



NVIDIA 股價分析

第四組

組員

410650617 吳紫柔

410650641 呂芊樺

410651151 羅恩雅

410651243 劉羿暉

目錄

第壹章 緒論	3
第一節 前言	3
第二節 研究目的與動機	3
第三節 研究方法	4
第貳章 原始資料分析	5
第一節 敘述統計	5
第二節 時間序列圖	6
第三節 ACF/PACF 圖	7
第參章 資料平穩化	8
第一節 一階差分	8
第二節 季節性差分	10
第肆章 配適 ARIMA 和 SARIMA 模型	12
第一節 殘差是否符合白噪音假設	12
第二節 比較 AIC 與 BIC	13
第三節 最佳配適模型	14
第伍章 預測與結論	16
第一節 預測	16
第二節 結論	17
第陸章 附錄	18
程式碼	18

第壹章 緒論

第一節 前言

在現代經濟的發展中，股票市場不僅僅是經濟活動中的一項重要指標，更影響著整個經濟體系的穩定。股票市場波動頻繁，其背後的複雜性涉及多方面因素相互交織，共同塑造了股票市場的獨特格局。在這樣的背景下，投資者如何預測和應對這些變化，成為了財務管理和投資決策中的重要課題。

隨著人工智慧（AI）技術的迅速發展，AI 晶片的需求量在近年來呈現顯著上升趨勢。2024 年，這一需求更是達到了新的高峰。AI 晶片被廣泛應用於各個領域，包括自動駕駛、智能家居、醫療診斷、語音識別等，驅動了市場對高性能計算和數據處理能力的迫切需求。這一趨勢不僅推動了相關技術的進一步發展，也使得 AI 晶片製造商的股價節節攀升。公司例如：輝達（NVIDIA）、AMD、Intel 等，在 AI 晶片領域的市場佔有率和技術優勢，使得它們成為投資者眼中的熱門選擇。

第二節 研究目的與動機

近年來掀起的 AI 熱潮，許多軟體與技術相繼出世，使 AI 晶片備受關注，輝達作為該市場的領軍企業，其股價表現成為投資者和研究者關注的焦點。出於好奇其過去股市及更好地理解 AI 行業的發展趨勢，本研究旨在觀察其過去十年的股價變化情形，並試圖找出可能影響股價變化的因素。並透過建立時間序列模型，對未來股價變化進行預測，希望能夠更好地理解 AI 行業的發展，以提供投資參考。

第三節 研究方法

於 Yahoo Finance 選取輝達 2014 年 5 月到 2024 年 4 月的月收盤價(共 120 筆)作為研究資料。隨後，把前 110 筆資料作為訓練集，剩餘 10 筆做為測試資料集，預測末 10 個月的股價。

首先，由訓練資料建立時間序列圖及 ACF、PACF 圖，並檢測其平穩性。接著，判斷是否要進行一階差分和季節性差分，並繪製時間序列圖及 ACF、PACF 圖確認模型平穩性。隨後，由圖選定 q, Q, p, P 範圍並建立 ARIMA 和 SARIMA 模型，列出殘差檢定符合白噪音假設的模型，並透過 AIC 和 BIC 選定最佳模型。最後，利用最佳模型對測試集資料進行預測。

