，則我們可大略判斷其應為 $AR(p)$ model。

3.7 透過資訊評選準則判斷時間序列模型階數

除了觀察法之外，我們也可以透過資訊頻選準則(information criteria)來決定最適模型之期數，其中，常見的資訊評選準則如赤池資訊評選準則(Akaike Information Criteria, AIC)與貝氏資訊評選準則(Bayesian Information Criteria, BIC)。

赤池資訊評選準則(Akaike Information Criteria, AIC):

$$AIC = T \ln \left(\frac{\sum_t \hat{\epsilon}_t^2}{T} \right) + 2K$$

其中， T 為樣本數， $\hat{\epsilon}_t^2$ 為模型進行適配後的殘差， K 為模型參數。

貝氏資訊評選準則(Bayesian Information Criteria, BIC):

$$BIC = T \ln \left(\frac{\sum_t \hat{\epsilon}_t^2}{T} \right) + K \ln T$$

其中， T 為樣本數， $\hat{\epsilon}_t^2$ 為模型進行適配後的殘差， K 為模型參數。

對於 $ARMA(p,q)$ 模型而言，我們可以依靠觀察法或研究者自身經驗等猜測模型最適可能期數，並透過計算 AIC 與 BIC 的值，輔助進行挑選模型最適期數，當 AIC 和 BIC 的值越小，則代表模型越好。而我們可以發現到，AIC 和 BIC 之最大差別在於公式後方懲罰項的部分，差別在於 2 與 $\ln T$ 的部分，若 $T \geq 8$ ，則 $\ln T > 2$ ，也就是說，相較於 AIC 而言，BIC 給予的懲罰項較大，因此，相較 AIC 而言，BIC 會挑選較精簡的模型。

在此研究中，我透過觀察法首先判斷應使用 $ARIMA(1,0,0)$ 模型後，並額外考慮了 $ARIMA(2,0,0)$ 、 $ARIMA(3,0,0)$ 、 $ARIMA(1,0,1)$ 、 $ARIMA(1,0,2)$ 、 $ARIMA(1,0,3)$ 、 $ARIMA(2,0,1)$ 、 $ARIMA(2,0,2)$ 、 $ARIMA(2,0,3)$ 等模型，再透過 R 語言計算各模型的 AIC 與 BIC 值，最後，我選擇採用 BIC 資訊選取模型最適期數，故本研究是採取較精簡之模型進

行分析，詳細過程與圖示將會在第三章節 - 實證分析與結果中呈現。

3.8 模型評估指標

在此研究中，我使用平均絕對誤差(Mean Absolute Error, MAE)與均方根誤差(Root Mean Squared Error, RMSE)作為模型評估指標。

平均絕對誤差(Mean Absolute Error, MAE):

$$MAE = \frac{1}{T} \sum_{t=1}^T |Y_t - \hat{Y}_t|$$

其中，T 為樣本數， Y_t 為資料實際值，在此研究中，此部分設定為驗證集(validation set)之資料實際值。 \hat{Y}_t 為模型預測值。

根據公式，MAE 為觀察實際值與預測值之間絕對誤差的平均值，能夠反映預測值誤差的實際狀況，若 MAE 值越接近零，則說明模型預測誤差程度越低。

均方根誤差(Root Mean Squared Error, RMSE):

$$RMSE = \left[\frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 \right]^{\frac{1}{2}}$$

其中，T 為樣本數， Y_t 為資料實際值， \hat{Y}_t 為模型預測值。

根據公式，相較於 MAE 而言，RMSE 為將實際值與預測值之差距取平方後取平均值，再開平方根。可以發現到，因為將差距取平方，故會將預測誤差較大部分放大。若 RMSE 值越接近零，則說明模型預測誤差程度越低。

第四章 實證分析與結果

在此章節中，我將根據 1.2 小節研究架構與流程，進行實證分析，其中包含資料準備、資料處理、敘述統計分析、檢驗時間序列資料定態性、

ARIMA 模型期數判斷、設定資料訓練集與驗證集、模型擬合與預測分析，並使用 Ljung-Box test, Jarque-Bera test, White test 檢驗模型殘差之序列相關性、同質性與常態分配假設，最後再計算 MAE、RMSE，做為模型評估指標，用以評判模型之預測程度。

4.1 資料準備與處理

此研究的資料來自於美國聯邦儲備系統經濟資料庫(Federal Reserve Economic Data, FRED)之美國實質 GDP 季資料，時間範圍為 1947 年 1 月 1 日至 2024 年 7 月 1 日。其中，此經濟數據代號為 GDPC1，代表 Real Gross Domestic Product，即實質國內生產毛額(Real GDP)。

探索性資料分析(Exploratory Data Analysis, EDA)在資料科學領域中是相當重要的步驟，做好此步驟能夠讓研究者在起步時能辨識出資料的特徵與缺失值，以及其他可能會影響到後續分析的錯誤。準備好資料後，首先，我先將原始資料進行視覺化，並進行初步瞭解，透過圖像分析判斷在原始資料中是否存在明顯固定趨勢或者季節性。

