統計諮詢 - 作業 8

國立成功大學統計學系暨數據科學研究所

廖傑恩（RE6094028）

2021-06-11

1 Exercise 18.4

1.1 問題敘述

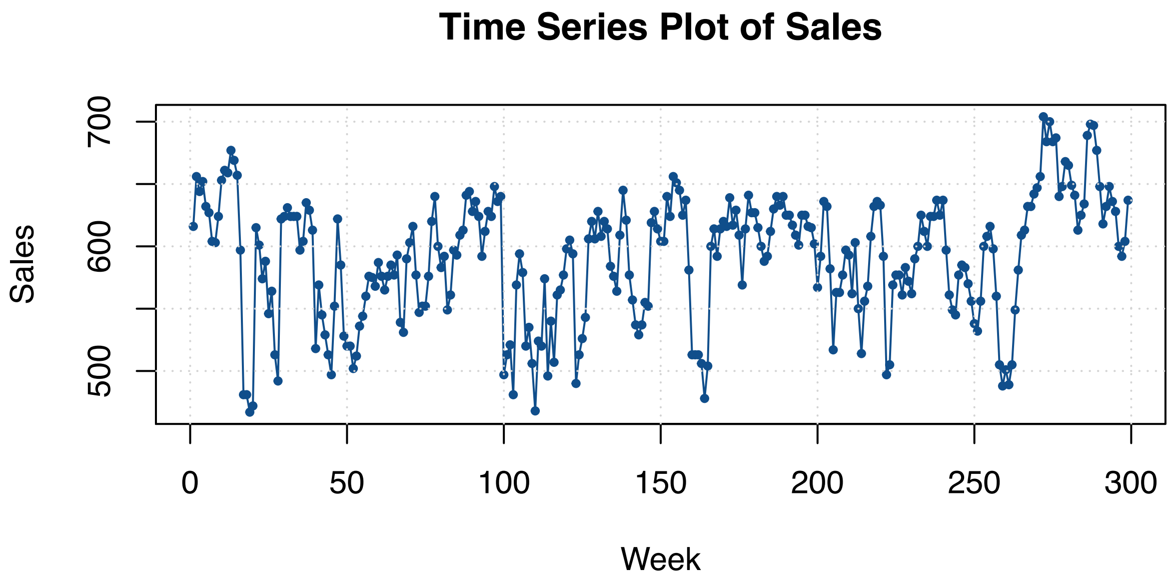
諮詢者想探討美國一種用於包裝藥品的塑膠容器，其銷售量隨著時間變化的趨勢。諮詢者也想了解為什麼無季節性的整合移動平均自迴歸模型（Auto-regressive Integrated Moving Average model, ARIMA）$ARIMA(p, 1, q)$ 不適用於這份資料。

1.2 資料介紹

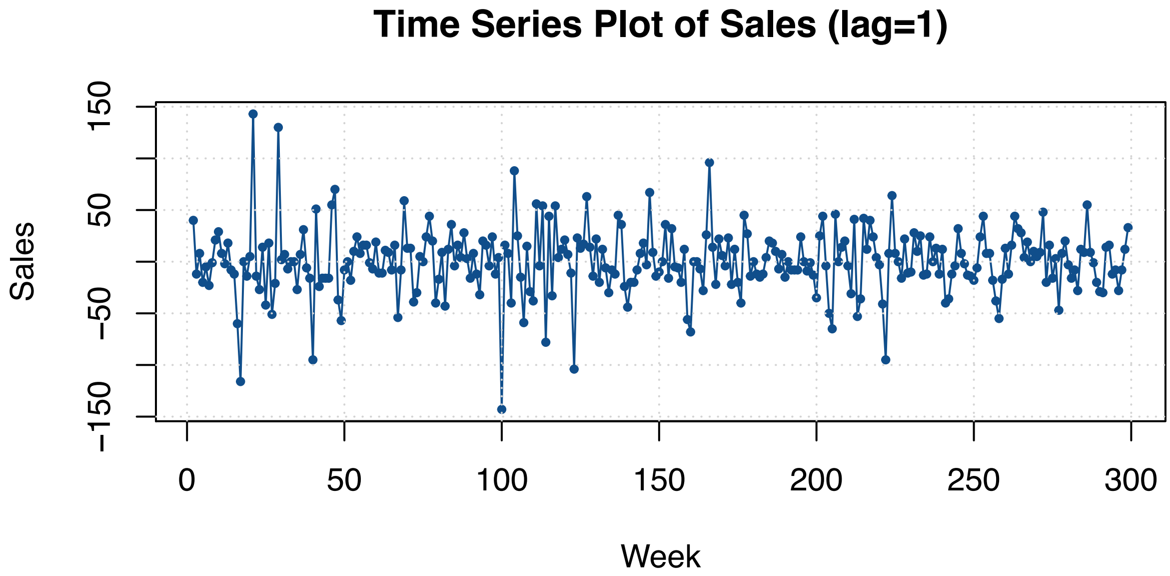
資料集原本由 Nicholls（1979）收集，Hand 等人（1994）後來重製。資料收集了美國一種用於包裝藥品的塑膠容器每週銷售量。資料有 299 列、行，每 1 列為一週銷售量，也就是說共有 299 週的銷售量。