第貳章 原始資料分析

第一節 敘述統計

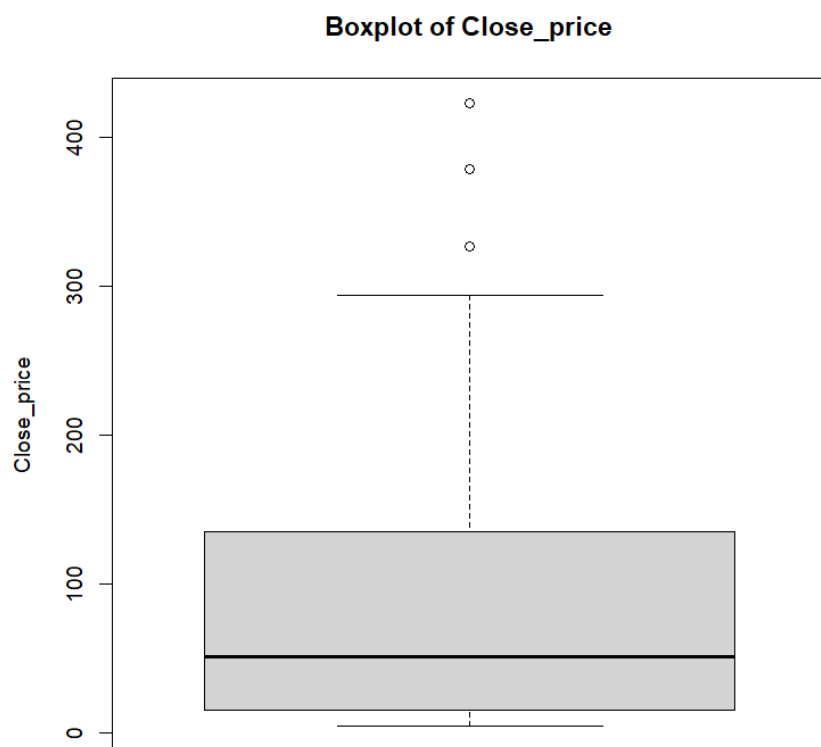


圖 2-1 盒形圖

表 2-1 敘述統計

統計值	股價(美元)
最小值	4.375
Q1	15.784
中位數	50.979
平均數	86.891
Q3	134.731
最大值	423.020

圖 2-1 和表 2-1，能明顯看出離群值存在，股價主要落在 15.784~ 134.731 美元區間，整體呈現明顯的右偏分布，價格變化較大。

第二節 時間序列圖

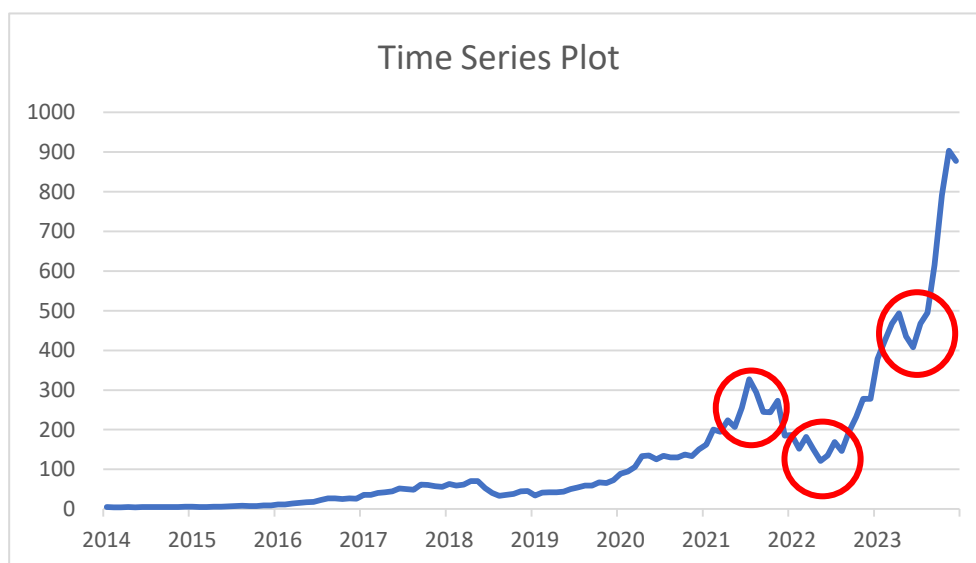


圖 2-2 原始資料的時間序列圖

由圖 2-2 可觀察到明顯的上升趨勢，因此初步判斷原始資料不平穩。

股市變化情形：

- 在新冠肺炎早期階段，輝達的股價像大多數公司一樣經歷了下跌。然而，由於疫情升溫對支援遠端工作、遊戲和資料中心技術的需求增加很快就逆轉上升。
- 2022 年股價下跌可歸因於持續的供應鏈中斷、美中貿易緊張局勢、先進半導體技術的出口限制及疫情後 PC 和遊戲硬體需求減少等原因。
- 2023 年，輝達的股價表現相當亮眼，但在 9 月份下滑，市值蒸發了近 1800 億美元。這主要受到「九月效應」的影響。九月效應是一個歷史性現象，指的是每年 9 月股市通常會表現不佳，這可能是由於基金公司在此期間結算年度，投資者進行投資組合調整，夏季交易量下降後的波動性增加等因素所致。在上半年表現非常強勁的情況下，9 月份股市下降顯得尤為突出。