在圖 1 中，我們可以清楚地發現到原始時間序列資料為非定態時間序列，即在給定不同時間 t 下，其時間序列 y_t 出現的範圍與自我相關性結構是不穩定的。

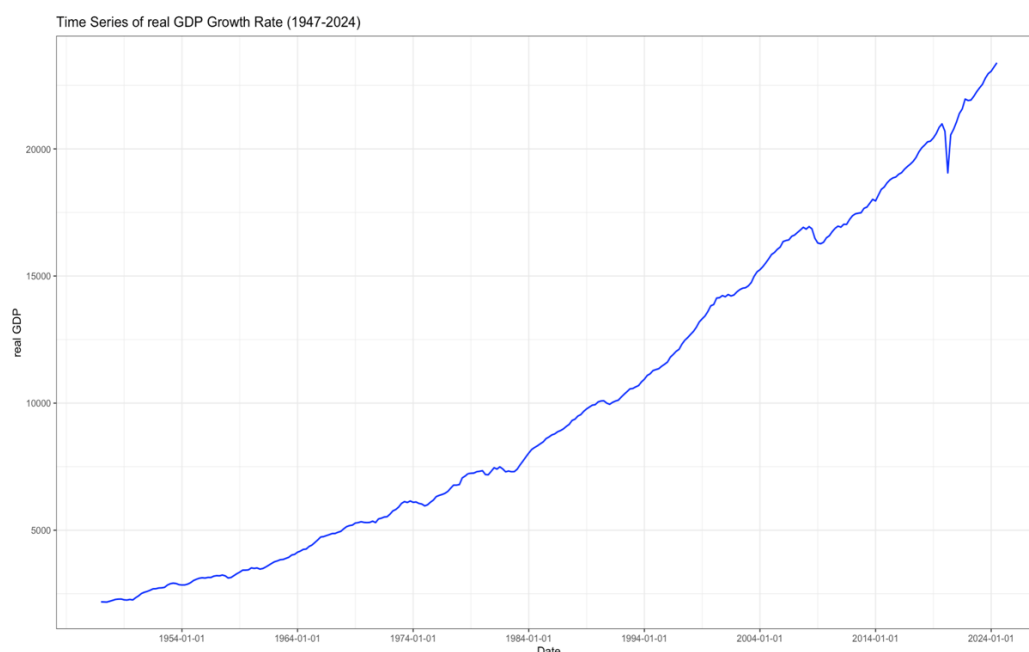


圖 1、美國實質 GDP 原始資料之時間序列圖

為了確保時間序列資料為定態序列，我們可以透過對非定態時間序列取自然對數與一階差分方式轉換為定態時間序列，並且可以進行 Augmented Dickey-Fuller test(以下稱 ADF test)來檢定時間序列資料是否具有單根(unit root)，若拒絕虛無假設，則代表被解釋變數 y_t 不具單根，此時為定態時間序列。

若時間序列資料為非定態，則需要透過取自然對數與一階差分方式轉換為定態時間序列，並透過 ADF test 來檢定時間序列資料是否具有單根(unit root)，若拒絕虛無假設，則代表被解釋變數 y_t 不具單根，此時為定態時間序列；若無法拒絕虛無假設，則代表被解釋變數 y_t 具單根，此時為非定態時間序列。

此外，針對定態時間序列，單純以時間序列圖而言，我們應該期待看見平均回復(Mean Reversion)，變異程度為常數(Constant Variance)、無固定趨勢(No Deterministic Trend)、無季節性(No seasonality)等特性。

其中，平均回復代表當定態時間序列在隨著時間增加的同時，會傾向出現在平均值附近。儘管有時會出現與平均值的偏差狀況，例如 2008 年的金融風暴或 2020 年的 Covid-19 疫情，但定態時間序列總是能夠回到平均值；變異程度為常數代表時間序列資料會出現波動，但此波動的振幅度能保持相對一致，代表隨著時間增加時，變異數為常數；無固定趨勢代表定態時間序列不會出現持續向上或者持續向下的趨勢；無季節性代表定態時間序列儘管會出現波動，但此波動不會是持續且固定的模式，即季節性。

在圖 2 中，我對美國實質 GDP 原始資料進行 ADF test，由於根據時間序列圖呈現，我們預期其應為非定態時間序列，故假設檢定應出現無法拒絕虛無假設(被解釋變數 y_t 具單根)，無證據支持對立假設(被解釋變數 y_t 不具單根)的情況，而根據 $P\text{-value} = 0.983$ 的結果，我們可以確定在顯著水準 $\alpha = 0.05$ 下，此時無法拒絕虛無假設，意即被解釋變數 y_t 具單根，此為非定態時間序列。

```
> adf.test(real.gdp)

Augmented Dickey-Fuller Test

data: real.gdp
Dickey-Fuller = -0.46679, Lag order = 6, p-value = 0.983
alternative hypothesis: stationary
```

圖 2、針對美國實質 GDP 原始資料(非定態時間序列)進行 ADF test

因此，接下來，我針對原始資料進行取自然對數與一階差分方式轉換為定態時間序列，並且畫出時間序列圖之後，再做一次 ADF test。在圖 3 中，可以發現到，此轉換後的時間序列在隨著時間增加的同時，會傾向出現在平均值附近。儘管有時出現與平均值的偏差狀況，例如 2008 年的金融風暴或 2020 年的 Covid-19 疫情衝擊，但此時間序列總是能回到平均值，即我們在前文說的均值回復(Mean Reversion)，但是很明顯的，此時間序列變異程度應不為常數，另外，根據此時間序列圖而言，我們也可以發現無固定趨勢(No Deterministic Trend)與無季節性(No seasonality)等特性。