1.3 資料探索

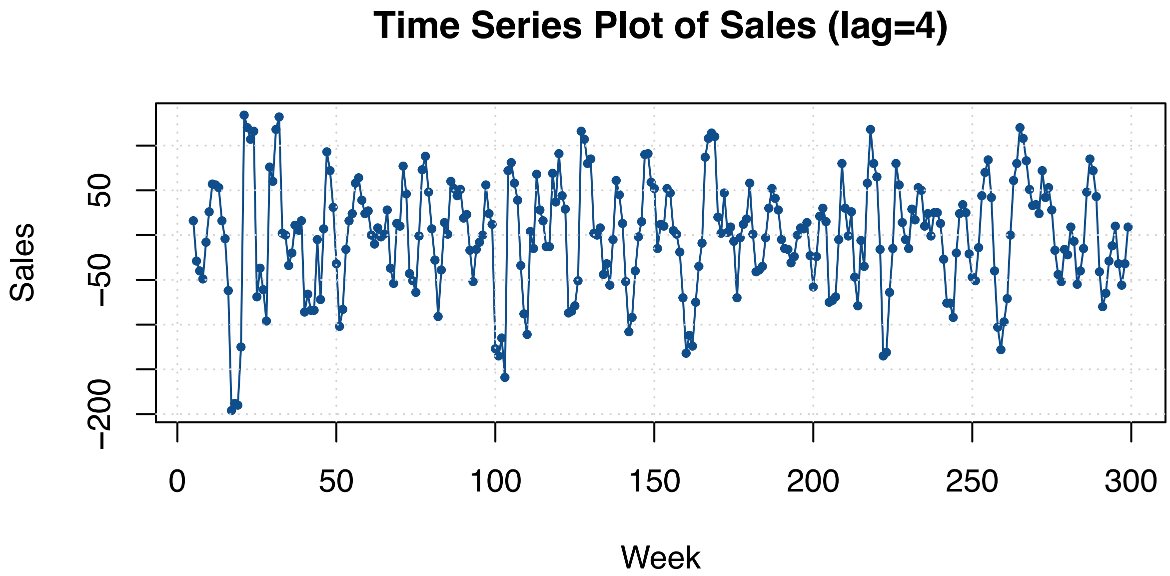
我們將資料繪製成時間序列圖，如下圖。



我們將原始序列資料進行一階差分後繪製成時間序列圖，如下圖。



我們將原始序列資料進行季節性差分後繪製成時間序列圖，如下圖。

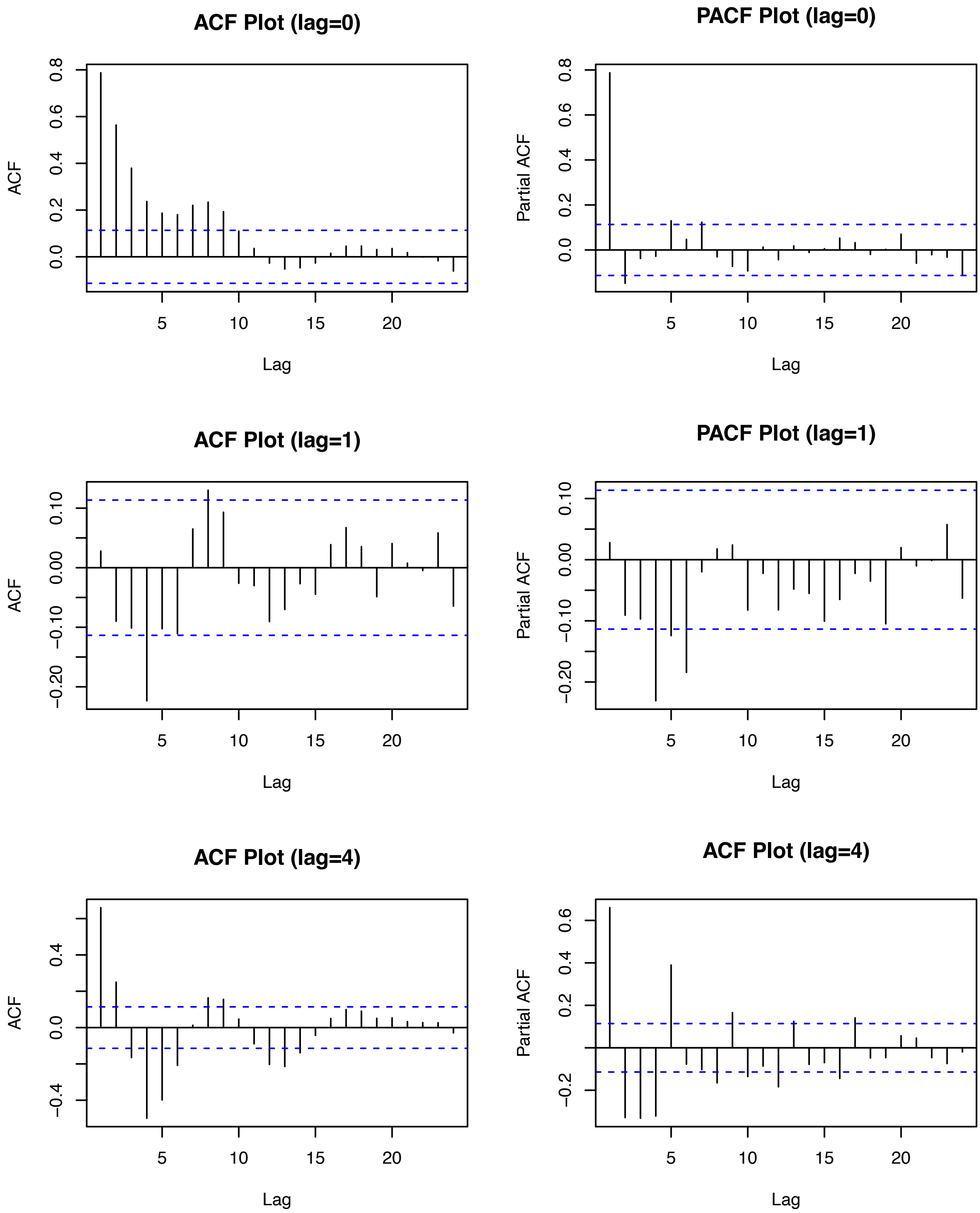


1.4 資料分析

我們以時間序列分析方法來分析此資料集。

1.4.1 自迴歸函數

我們以自迴歸函數（auto-correlation function, ACF）與部分自迴歸函數（partial auto-correlation function, PACF）這兩項基本的工具觀察原始序列資料、進行一階差分後與進行季節性差分的序列資料，如下 6 圖：



1.4.2 平穩性檢定

我們接著分別對原始序列、進行一階差分後的序列與進行季節性差分後的序列進行 Dickey-Fuller 檢定，以確認序列資料是否具有平穩性（stationarity），檢定的假設為 H_0 : 不平穩 v.s. H_1: 平穩，令顯著水準為 0.05。檢定結果如下表：