第三節 ACF/PACF 圖

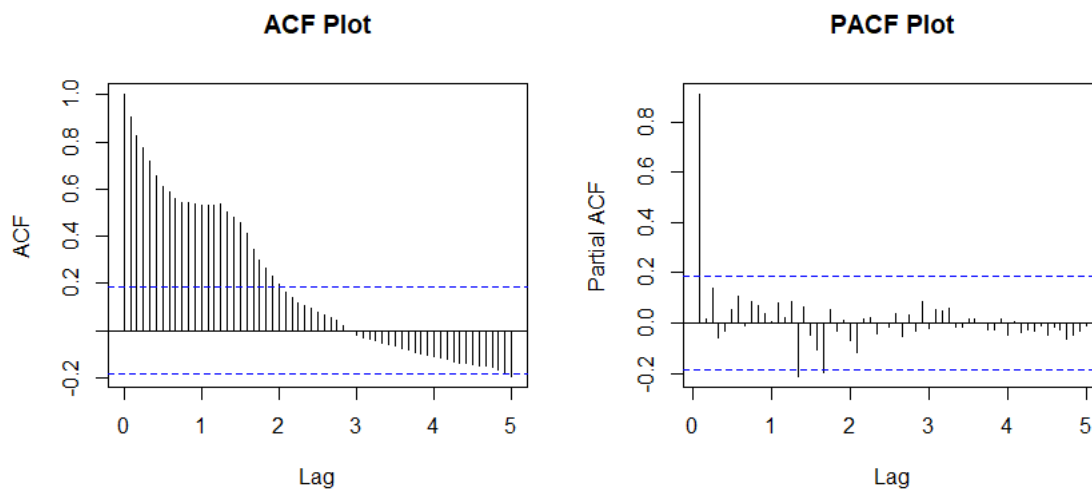


圖 2-3 原始資料的 ACF 和 PACF 圖

由圖 2-3 可見，原始資料的 ACF 圖自相關係數下降緩慢，因此我們推斷資料不平穩。隨後，進行單根檢定再次確認。

單根檢定：

$H_0: \phi = 1$ ，即資料不平穩

$H_1: |\phi| < 1$ ，即資料平穩

R 程式結果如下：

Augmented Dickey-Fuller Test

k=0

Dickey-Fuller = -0.30256

Lag order = 0

p-value = 0.9891

alternative hypothesis: stationary

Phillips-Perron Unit Root Test

Dickey-Fuller Z(alpha) = -6.4525

Truncation lag parameter = 4

p-value = 0.7424

alternative hypothesis: stationary

兩檢定結果 P-value 皆大於 0.05，故不拒絕虛無假設，代表資料不平穩，需進行差分。

第參章 資料平穩化

第一節 一階差分

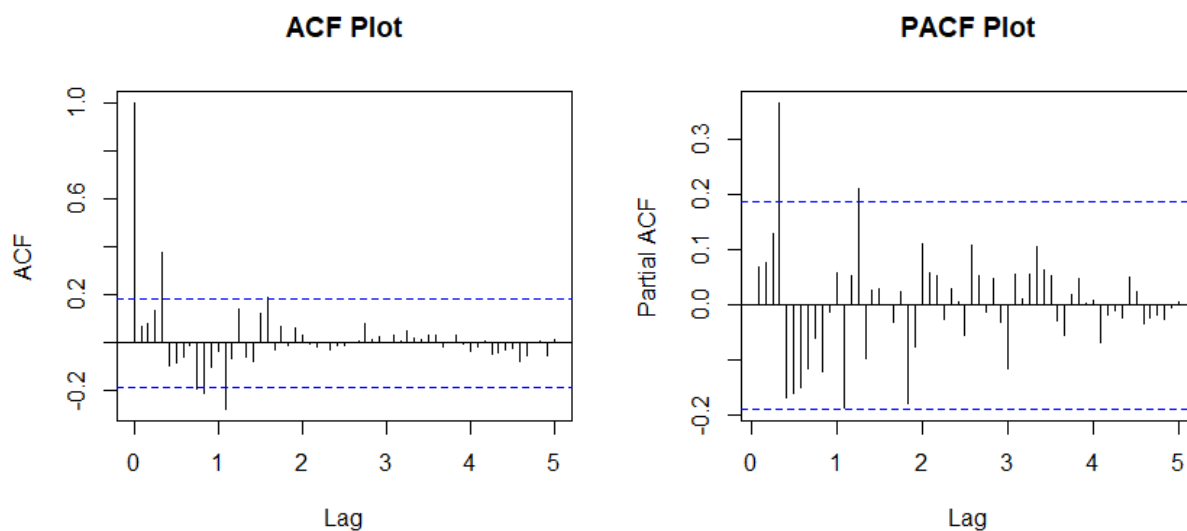


圖 3-1 一階差分後的 ACF 和 PACF 圖

由圖 3-1 可見，一階差分後的 ACF 圖自相關係數下降速度快了許多，因此我們推測資料可能已呈現平穩。

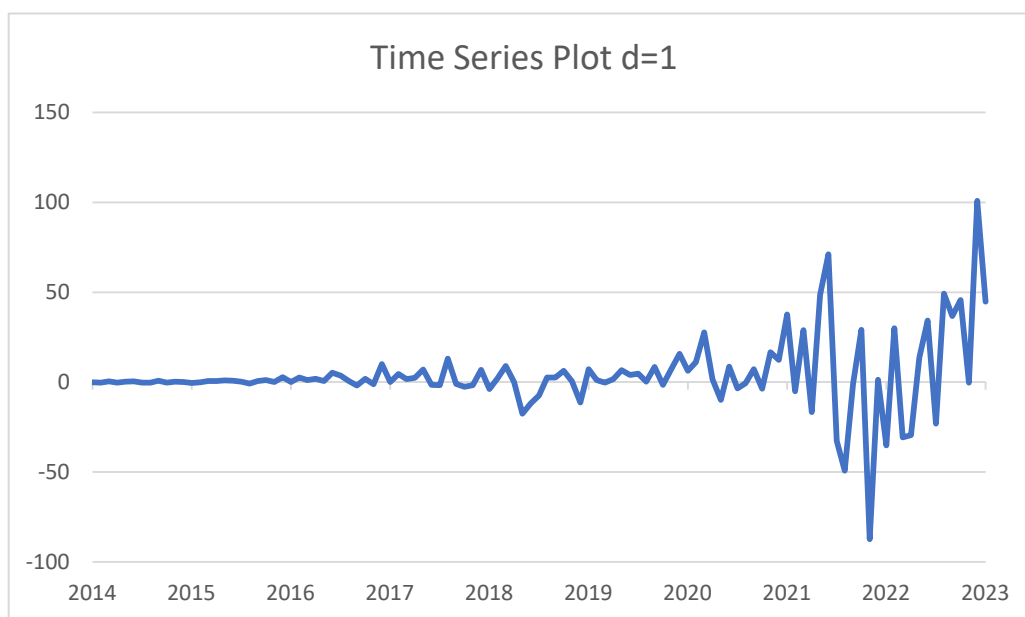


圖 3-2 一階差分後的時間序列圖

我們將原始資料的時間數列圖(圖 2-1)與一階差分後的时间數列圖(圖 3-2)進行比較，可觀察到原始資料的向上趨勢已獲得很好的改善，但差分後時間序列圖在 2021 年後仍有比較大的起伏，可能存在變異數異質性問題，因此我們進行單根檢定，更精確的確定資料是否平穩。