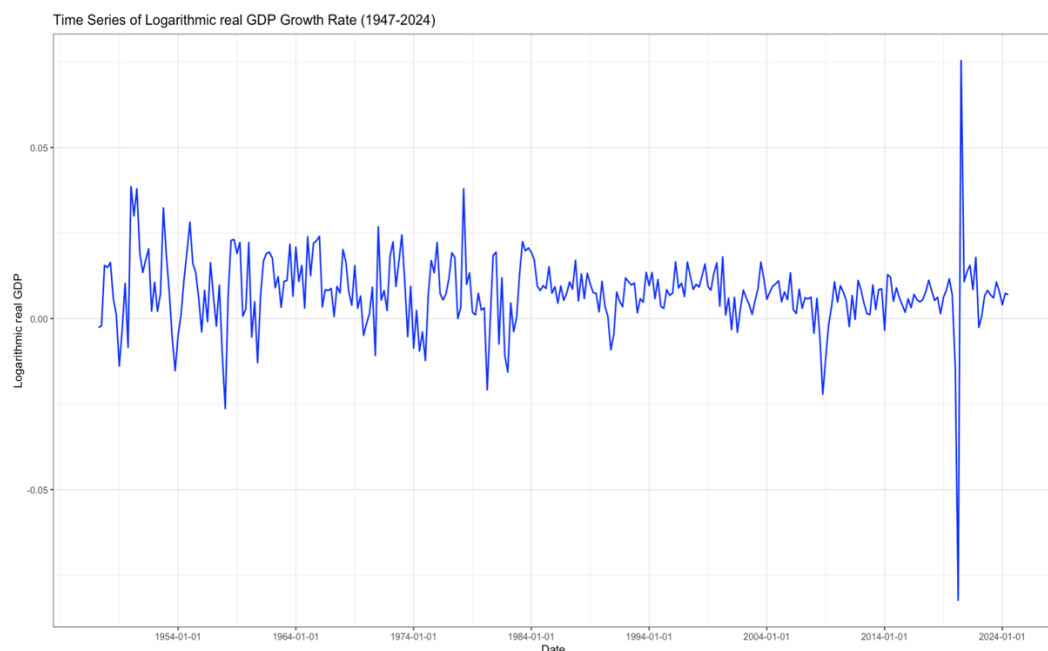


圖 3、差分後美國實質 GDP 資料之時間序列圖

為了確定我們在取自然對數與一階差分後的此時間序列究竟是否為定態時間序列，在圖 4 中，我便對差分後的美國實質 GDP 資料再做一次

ADF test，可以發現到，根據 $P\text{-value} = 0.01$ 的結果，我們可以確定在顯著水準 $\alpha = 0.05$ 下，此時拒絕虛無假設，有證據支持對立假設，意即被解釋變數 y_t 不具單根，此為定態時間序列。

```
> adf.test(log_real_gdp)

Augmented Dickey-Fuller Test

data: log_real_gdp
Dickey-Fuller = -7.3829, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

圖 4、針對差分後美國實質 GDP 資料(定態時間序列)進行 ADF test

4.2 敘述統計分析

現在，我們已經將美國實質 GDP 原始資料(非定態時間序列)轉換為差分後美國實質 GDP 資料(定態時間序列)，而在進入 ARIMA 模型期數的選擇之前，我們先針對原始資料與差分後資料去計算出敘述統計量，包含平均數(Mean)、標準差(Standard deviation)、偏態係數(Skewness)、峰態係數(Kurtosis)，這樣做的目的是為了在建立時間序列模型之前，我們能透過這些敘述統計量得到資料的特性資訊。

Table: Summary Statistics of Real GDP (Raw Data) and Log Real GDP (Processed Data)

Statistic	Real_GDP	Log_Real_GDP
Mean	10104.4461736	0.0076504
Standard Deviation	6176.0623272	0.0111694
Skewness	0.4649338	-1.0166491
Kurtosis	-1.0811220	17.0856277

圖 5、美國實質 GDP 原始資料與差分後美國實質 GDP 資料之敘述統計量

