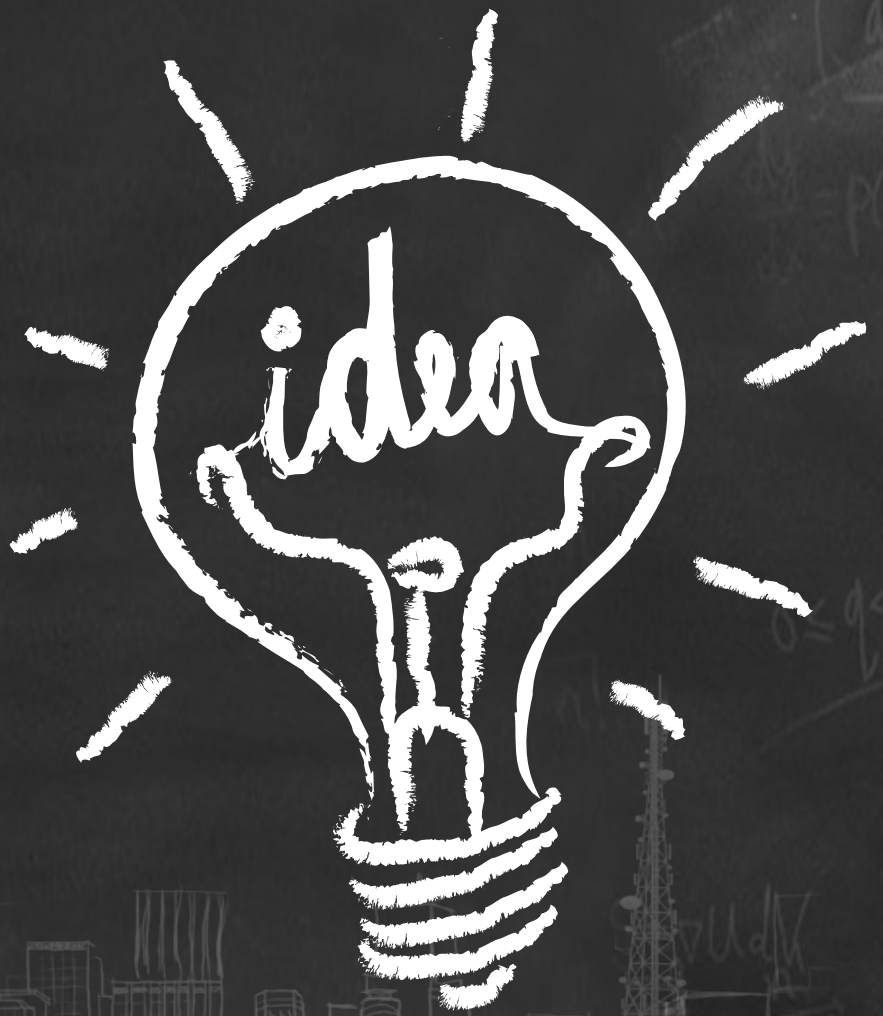
時間序列分析期中報告

台灣月出生人口分析
（2000~2023）



組員：107508006 歐西四 陳葳芃
108405056 廣告四 楊岳錩

報告大綱
Contents

- 資料介紹及來源
- Raw data time series的探索與insight
- Raw data 的ACF 、PACF
- 一階差分後的比較
  time series 、 ACF 、 PACF
- 小結———配適模型選取

資料介紹及來源

	time	amount	month
1	2000-01-01	24042	01
2	2000-02-01	22263	02
3	2000-03-01	26259	03
4	2000-04-01	22587	04
5	2000-05-01	25019	05
6	2000-06-01	22937	06
7	2000-07-01	24436	07
8	2000-08-01	25771	08
9	2000-09-01	26915	09
10	2000-10-01	28710	10

https://nstatdb.dgbas.gov.tw/dgbasall/webMain.aspx?sdmx/a130203010/1...M.&startTime=2000-01&endTime=2023-03