檢定對象	檢定統計量	p 值
原始序列	-4.1621	0.01
一階差分	-9.4201	0.01
季節性差分	-7.1692	0.01

在三個檢定中，p 值都小於顯著水準，因此我們都拒絕 H_0，表示有顯著的證據顯示原始序列、一階差分後的序列和季節差分後的序列達平穩。

1.4.3 時間序列模型定義

由上表以及 ACF 與 PACF 圖，可看出一階差分後時間序列較平穩，但是由 ACF 圖可發現可能有季節性問題，在 lag 4 有突出，表示原始資料與 4 期後的資料可能存在一定程度的自相關，而在一階差分序列的 ACF 與 PACF 圖中，lag 8 與 lag 12 等後續間隔為 4 的也有突出，因此我們配適考慮含有季節性的 $ARIMA(p,d,q) \times (P,D,Q)_s$：

$$\Phi_P(B^S)\Phi_p(B)\Delta_s^D\Delta^d y_t = \theta_Q(B^s)\theta_q(B)\epsilon_t$$

- y_t 為第 t 週的銷售量
- $\Phi_P(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{s,2} - ... - \Phi_2 B^{s,P})$ 為 P 階季節性 AR
- $\Phi_p(B) = (1 - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_2 B^p)$ 為 p 階非季節性 AR
- $\Delta_s^D = (1 - B^S)^D$ 為季節性差分
- $\Delta^D = (1 - B)^d$ 為非季節性差分
- $\theta_Q(B^s) = (1 - \theta_1 B^s - \theta_2 B^{s,2} - ... - \theta_Q B^{s,Q})$ 為 Q 階季節性 MA
- $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^{s,q})$ 為 Q 階季節性 MA
- ϵ_t 為第 t 週的白噪音（white noise）

1.4.4 資料切割

為了驗證模型預測表現，我們將資料集切割為兩個子資料集，以前 269 筆（90%）資料為訓練集，最後 30 筆（10%）資料為測試集，我們使用訓練集配適模型，以測試集來評估模型的預測能力。

1.4.5 配適結果

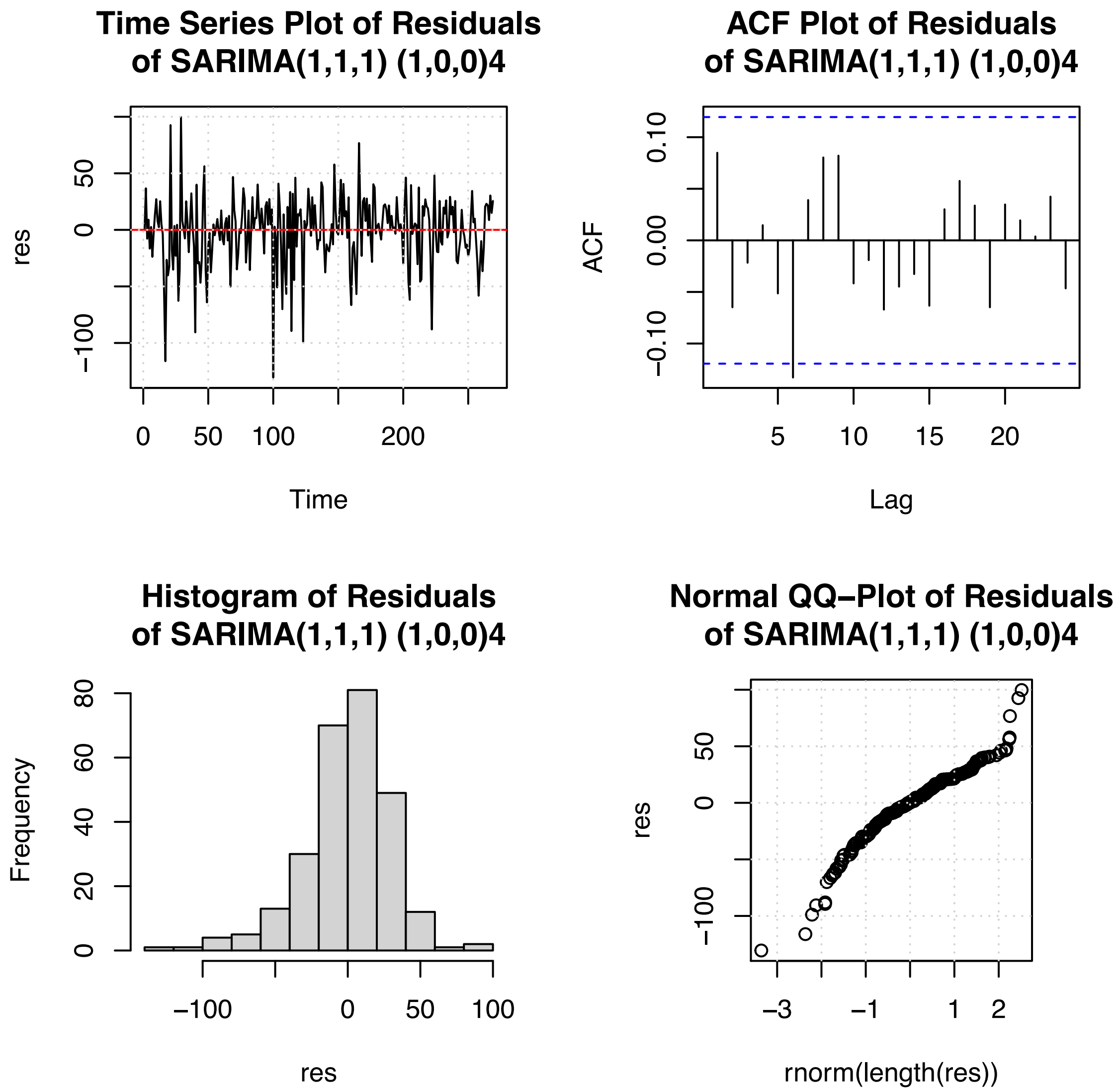
我們嘗試各種參數的模型，並對各模型的殘差進行診斷，以下列出 2 個有通過殘差診斷的候選模型之係數估計結果。