在圖 5 中，我將美國實質 GDP 原始資料與差分後美國實質 GDP 資料之敘述統計量透過 R 語言計算出來，並整理成表格。其中，針對原始資料之敘述統計量，我們可以發現在標準差部分，原始資料(非定態時間序列)非常地大；而差分後資料(定態時間序列)則相對地十分小。關於偏態係數(Skewness)部分，原始資料呈現右偏(Right-skewed)，代表有高於平均的成長率，我認為這反應了實質 GDP 在長期趨勢下呈現具上升趨勢的狀況，儘管仍然會發生經濟衰退事件，如 2008 年的金融危機與 2020 年的 Covid-19

全球疫情，但以長期而言，大多數國家的實質 GDP 仍會呈現上升趨勢；而差分後資料呈現左偏(Left-skewed)，代表有低於平均的成長率，我認為根據圖 5 的時間序列圖，可看出即使在進行資料差分轉換後，2020 年 Covid-19 的全球疫情影響仍十分巨大，應為差分後資料呈現左偏之主因。最後，關於峰態係數部分，原始資料呈現低闊峰(Platykurtic)，意即資料分配相當集中，而離群值(Outliers)出現的情況不多；相對來說，差分後資料呈現高狹峰(leptokurtic)，即資料具有厚尾(Heavy tail)特性，代表有較大機率出現離群值。

4.3 透過觀察法與資訊評選準則判斷 ARIMA 模型期數

4.3.1 自相關函數(ACF)與偏自相關函數(PACF)

若我們想判斷 ARIMA，或者 ARMA 模型之 p, q 期數，則最直觀與簡易的想法即透過觀察自相關函數(ACF)與偏自相關函數(PACF)圖形之呈現狀況進行判斷。

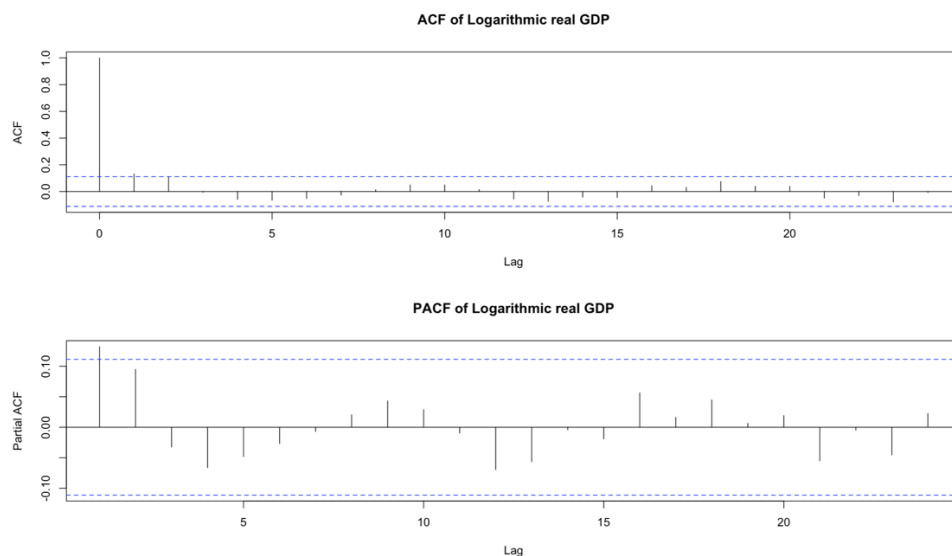


圖 6、差分後美國實質 GDP 資料之 ACF(上)和 PACF(下)圖

根據圖 6，我呈現了將美國實質 GDP 進行一階差分後的 ACF 與 PACF 圖表，我們可以發現在下方的 PACF 圖呈現在第一期後出現截斷現象，以及上方的 ACF 圖呈現之未顯著截斷現象，我們可以先大致猜測應適合使用 ARIMA(1,0,0)模型。

另外，在圖 7 的部分為美國實質 GDP 原始資料的 ACF 與 PACF 圖，

我們可以發現到，對於定態時間序列(如圖 6)而言，其 ACF 會快速地下降至零附近；而對於非定態時間序列(如圖 7)，其 ACF 則會相對緩慢地下降。

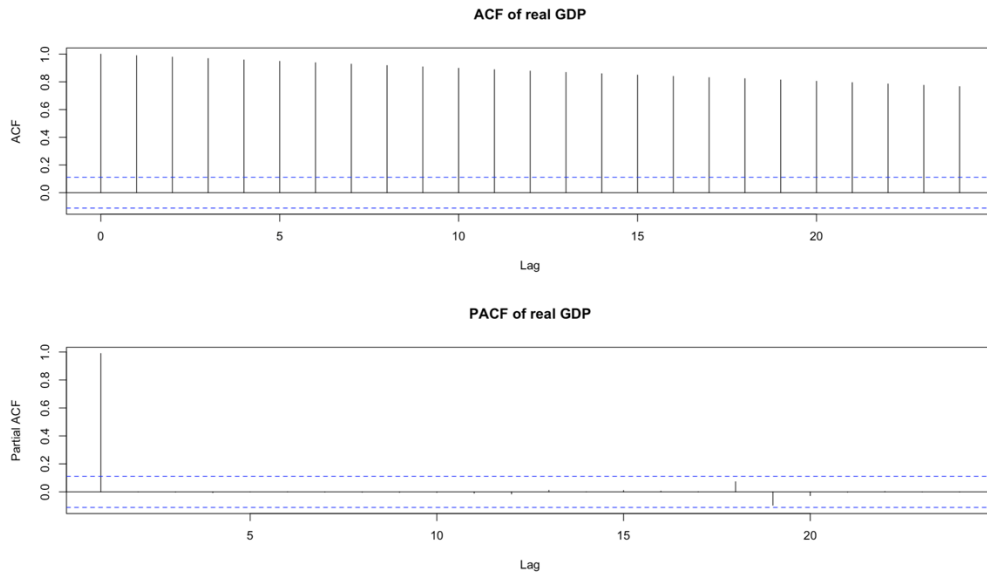


