

## 財務計量應用課程期末報告

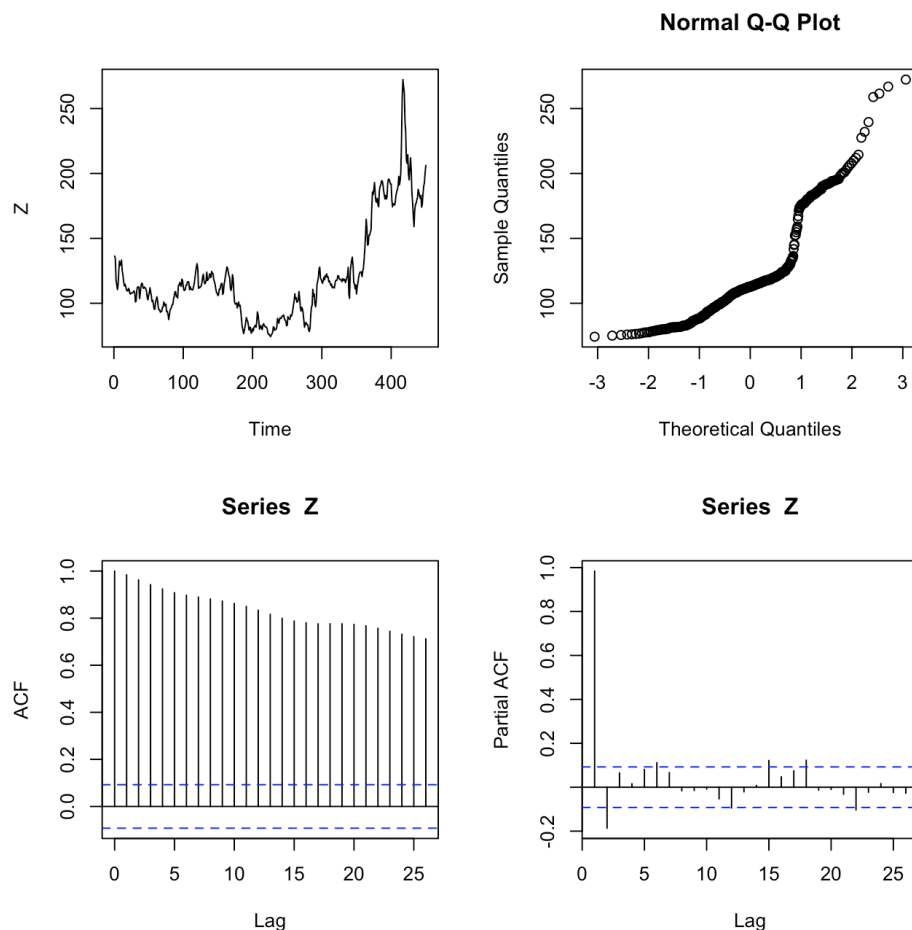
### 一. 時間數列資料

#### 1. 資料敘述

本次分析的資料集是一個真實的數據集，該數據集紀錄了1980年1月至2017年6月期間，美國牛肉原物料每月的價格資料，該價格資料以美分/磅為單位。此分析的主要目的是為探究該資料是否為穩定的時間序列資料，分析將通過建立模型，利用迴歸分析、PACF及ACF圖形並計算AIC來判斷該時間數列資料為何種模型，完成判斷後還將進一步對此資料作出五期的預測。