模型	係數	估計值	標準誤
$ARIMA(1,1,1)\ (1,0,0)_4$	AR1	0.7811	0.0401
	MA1	-1.000	0.0110
	SAR1	-0.1884	0.0613

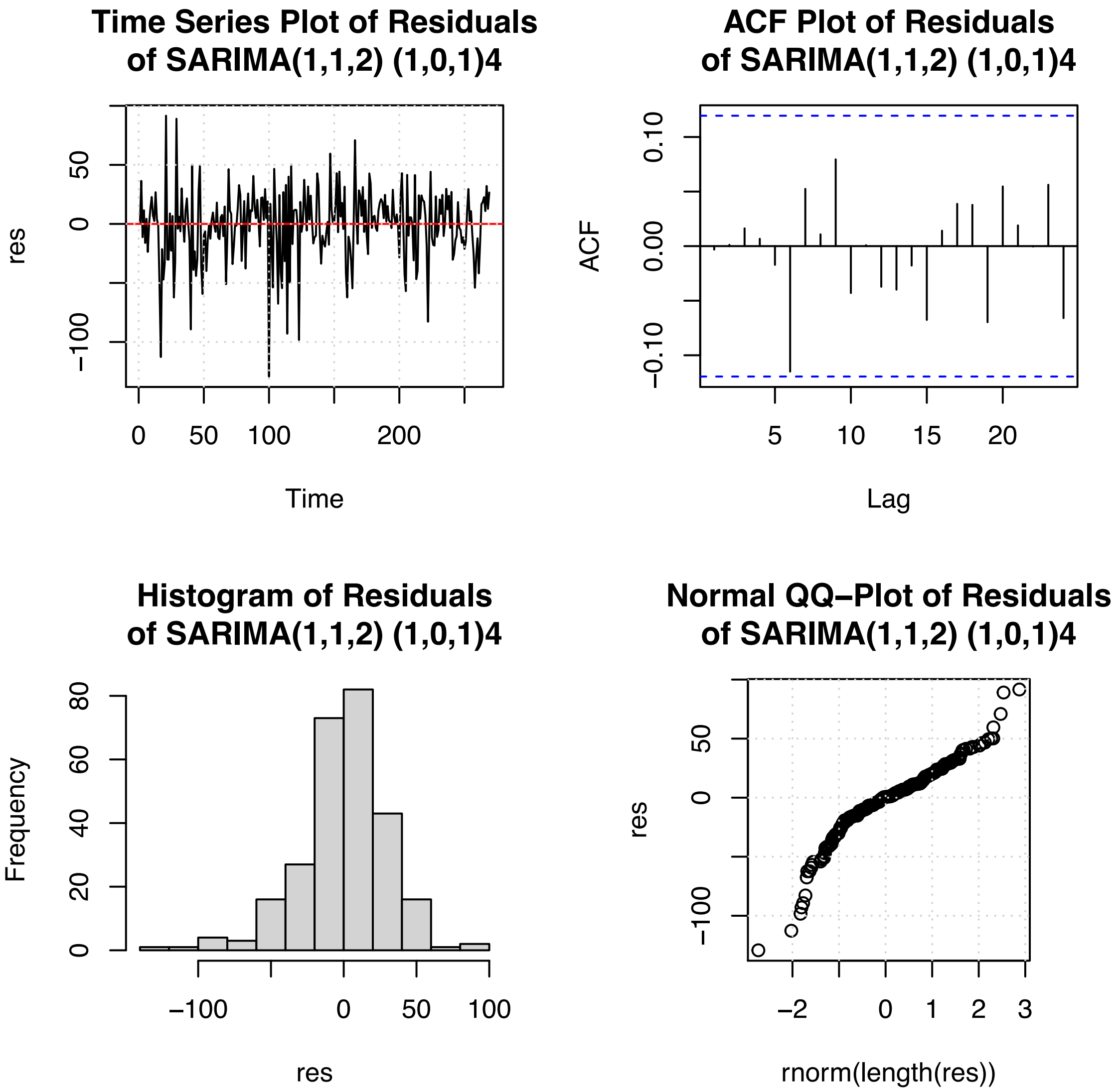
模型	係數	估計值	標準誤
$ARIMA(1,1,2)\ (1,0,1)_4$	AR1	0.7152	0.0582
	MA1	-0.8523	0.0808
	MA2	-0.1477	0.0800
	SAR1	-0.5688	0.1942
	SAR2	0.4167	0.2129