單根檢定：

$H_0: \phi = 1$ ，即資料不平穩

$H_1: |\phi| < 1$ ，即資料平穩

R 程式結果如下：

Augmented Dickey-Fuller Test

k=0

Dickey-Fuller = -9.7157

Lag order = 0

p-value = 0.01

alternative hypothesis: stationary

Phillips-Perron Unit Root Test

Dickey-Fuller Z(alpha) = -120.76

Truncation lag parameter = 4

p-value = 0.01

alternative hypothesis: stationary

兩檢定結果 P-value 皆為 0.01 小於 0.05，故拒絕虛無假設，代表資料已平穩。

第二節 季節性差分

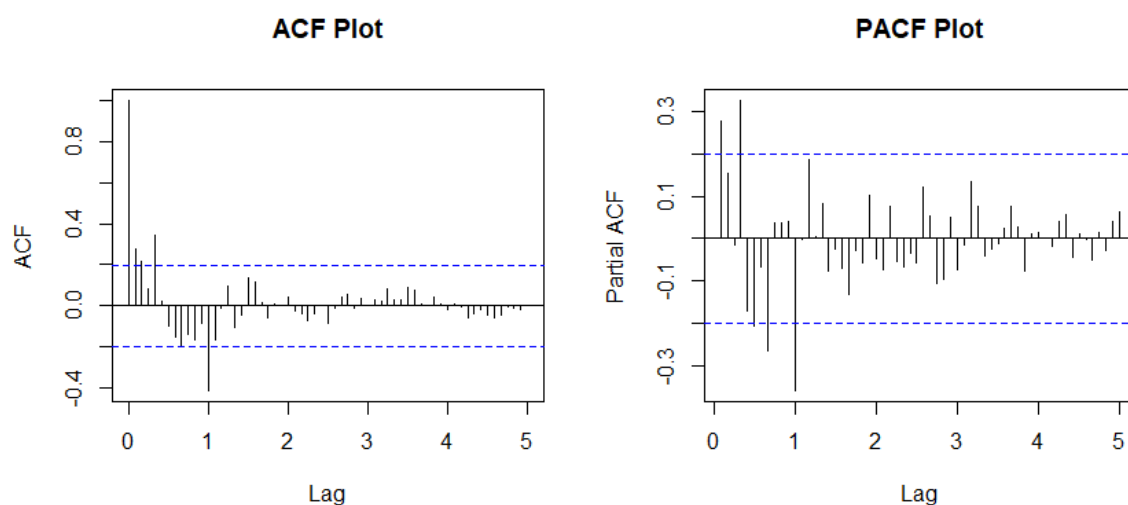


圖 3-3 季節性差分後的 ACF 和 PACF 圖

若再做季節性差分，由圖 3-3 發現其結果與一階差分後差不多，ACF 和 PACF 收斂情形也未使模型更簡化，且因為一階差分後資料已平穩，因此我們認為不需要再做季節性差分。

觀察一階差分後的 ACF 圖與 PACF 圖選擇 q , Q , p , P 之範圍：

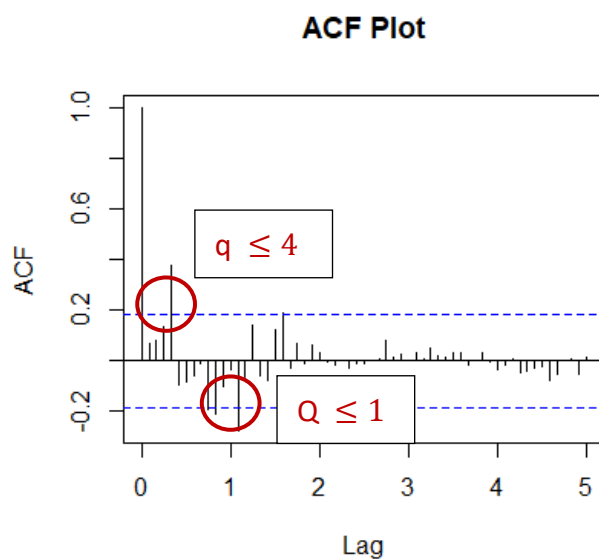


圖 3-4 一階差分後的 ACF 圖

由圖 3-4 可以看出在 $q=4$ 時，ACF 超過 $\pm \frac{1.96}{\sqrt{110}} \approx \pm 0.1869$ 的區間，並且在 lag 1 後切斷。因此判斷 $q \leq 4$ 、 $Q \leq 1$ 。