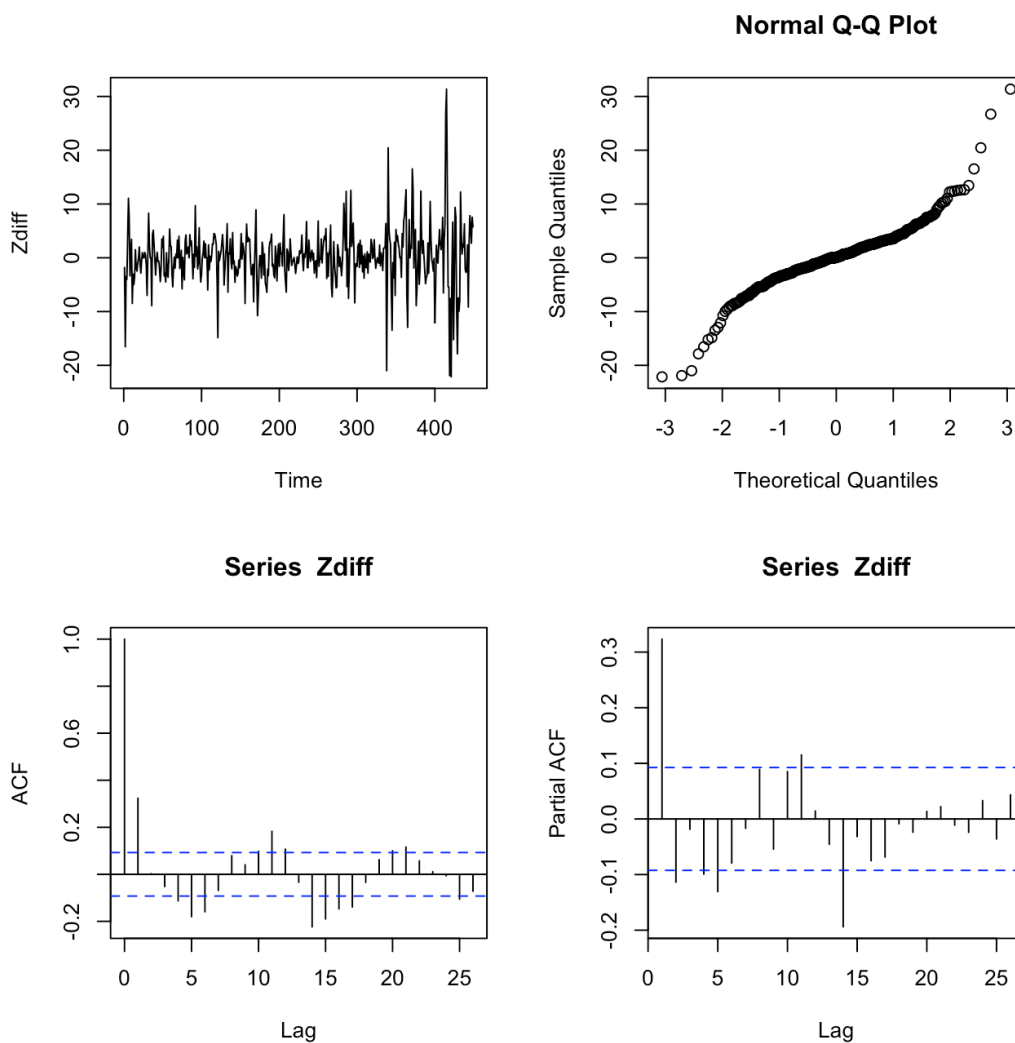
#### 2. 資料分析

分析首先利用美國牛肉原物料月價格的原始資料做了Augmented Dickey-Fuller Test 得到  $p\text{-value} = 0.6603$ ，並做出時間序列圖形如下：



### 原始資料時間數列圖

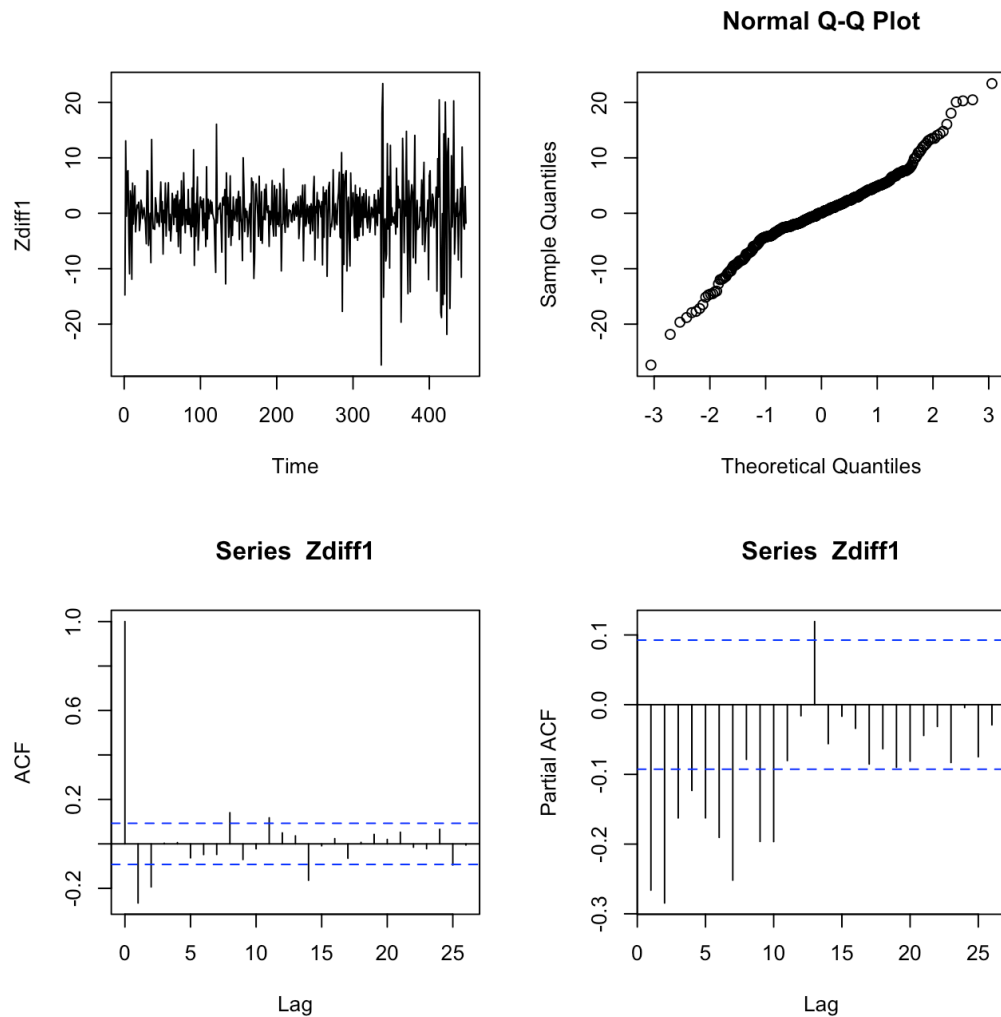
由於檢定的  $p\text{-value} = 0.6603 > 0.05$ ，且該圖形並不平穩，表示該時間數列並不穩定。因此我們對此資料做了一次差分，差分後得到的時間序列圖形如下：