1.4.6 殘差診斷

我們對兩模型進行診斷，首先先繪製殘差圖：



上圖為 $ARIMA(1,1,1)\ (1,0,0)_4$ 的殘差圖，可以發現，殘差的平均大約為零，ACF 圖中大於 0 的 lag 皆無超出區域，表示殘差之間並無相關性，直方圖與 QQ 圖顯示殘差應服從常態分配。



上圖為 $ARIMA(1,1,2)\ (1,0,1)_4$ 的殘差圖，可以發現，殘差的平均大約為零，ACF 圖中大於 0 的 lag 皆無超出區域，表示殘差之間並無相關性，直方圖與 QQ 圖顯示殘差應服從常態分配。

接著我們透過檢定來確認殘差是否符合假設（assumption）。顯著水準均設定為 0.05，檢定與其假設如下：

1. 以 t 檢定法檢驗殘差期望值是否為零：$H_0 : E(\epsilon_t) = 0\ v.s.\ H_1 : E(\epsilon_t) \neq 1$
2. 以 Kolmogorov-Smirnov 檢定法檢驗殘差常態性：$H_0 : \epsilon_t \sim ND\ v.s.\ H_1 : \epsilon_t\ does\ not \sim ND$
3. 以 Ljung Box 檢定法檢驗殘差獨立性：$H_0 : \epsilon_t\ independent\ v.s.\ H_1 : \epsilon_t\ not\ independent$