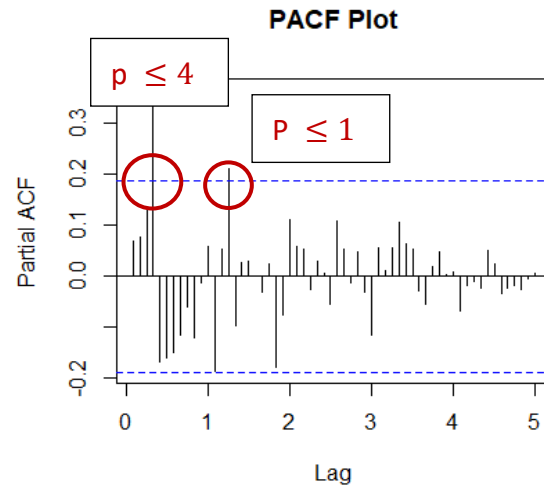


圖 3-5 一階差分後的 PACF 圖

由圖 3-5 可以看出在 $p=4$ 時，PACF 超過 $\pm \frac{1.96}{\sqrt{110}} \approx \pm 0.1869$ 的區間，並且在 lag 1 後切斷。因此判斷 $p \leq 4$ 、 $P \leq 1$ 。共有 $(4+1)(1+1) \times (4+1)(1+1) = 100$ 組模型需配適。

第肆章 配適 ARIMA 和 SARIMA 模型

第一節 殘差是否符合白噪音假設

判斷依據：

(一) 標準化殘差圖

殘差在正負之間隨機性跳動且無明顯趨勢。

(二) ACF 殘差圖

在大樣本性質下，資料的時間數列為純隨機，滿足白噪音條件：

$$1. E(W_t) = 0$$

$$2. \text{Var}(W_t) = \sigma^2, \quad \forall t$$

$$3. \text{Cov}(W_t, W_s) = 0, \quad \forall t \neq s$$

則觀察期數為 n 的時間數列，其 ACF 殘差值 $\hat{\rho}_e(h) \sim N(0, \frac{1}{n})$ 。

本研究訓練樣本 $n=110$ ，故由 ACF 殘差圖判斷時，需落在 $\pm \frac{1.96}{\sqrt{110}} \approx \pm 0.1869$

的區間內。

(三) Ljung-Box-Pierce 檢定

H_0 : 資料是獨立分佈，即沒有自相關性。

H_1 : 資料不是獨立分佈，即有自相關性。

檢定統計量為：

$$Q = n(n+2) \sum_{h=1}^H \frac{\hat{\rho}_e(h)}{n-h}$$

Q 統計量近似卡方分配。若 $P\text{-value} \geq 0.05$ 則不拒絕虛無假設，即判定殘差沒有自相關性，符合白噪音的殘差假設。

經上述判斷，100 組模型中，符合白噪音假設的模型共有 38 組；不符合白噪音假設的模型共有 62 組。

第二節 比較 AIC 與 BIC

列出符合白噪音假設的 38 組模型，對其 AIC、BIC 進行比較：

表 4-1 符合噪音假設的模型

模型	AIC	BIC	q+p+Q+P
ARIMA(0, 1, 4)	954.8337	968.2904	4
ARIMA(1, 1, 3)	963.3874	976.8442	4
ARIMA(1, 1, 4)	955.9997	972.1478	5
ARIMA(2, 1, 4)	957.4852	976.3246	6
ARIMA(3, 1, 2)	963.9111	980.0592	5
ARIMA(3, 1, 4)	956.1952	977.726	7
ARIMA(4, 1, 0)	955.7811	969.2378	4
ARIMA(4, 1, 1)	956.6984	972.8465	5
ARIMA(4, 1, 2)	957.7286	976.568	6
ARIMA(4, 1, 3)	957.7478	979.2786	7
ARIMA(4, 1, 4)	956.4612	980.6833	8
SARIMA(0, 1, 4) × (0, 0, 1)₁₂	956.8263	972.9744	5
SARIMA(1, 1, 4) × (0, 0, 1)₁₂	957.9814	976.8208	6
SARIMA(2, 1, 4) × (0, 0, 1)₁₂	959.4789	981.0097	7
SARIMA(3, 1, 2) × (0, 0, 1)₁₂	965.8042	984.6437	6
SARIMA(3, 1, 4) × (0, 0, 1)₁₂	958.0483	982.2704	8
SARIMA(4, 1, 0) × (0, 0, 1)₁₂	957.6392	973.7873	5
SARIMA(4, 1, 1) × (0, 0, 1)₁₂	958.6036	977.443	6
SARIMA(4, 1, 2) × (0, 0, 1)₁₂	959.6883	981.219	7
SARIMA(4, 1, 3) × (0, 0, 1)₁₂	959.6531	983.8752	8
SARIMA(4, 1, 4) × (0, 0, 1)₁₂	958.3801	985.2936	9

SARIMA(0, 1, 4) × (1, 0, 0)₁₂	956.8241	972.9722	4
SARIMA(1, 1, 4) × (1, 0, 0)₁₂	957.9762	976.8157	5
SARIMA(2, 1, 4) × (1, 0, 0)₁₂	959.4769	981.0077	6
SARIMA(3, 1, 2) × (1, 0, 0)₁₂	965.7867	984.6261	5
SARIMA(3, 1, 4) × (1, 0, 0)₁₂	958.0005	982.2226	7
SARIMA(4, 1, 0) × (1, 0, 0)₁₂	957.6122	973.7603	4
SARIMA(4, 1, 1) × (1, 0, 0)₁₂	958.5849	977.4243	5
SARIMA(4, 1, 2) × (1, 0, 0)₁₂	959.679	981.2097	6
SARIMA(4, 1, 3) × (1, 0, 0)₁₂	959.6334	983.8555	7
SARIMA(4, 1, 4) × (1, 0, 0)₁₂	958.3421	985.2556	9
SARIMA(0, 1, 0) × (1, 0, 1)₁₂	981.3532	989.4273	2
SARIMA(0, 1, 3) × (1, 0, 1)₁₂	977.4472	993.5953	5
SARIMA(1, 1, 0) × (1, 0, 1)₁₂	959.0751	977.9145	3
SARIMA(1, 1, 3) × (1, 0, 1)₁₂	961.019	987.9325	6
SARIMA(3, 1, 1) × (1, 0, 1)₁₂	977.5183	996.3577	6
SARIMA(4, 1, 1) × (1, 0, 1)₁₂	960.061	981.5918	7
SARIMA(4, 1, 2) × (1, 0, 1)₁₂	961.1218	985.3439	7