圖 7、美國實質 GDP 原始資料之 ACF(上)和 PACF(下)圖

4.3.2 赤池資訊評選準則(AIC)與貝氏資訊評選準則(BIC)

除了觀察 ACF 與 PACF 圖之外，我們還可以透過赤池資訊評選準則(AIC)與貝氏資訊評選準則(BIC)之值進行 ARIMA 或者 ARMA 模型之最適期數(p, q)判斷，當資訊評選準則越小，則該模型越佳，並且相較於 AIC，就兩者公式本身的意義而言，可見 BIC 給予的懲罰項較大，故相較於 AIC，BIC 會挑選較精簡的模型，意即參數較少之模型。

由於根據 4.3.1 小節中圖 6 的 ACF 與 PACF 圖呈現之資訊，我們已先行猜測應適合使用 ARIMA(1,0,0)模型，故接下來，我額外考慮了 ARIMA(2,0,0)、ARIMA(3,0,0)、ARIMA(1,0,1)、ARIMA(1,0,2)、ARIMA(1,0,3)、ARIMA(2,0,1)、ARIMA(2,0,2)、ARIMA(2,0,3)等模型，並透過 R 語言將上述考慮之模型針對訓練集資料(Training set)進行擬合(Fitting)後，再計算各模型的 AIC 與 BIC 值，其結果如圖 8，我們可以發現，若根據 AIC 準則判斷最適模型期數，則我們應該選擇 ARMA(2,2)；若根據 BIC 準則判斷最適模型期數，則我們應該選擇 ARMA(1,0)，此結果可證實在前一段之論述，確實 BIC 會傾向挑選較精簡的模型。最後，我選擇採用 BIC 準則來選取模型最適期數，即使用 ARMA(1,0)模型，故本研究是

採取較精簡之模型進行分析。

	Models	AIC	BIC
1	ARMA(1,0)	-1857.448	-1846.307
2	ARMA(2,0)	-1858.231	-1843.376
3	ARMA(3,0)	-1856.567	-1837.998
4	ARMA(1,1)	-1856.836	-1841.981
5	ARMA(1,2)	-1856.804	-1838.235
6	ARMA(1,3)	-1854.751	-1832.468
7	ARMA(2,1)	-1856.378	-1837.809
8	ARMA(2,2)	-1864.201	-1841.919
9	ARMA(2,3)	-1855.408	-1829.412

圖 8、計算不同 ARMA 模型的 AIC 與 BIC