### 第一次差分後資料時間序列圖

第一次差分後的資料檢定的  $p\text{-value} = 0.01$  ( $p\text{-value smaller than printed } p\text{-value}$ )  $< 0.05$ ，做出的時間序列圖形雖相比原始資料得到的圖形有明顯的改善，較差分前更穩定，但我們認為數列還不是非常穩定，且 ACF 的下降趨勢並不是很明顯，需再做一次差分。

第二次差分後的圖形如下：



第二次差分後資料時間序列圖

第二次差分後的資料檢定的  $p\text{-value} = 0.01$  ( $p\text{-value smaller than printed } p\text{-value}$ )  $< 0.05$ ，做出的時間序列圖形第一次差分後的資料得到的圖形更穩定，ACF 圖形呈快速下降趨勢，表示數列已穩定。

然後，通過計算 AIC 的方式，找到當  $p=2, q=2$  時，aic 是最小值，此時  $aic=2711.791$ ，因此我們認為該模型為 ARIMA(2,2,2)。

確認模型後，我們進一步做了五期的預測，得到的預測結果為：207.0437, 206.0810, 204.7886, 203.7814, 203.2059。

### 3. 結論

在對該美國牛肉原物料每月的價格資料進行兩次差分後，可得到較穩定的時間數列圖形，同時通過 AIC 也可判斷該模型屬於 ARIMA (2, 2, 2)。另外該資料五期的預測結果為：207.0437, 206.0810, 204.7886, 203.7814, 203.2059。

## 二. 類別資料

### 1. 資料敘述

本次分析的資料集是一個真實的數據集，該數據集篩選了台灣成立年份距今在 10 年內的食物，醫藥，化工公司 2018 年的獲利能力資料，包含營收，eps 等。此分析的主要目的是為了用現有的資料來預測將來的獲利能力，分析將通過建立模型，利用邏輯迴歸分析和 svm 來完成資料的分類以及根據訓練資料來測試和預測，並且比較哪種方法更加的合適。其變數相關資料如下：

y: EPS 年增 / 減 (增是 1, 減是 0)

x1: 營收

x2: 淨利率 (%)

x3: ROA (%)

x4: ROE (%)

### 2. 資料分析

在進行資料分析的第一步，先將資料分為了 train 資料與 test 資料。前者為原始資料的三分之二，代表解釋能力；後者為原始資料的三分之一，代表預測能力。資料將被隨機分割 50 次以保證有較高的準確率，然後儲存訓練結果矩陣與測試結果矩陣。該資料中，訓練資料有 104 筆，測試資料有 52 筆。

參數設置完成後，為了得到最好的解釋能力與預測能力，分別進行 Logistic Regression 及 SVM 方法的分析。Logistic Regression 與 SVM 分析的解釋能力結果分別為 110.00 與 109.98；預測能力分別為 48.26 與 51.52。解釋能力上，兩者相差不大，Logistic Regression 分析方法稍高於 SVM 方法；而預測能力上 SVM 明顯較好。因此我們認為最佳的分析方法為 SVM。

### 3. 結論

通過第二部分的資料分析結果可得知，在使用 Logistic Regression 方法分析該資料時，解釋能力較 SVM 好；而在使用 SVM 方法分析該資料時，預測能力較

Logistic Regression 更好。該資料在 Logistic Regression 模型下得到的解釋能力為 110.00，預測能力為 48.26；在 SVM 模型下得到解釋能力為 109.98，預測能力為 51.52。解釋能力上，兩者相差不大，Logistic Regression 分析方法只是稍高於 SVM 方法；而預測能力上 SVM 明顯較好。因此我們認為最佳的分析方法為 SVM。