由表 4-1 可以看出 ARIMA(0,1,4)模型 AIC =954.8337 和 BIC =968.2904 皆為符合白噪音假設模型組合中 AIC、BIC 最小值，且其模型複雜度較低，故選此做為最佳模型。

第三節 最佳配適模型

由 R 程式結果，最佳模型為：

$$(1-B)X_t = (1 + 0.087B_1 + 0.0949B_2 + 0.1562B_3 + 0.5043B_4)W_t, W_t \sim WN(0, 336.6)$$

表 4-2 最佳配適模型

MA(p)	MA(1)	MA(2)	MA(3)	MA(4)
θ_p	0.0870	0.0949	-0.1562	0.5043
SE(標準誤)	0.0897	0.0874	0.0813	0.0844
$\hat{\sigma}^2 = 336.6$ $\log \text{likelihood} = -472.42$				

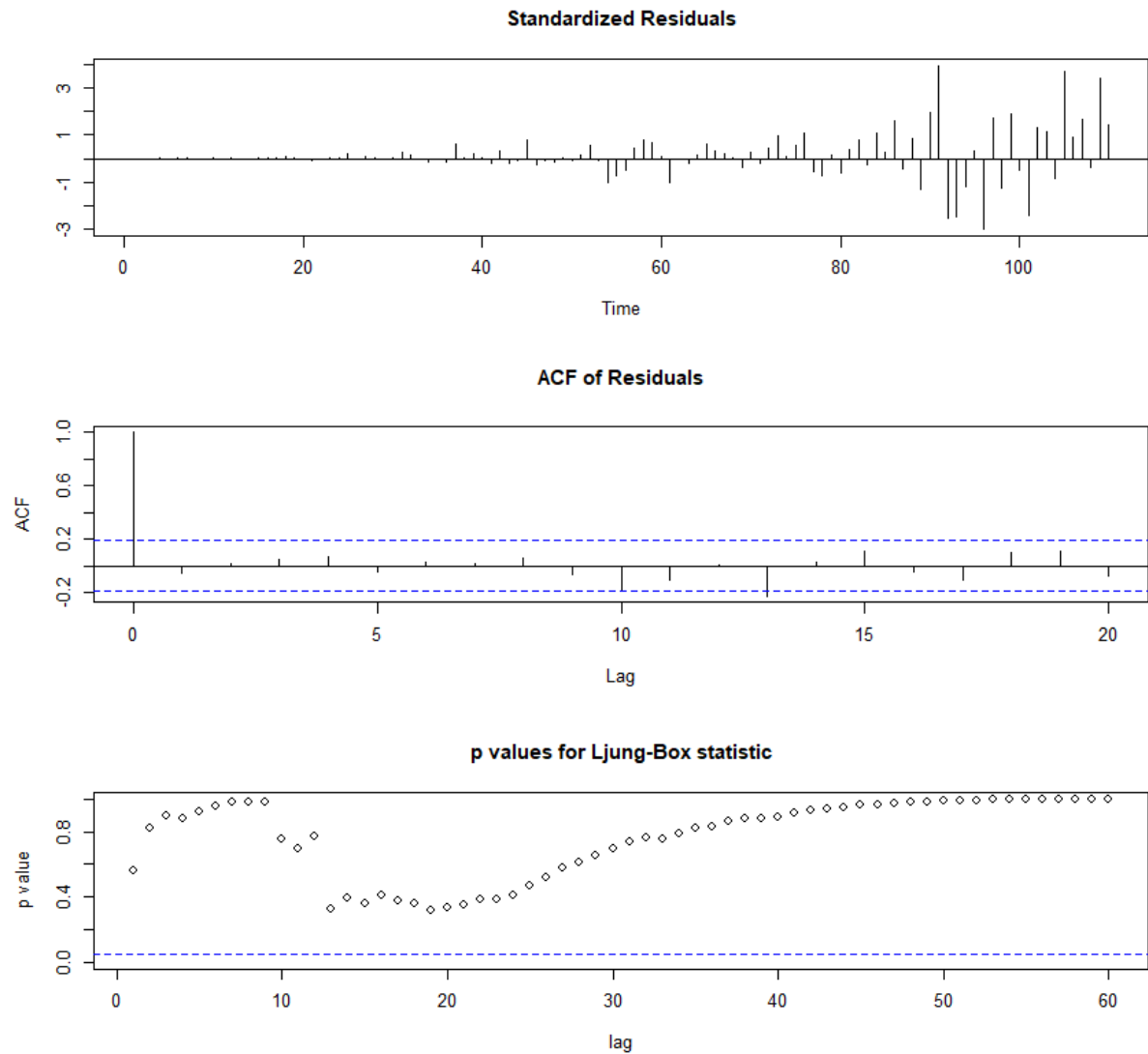


圖 4-1 最佳模型殘差檢定圖

由圖 4-1，最佳模型的判斷標準化殘差值在 0 附近隨機波動且無明顯趨勢，但有少數超過 ± 2 的離群值存在；ACF 殘差圖的自相關係數皆在區間內波動，表示無顯著的自相關；Ljung-Box-Pirece 檢定 P-value 皆大於 0.05，表示不能拒絕白噪音假設。