```
> cat("Model with Minimum AIC:", min_aic_model, "\n")
Model with Minimum AIC: ARMA(2,2)
> cat("Model with Minimum BIC:", min_bic_model, "\n")
Model with Minimum BIC: ARMA(1,0)
```

圖 9、不同模型下之 AIC 與 BIC 值最小的模型

值得一提的是，在 R 語言中，我們可以透過使用 `auto.arima()` 函式讓程式自動選擇最適模型期數，而在此部分，我同樣是讓程式針對訓練集資料進行擬合後，再回傳結果。而根據圖 10，我們可以發現程式同樣也選擇 ARMA(1,0) 模型，與我們前一段根據 BIC 準則選取之模型相同。另外，若使用 Python 程式語言，也可以使用 `statsmodels` 套件的 `auto_arima()` 函式來達到相同的效果。

若在大樣本的情況下(如本研究樣本 $T = 310$)，根據 AIC 與 BIC 兩者本身公式而言，則應採取 BIC 準則建議的模型最適期數；若在小樣本情況下，則應採取 AIC 準則所建議的模型最適期數。

```

> arima_model <- auto.arima(train_data$log_real_GDP)
> arima_model
Series: train_data$log_real_GDP
ARIMA(1,0,0) with non-zero mean

Coefficients:
          ar1      mean
      0.1322  0.0077
s.e.  0.0569  0.0007

sigma^2 = 0.0001257:  log likelihood = 931.72
AIC=-1857.45   AICc=-1857.37   BIC=-1846.31

```

圖 10、透過 auto.arima()自動選擇最適模型

4.4 設定資料訓練集與驗證集

本研究中使用到的資料為美國實質 GDP 之時間序列資料，其時間範圍自 1947 年 1 月 1 日到 2024 年 7 月 1 日，且為季資料。關於訓練集資料(training set)，我設定其時間範圍自 1947 年 1 月 1 日至 2022 年 10 月 1 日；而驗證集資料(validation set)則設定自 2023 年 1 月 1 日至 2024 年 7 月 1 日。

4.5 模型擬合與預測分析

在決定好模型最適期數，並且我們也將差分後美國實質 GDP 資料分成訓練集與驗證集之後，我使用設定好之 ARMA(1,0)模型對訓練集資料進行擬合，再針對驗證集資料進行預測，然後畫出在驗證集資料中，ARMA(1,0)模型預測值與實際值之位置，以便於檢視模型預測成果。

將資料分成訓練集與驗證集的理由在於，我們希望模型能夠在熟悉的資料上進行訓練，並且後續在未知的資料上驗證其精確程度，以確保模型可靠性。

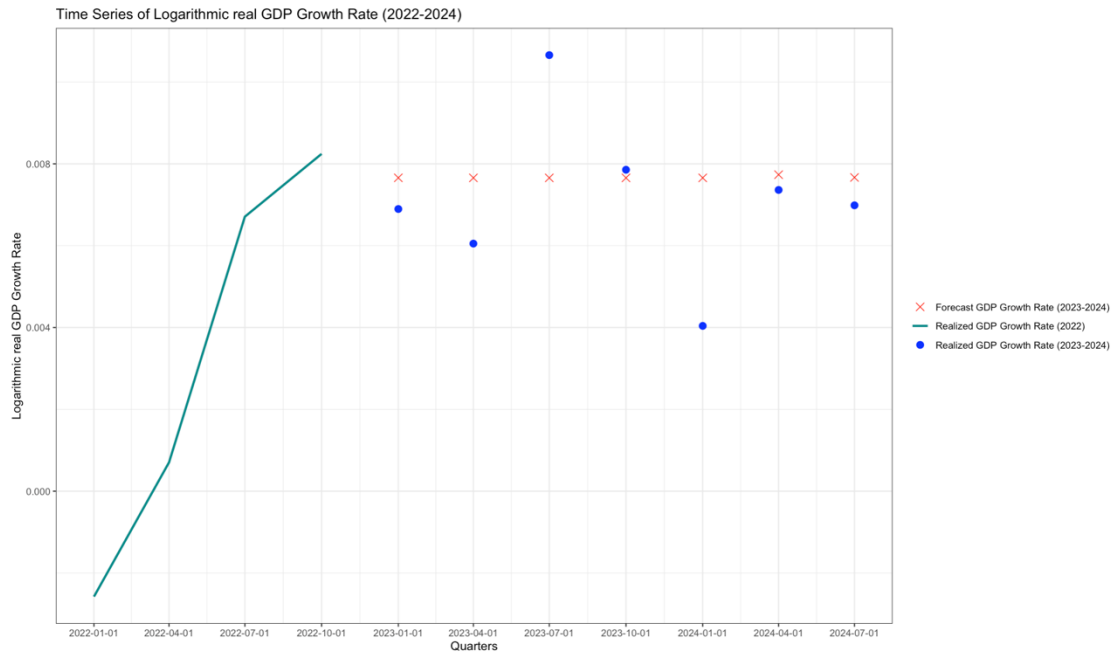


圖 11、ARMA(1,0)模型於驗證集資料之預測值與實際值時間序列圖

在圖 11 中，綠色折線為訓練集資料，代表美國實質 GDP 之實際走向；藍色圓點為驗證集資料，以點表示美國實質 GDP 於 2023 年 1 月 1 日至 2024 年 7 月 1 日之實際值；紅色叉叉為 ARMA(1,0)模型針對驗證集資料之預測值。

我們可以發現到，ARMA(1,0)所預測值較為穩定，考慮到模型為較精簡之模型，只考慮了一階自我迴歸項與常數項，代表我們僅拿前一期滯後期考慮對未來解釋變數的影響，因此也呈現在圖 11 所表現出的穩定預測值。若我們考慮針對較穩定的時間序列資料，則此模型能夠適用在此類資料上；但若我們考慮針對具劇烈波動的時間序列資料時，則此模型可能無法有效反應短期波動。

4.6 模型檢定分析與評估指標

接下來，我透過 R 語言進行 Ljung-Box test, Jarque-Bera test，以及透過 Python 程式語言進行異質性檢定(Heteroskedasticity test)，此外，再根據平均絕對誤差(Mean Absolute Error, MAE)與均方根誤差(Root mean Squared Error, RMSE)作為模型評估指標。

```
> ljung_box_test <- Box.test(residuals, lag = 17, type = "Ljung-Box")
> ljung_box_test

Box-Ljung test

data: residuals
X-squared = 10.905, df = 17, p-value = 0.8614
```

圖 12、Ljung-Box test 檢測模型殘差是否具序列相關性之結果

在圖 12 中，我透過 `residuals()` 將 ARMA(1,0) 模型的殘差提取出來，並進行 Ljung-Box test，而根據 $P\text{-value} = 0.8614 > \text{顯著水準 } \alpha = 0.05$ 之結果，代表無法拒絕虛無假設，無證據支持對立假設，意即 ARMA 模型之殘差無序列相關。理想上，我們會希望時間序列模型在能夠有效捕捉資料中的自我相關結構下，殘差不該存在固定趨勢。因此，此檢定結果對模型而言是一個好的指標。

```
> jarque.bera.test(residuals)