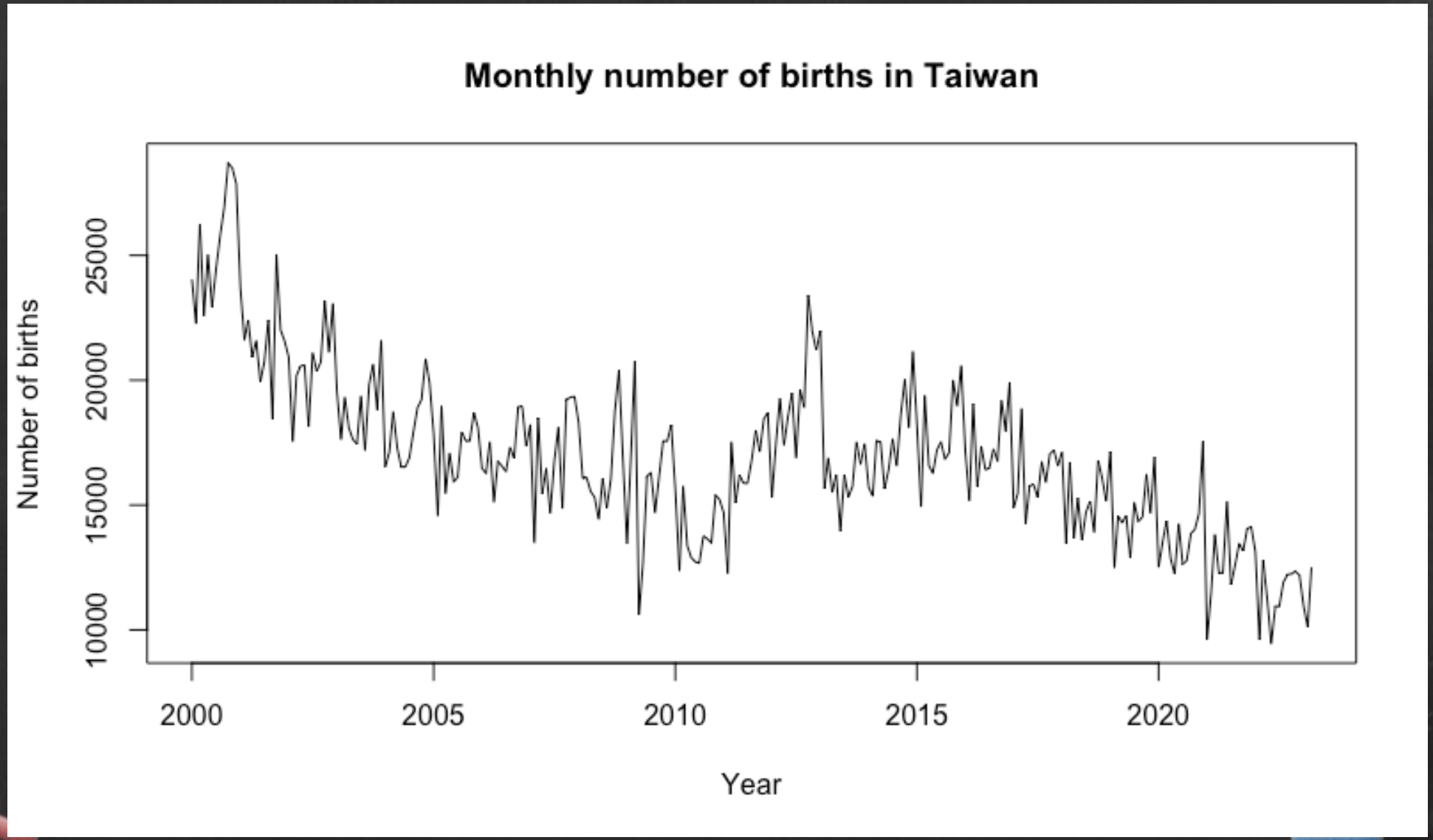
台灣月出生人口統計

本份資料源自中華民國統計資訊網——總體統計資料庫，紀錄自2000年1月以來每月出生人數，資料目前更新至2023年3月。

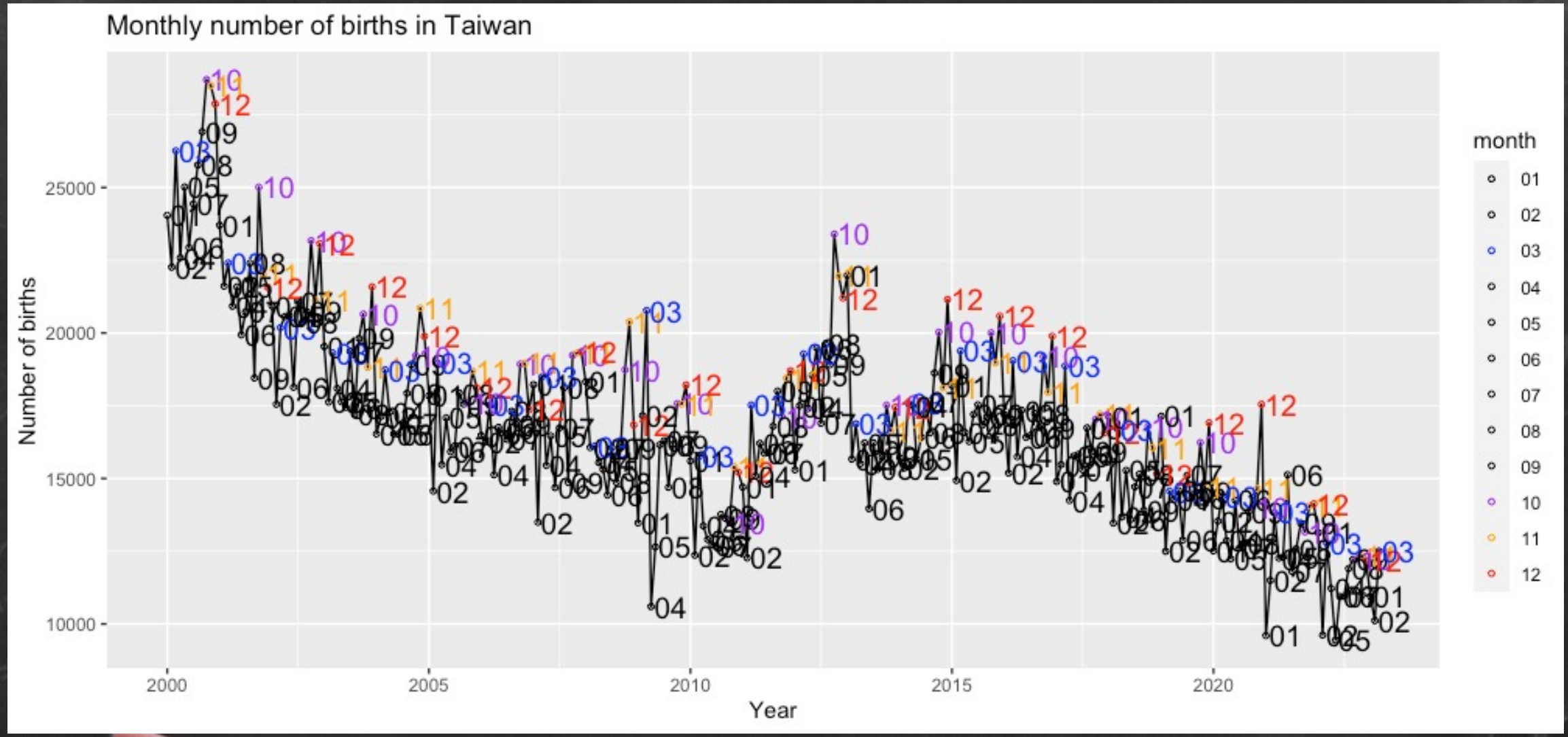
樣本個數：279