第五章 預測與結論

第一節 預測

最佳模型：

$$(1-B)X_t = (1 + 0.087B_1 + 0.0949B_2 + 0.1562B_3 + 0.5043B_4)W_t, W_t \sim WN(0, 336.6)$$

為測試此模型的準確性，選取研究資料的末 10 筆資料建立測試集，預測 2023 年 7 月到 2024 年 4 月的股價，以檢測此模型的預測能力。

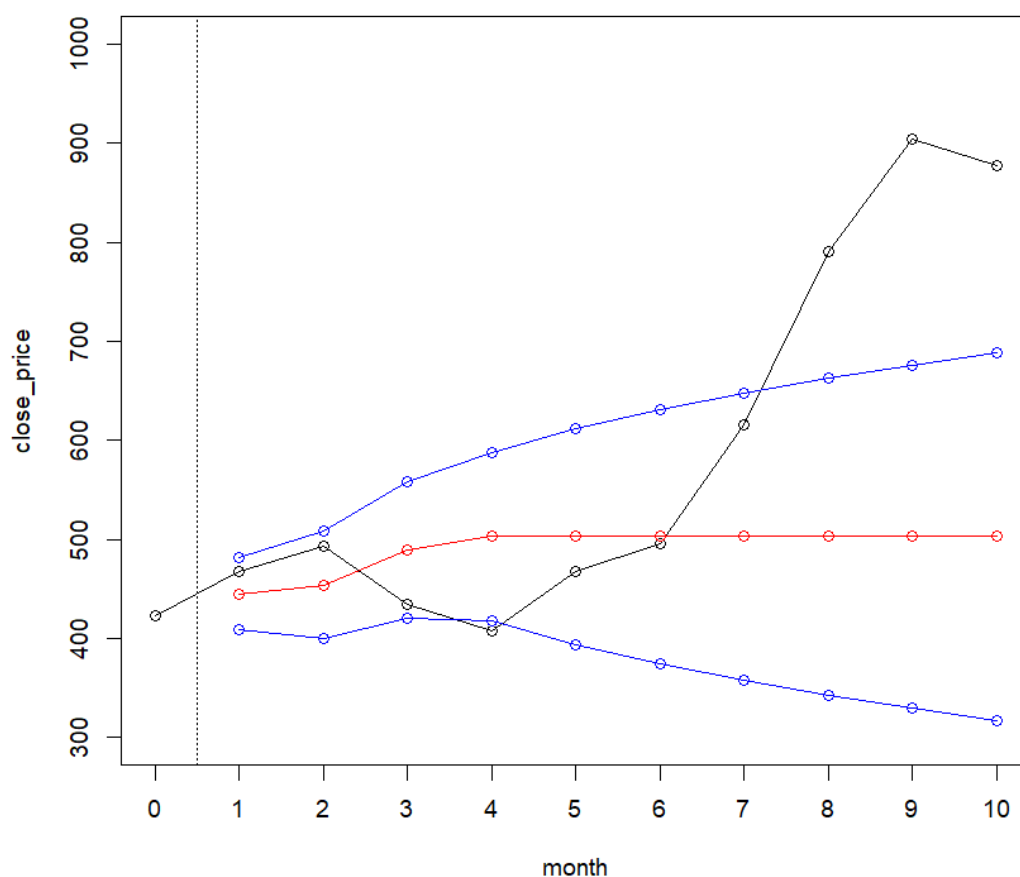


圖 5-1 預測圖

由圖 5-1 可知黑色線條為測試集資料，紅色為預測值，由此判斷近期預測有在 95%信賴區間內，尤其在第六個月時重疊，預測最佳。然而，在第八個月開始，預測的準確度開始下降，這可能是由於長期預測存在更多的不確定性和難以預測的因素，例如外部環境的變化、市場波動等。

第二節 結論

隨科技進步人們對 AI 晶片的需求增加，尤其近兩年來顯著上升與輝達股市持續飆升之間可能存在關聯性。由圖 5-2 可見過去一年間股價翻了將近三倍，且研究截止前(6 月 5 日收盤價：1224.4)再次突破新高。由於股價變動速度快，欲做短期預測用近期的日資料可能更準確；而長期預測則很難精準。

雖然人們普遍看好 AI 市場前景，但也有專家推測儘管 AI 熱潮還會延燒一陣子，多家新創公司已漸漸開始轉變方向，對輝達 GPU 的需求可能降低，所以欲預測股市長期走勢仍然存在風險和不確定性。股市可能會受到多種因素的影響，包括：經濟數據和指標、企業財報、政府政策和法規變動、國際事件等，因此單憑過去的資料預測未來的股市走勢並不是完全可靠的，所以在做出任何投資決策之前，需追蹤市場動態並考慮多方面的因素，以降低投資風險。