Jarque Bera Test

data: residuals
X-squared = 4866, df = 2, p-value < 2.2e-16
```

圖 13、Jarque-Bera test 檢測模型殘差是否服從常態分配之結果

在圖 13 中，我進行了 Jarque-Bera test，而根據 $P\text{-value} = 2.2e-16 < \text{顯著水準 } \alpha = 0.05$ 之結果，代表拒絕虛無假設，有證據支持對立假設，意即模型殘差不服從常態分配。

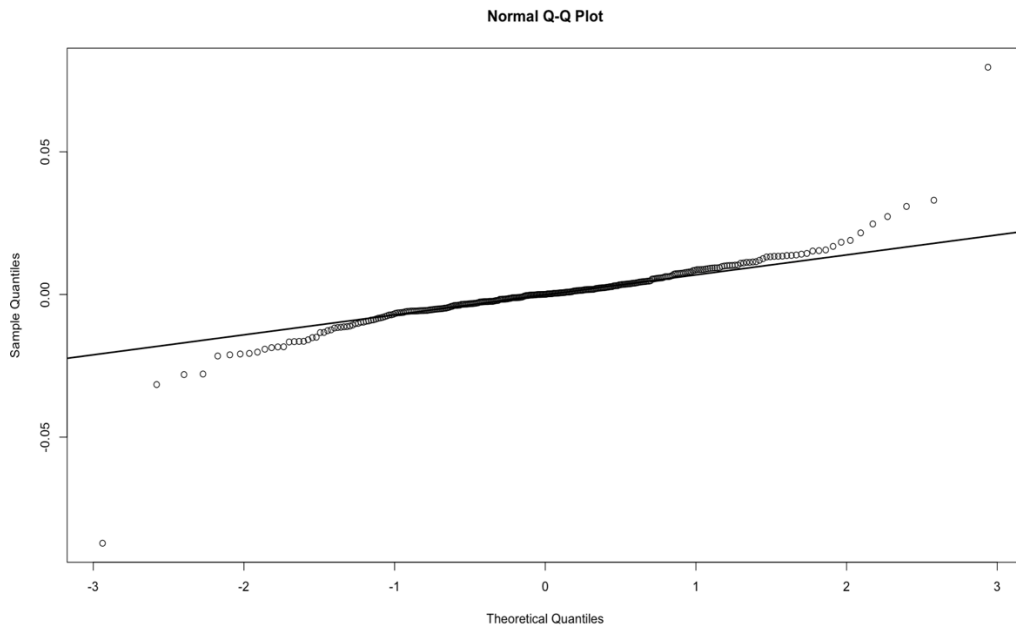


圖 14、針對模型殘差之 Normal Q-Q plot

在圖 14 中，我針對模型的殘差畫出常態分配(Normal)分位圖(Quantile-Quantile plot, Q-Q plot)，其中，X 軸為常態分配理論分佈值，Y 軸為樣本實際分佈值，可以發現到，在 45 度線之常態理論線周圍，存在離群值，且樣本一開始時累積速度慢，但因為資料皆聚集在中央，故到了樣本累積 $\geq 50\%$ 時，累積的速度變快，代表說此資料為高峽峰(leptokurtic)，具厚尾(heavy tail)特性。

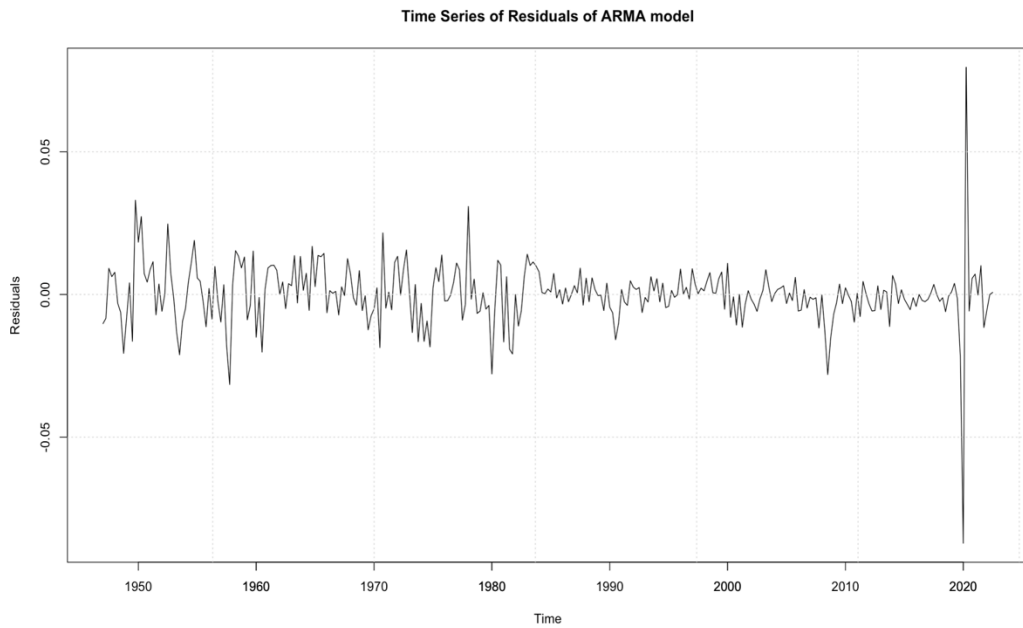


圖 15、模型殘差時間序列圖

在圖 15 中，我將模型的殘差畫出時間序列圖，以便於進行殘差分析，可以發現到，模型的殘差看起來皆落在 0 附近震盪，若我們認為時間序列模型能夠有效捕捉資料中的自我相關結構，則殘差負責的是無法解釋的部分，說明了以平均而言，此 ARMA(1,0)模型的預測是準確的，然而，針對 2020 年 Covid-19 的顯著偏差部分，可能表示模型對於外生變數的變動，並無法有效地捕捉此部分，但考慮到我們的模型僅考慮了一階自我迴歸項與常數項，模型在後續針對特定問題調整的需求仍然存在。