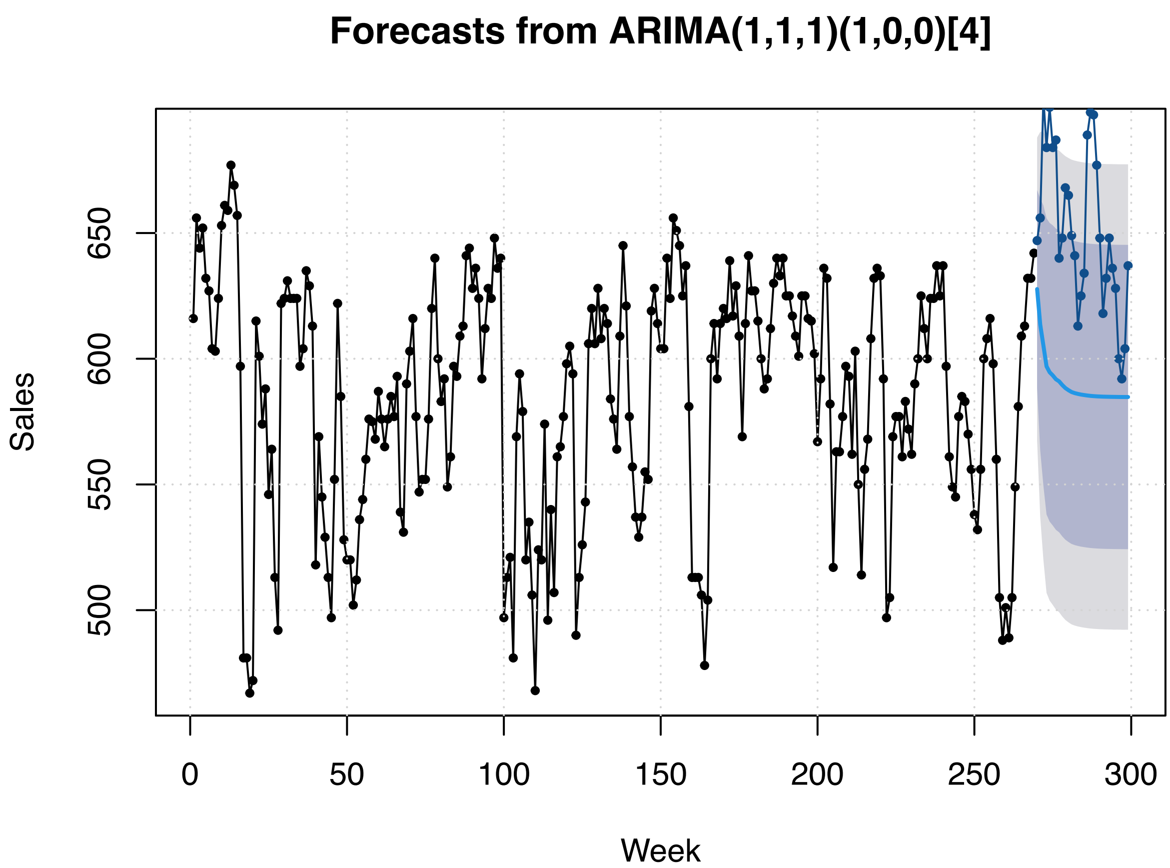
下表是針對兩模型進行三種檢定的結果：

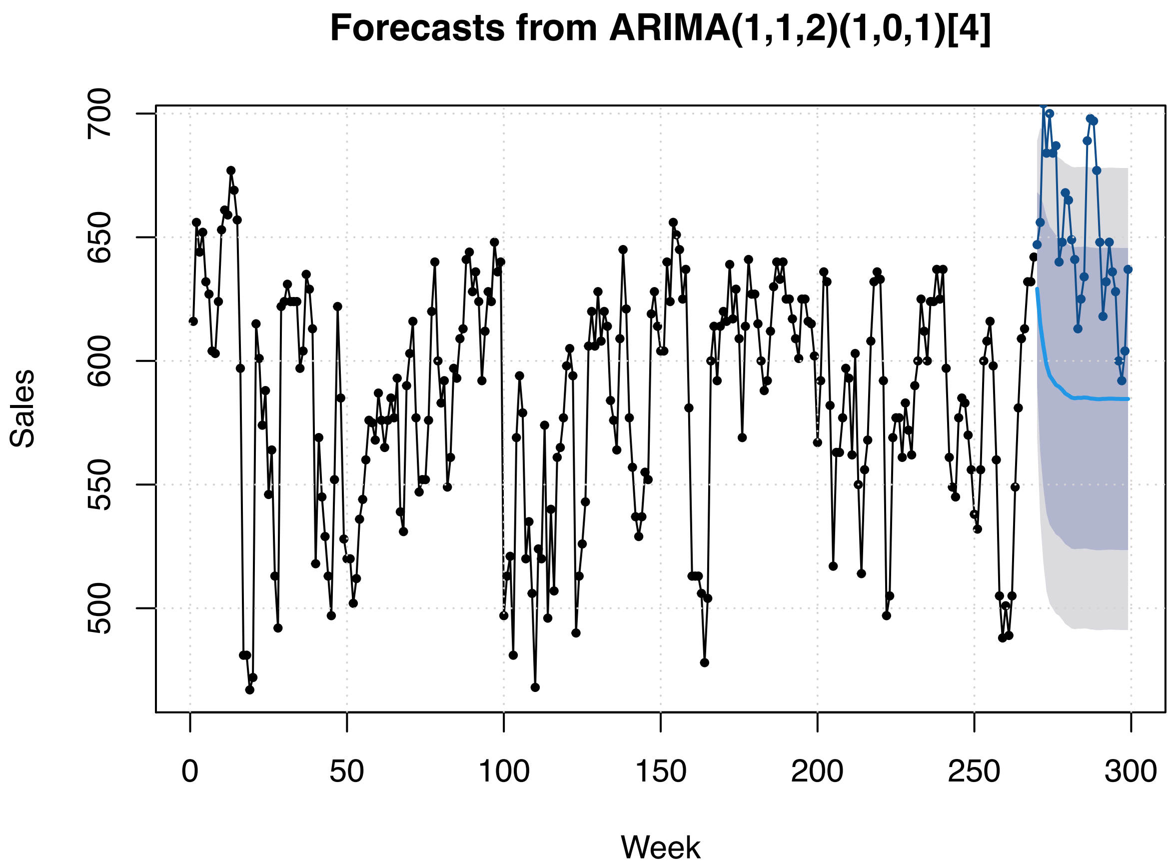
模型	檢定方法	檢定統計量	p 值
$ARIMA(1,1,1)\ (1,0,0)_4$	One-sample t test	$t = -0.5380$	0.5910
	Kolmogorov-Smirnov	$D = 0.5316$	0.0814
	Ljung Box	$\chi^2 = 21.3655$	0.6721
$ARIMA(1,1,2)\ (1,0,1)_4$	One-sample t test	$t = -0.5503$	0.5826
	Kolmogorov-Smirnov	$D = 0.5019$	0.1165
	Ljung Box	$\chi^2 = 14.7211$	0.9477

針對兩模型進行三種檢定的檢定統計量之 p 值都大於顯著水準，因此我們這 6 組檢定中都不拒絕 H_0，也就是說，兩模型殘差期望值都為 0、服從常態分配且無自相關，通過診斷。

1.4.7 模型比較

我們將配適好的兩模型用來預測，並評估預測表現，其預測圖如下兩圖所示。圖中著色的時間範圍為測試資料集的時間範圍，模型在訓練時並無這段時間的資訊。深色範圍為 80% 信賴區間，淺色範圍為 95% 信賴區間，淺藍線為模型預測值，深藍色點為測試資料集範圍內的真實值。可以發現，在兩圖中只有部分的真實值落在 80% 信賴區間內，有些真實值的點甚至在 95% 信賴區間外，顯示模型預測能力不佳。





我們也以模型在訓練集上的赤池訊息量準則（Akaike information criterion, AIC）以及在訓練集與測試集上的均方誤差（mean square error, MSE）作為評估指標，結果如下表所示。可以發現在訓練集 AIC 上，兩模型差異很小，而兩模型在測試集上的 MSE 都遠高於其在訓練集上的 MSE，顯示兩模型泛化的預測能力都不佳。在測試集上的 MSE，第一個模型比第二個模型低，因此我們選擇第一個模型作為最後的模型。

模型	訓練集 AIC	訓練集 MSE	測試集 MSE
$ARIMA(1,1,1)\ (1,0,0)_4$	2605.01	929.18	4702.25
$ARIMA(1,1,2)\ (1,0,1)_4$	2602.98	907.88	4763.37

1.5 結論與建議

我們建立了季節性時間序列模型：$SARIMA(1,1,1)\ (1,0,0)_4$，其預測表現不佳，建議嘗試深度學習方法，如遞迴神經網路，預測表現通常較好。最後模型係數估計結果如下表：

模型	係數	估計值	標準誤
$ARIMA(1,1,1)\ (1,0,0)_4$	AR1	0.7811	0.0401
	MA1	-1.000	0.0110
	SAR1	-0.1884	0.0613