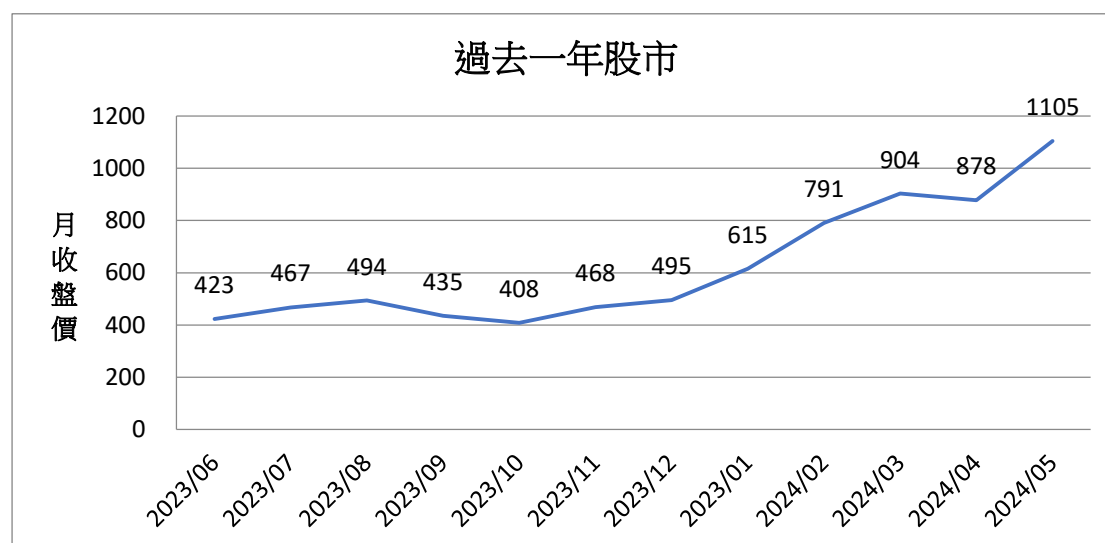


圖 5-2 過去一年股市

第陸章 附錄

程式碼

```
pacman::p_load(readxl, tidyverse, astsa, tseries)
data<-read_excel("C:\\Users\\niuno\\Desktop\\TS\\第四組資料.xlsx")
n=nrow(data)
data.t=data[1:(n-10),2]
summary(data.t)
boxplot(data.t$close, varwidth = TRUE,
        main = "Boxplot of Close_price",
        ylab = "Close_price")

#原始資料時間數列圖
data.temp <- ts(data.t, start = c(2014, 5), frequency = 12)
par(mfrow=c(1,1))
plot(data.temp, main = "Time Series Plot", ylab = "Close Price", xlab = "Time")

#畫 ACF 與 PACF 圖
par(mfrow=c(2,2))
acf(data.temp,60,main="ACF Plot")
pacf(data.temp,60,main="PACF Plot")
#單根檢定
adf.test(data.temp,k=0)
pp.test(data.temp)

#一階差分
data.d1 <- diff(data.temp)
#library(openxlsx)
#write.xlsx(data.d1, file = "C:\\Users\\niuno\\Desktop\\TS\\data.d1.xlsx",
#rowNames= F)
par(mfrow=c(1,1))
plot(data.d1, main = "Time Series Plot, d=1", ylab = "Close Price", xlab = "Time")
par(mfrow=c(2,2))
acf(data.d1,60,main="ACF Plot")
pacf(data.d1,60,main="PACF Plot")
#單根檢定
```

```

adf.test(data.d1,k=0)
pp.test(data.d1)

#季節性差分
data.s1 <- diff(data.d1,12)
par(mfrow=c(1,1))
plot(data.s1, main = "Time Series Plot, D=1", ylab = "Close Price", xlab = "Time")
par(mfrow=c(2,2))
acf(data.s1,60,main="ACF Plot")
pacf(data.s1,60,main="PACF Plot")
#單根檢定
adf.test(data.s1,k=0)
pp.test(data.s1)

#配適 ARIMA SARIMA 模型
data.fit=arima(data.t,order=c(0,1,4), seasonal = list(order = c(0, 0, 0), period=12))
data.fit
names(data.fit)
data.fit$coef
data.fit$sigma2
data.fit$aic
#檢驗配適後之殘差項
tsdiag(data.fit,60)

#求 AIC 與 BIC
AIC(data.fit)
BIC(data.fit)

#預測
data.pre <- predict(data.fit,n.ahead=10)
par(mfrow=c(1,1))
U<-data.pre$pred+2*data.pre$se
L<-data.pre$pred-2*data.pre$se
month<-110:120
x_labels <- 0:10
plot(month,data[month,]$Close,type="o",ylim=c(300,1000),ylab="production")
lines(data.pre$pred,col="red",type="o")
lines(U,col="blue",type="o")

```

```
lines(L,col="blue",type="o")
abline(v=110.5,lty="dotted")

# 增加底部 1 邊距
par(mar = c(5, 4, 4, 2) + 0.1)
plot(month, data[month, ]$Close, type = "o", ylim = c(300, 1000), ylab = "production",
xaxt = "n")
lines(data.pre$pred, col = "red", type = "o")
lines(U, col = "blue", type = "o")
lines(L, col = "blue", type = "o")
abline(v = 110.5, lty = "dotted")
axis(1, at = month, labels = x_labels, las = 1)
```