```

=====
SARIMAX Results
=====
Dep. Variable:      log_real_GDP      No. Observations:      303
Model:              ARIMA(1, 0, 0)    Log Likelihood          931.724
Date:               Fri, 20 Dec 2024  AIC                        -1857.448
Time:               05:17:47          BIC                     -1846.307
Sample:             04-01-1947        HQIC                    -1852.991
                  - 10-01-2022
Covariance Type:    opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
const          0.0077       0.001       8.310      0.000       0.006      0.009
ar.L1          0.1323       0.022       6.020      0.000       0.089      0.175
sigma2         0.0001    3.57e-06    34.960      0.000       0.000      0.000
=====
Ljung-Box (L1) (Q):      0.05      Jarque-Bera (JB):      4866.83
Prob(Q):                 0.82      Prob(JB):              0.00
Heteroskedasticity (H):  1.40      Skew:                  -0.40
Prob(H) (two-sided):     0.09      Kurtosis:              22.62
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
yangchen@ip30 Python_files %

```

圖 16、使用 Python 進行異質性檢定之結果

在圖 16 中，我針對異質性(Heteroskedasticity)檢定部分使用 Python 程式語言去實現，其中，針對 Heteroskedasticity test 部分，此檢定目的在檢測模型殘差項是否符合同質性假設，即 $Var(\epsilon_t) = \sigma^2$ ，而根據 Prob(H) = 0.09 結果，說明若在顯著水準 $\alpha = 0.05$ 下，則無法拒絕虛無假設，無證據支持對立假設，即殘差項符合同質性假設。

值得一提的是，根據圖 16 中，Ljung-Box test 的 Prob(Q) = 0.82，以及 Jarque-Bera test 的 Prob(JB) = 0 的結果，我們在使用 R 語言或者 Python 所得到的結果皆相同，即模型殘差不具序列相關性，且不服從常態分配。

```

Training set error measures:
              ME          RMSE          MAE          MPE          MAPE          MASE          ACF1
Training set 4.798875e-06 0.01117625 0.006963649 -139.4177 274.9235 0.7954509 -0.01256428
> accuracy(forecasted_values$mean, validation_data$log_real_GDP)
              ME          RMSE          MAE          MPE          MAPE
Test set -0.0005488619 0.001926398 0.001462316 -15.91158 24.67126

```

圖 17、模型評估指標 MAE, RMSE 結果

在圖 17 中，我分別計算出模型針對訓練集與驗證集之 MAE 與 RMSE 結果，可以發現到，在訓練集中，模型的 MAE = 0.007，RMSE = 0.0112；而在驗證集中，模型的 MAE = 0.0015，RMSE = 0.0019。代表模型在經過訓練集的擬合後，針對驗證集的預測程度提升。

第五章 結論及建議

本研究採用 ARMA(1,0)模型，針對美國自 1947 年 1 月 1 日至 2024 年 7 月 1 日之實質 GDP 資料季資料進行時間序列分析。在研究中，我們首先透過資料視覺化、探索性資料分析與敘述統計分析的方式，用以辨識時間序列資料是否存在特徵與是否具缺失值，以及其他可能會影響到後續研究分析的錯誤。此外，我們採用 MAE、RMSE 作為模型評估指標，根據研究結果，模型在經過訓練集的擬合後，應用在驗證集的預測程度提升，不過，由於 ARMA(1,0)模型僅考慮了一階自我迴歸項與常數項，因此對於較穩定的時間序列資料而言，則此能夠適用在此類資料上；但若考慮具劇烈波動的時間序列資料時，則此模型可能無法有效應用在此類資料上，因此，後續針對特定問題而改良與調整模型的需求仍存在。

事實上，根據圖 16 說明之模型統計檢驗結果，考慮到模型之一階自我迴歸係皆具有顯著 P-value，代表其在統計上是顯著的，並且根據 Ljung-Box test 之結果顯示，模型殘差不具序列相關性，因此，我認為模型對於本研究使用到之美國實質 GDP 資料進行預測是適當的，另外，關於模型殘差不服從常態分配之問題，我認為這取決於研究者對研究議題之應用。

第六章 參考文獻

一、 中文文獻

1. 簡均庭(2024)，「實質 GDP 的時間序列分析與預測」，碩士論文，元智大學資訊管理學系碩士班。
2. 鄭天德(2002)，「ARMA-TGARCH 模型之建立」，碩士論文，陽明交通大學經營管理研究所碩士班。

二、 英文文獻

1. Box, G., & Jenkins, G. M. (1976). "Analysis: Forecasting and Control". San Francisco, 1-171.
2. Miah, M. M., Tabassum, M., Rana, M. S. (2019). "Modelling and forecasting of GDP in Bangladesh: An ARIMA approach." *Journal of Mechanics of Continua*

and Mathematical Sciences 14(3): 150–166.

3. Dickey, D. A., & Fuller, W. A. (1981). "Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*" *Journal of the Econometric Society*, 1057-1072.
4. Nelson, Charles R., and Charles R. Plosser. "Trends and random walks in macroeconomic time series: some evidence and implications." *Journal of monetary economics* 10.2 (1982): 139-162.
5. Ljung, G. M., & Box, G. E. P. (1978). "On a measure of lack of fit in time series models." *Biometrika*, 65(2), 297–303.