變數個數：3
變數介紹：

	time	amount	month
類別	時間 time	數值 numerical	類別 categorical
說明	年月份	出生人數	月份

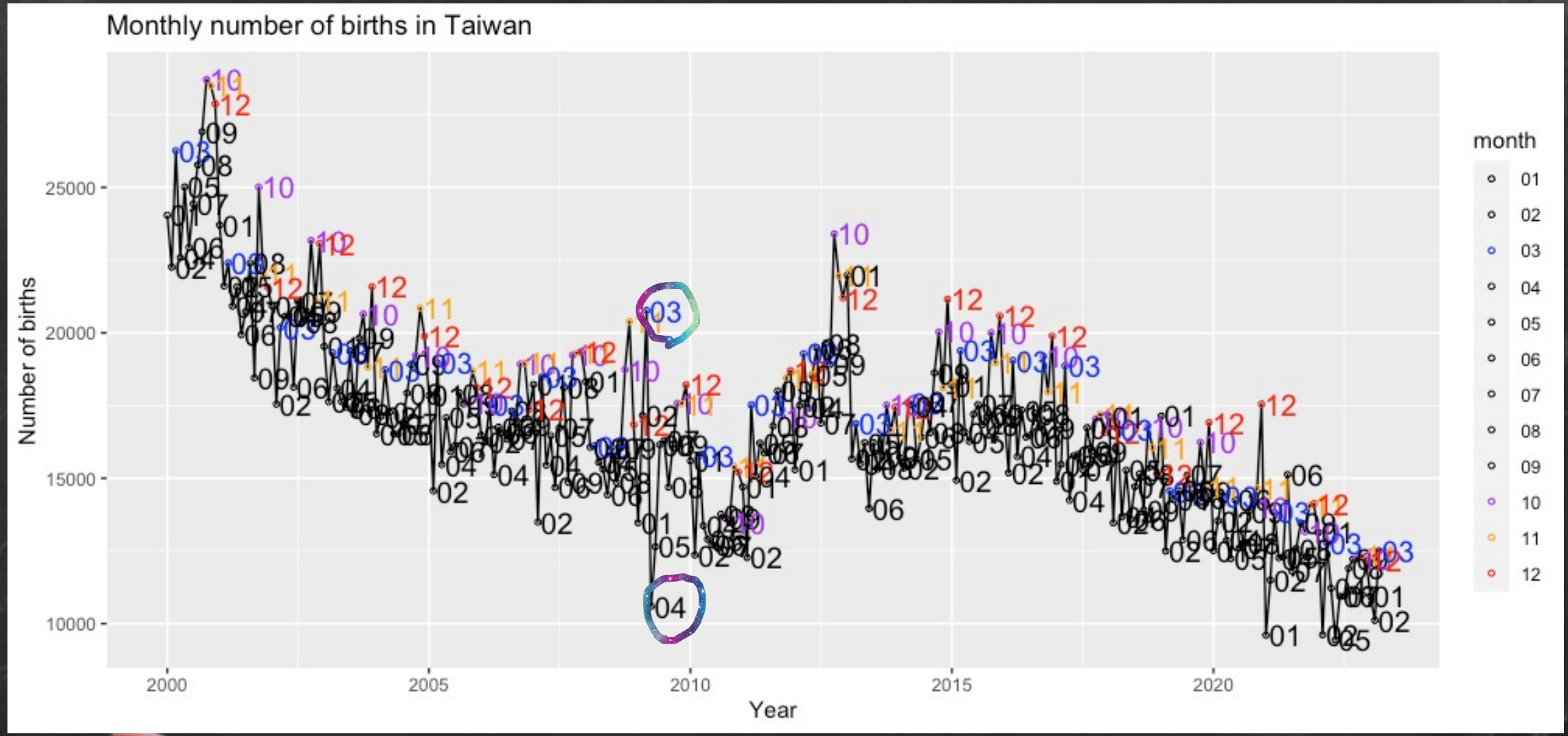
Raw data time series 探索



Raw data time series 探索 (ggplot)



Raw data time series 探索季節性與異常處



Raw data time series 的 insight

1. 出生人數高峰在10、12月
   推測：與情人相關特殊節慶影響
   Ex: 2月情人節、12月聖誕節。

2. 2009 3月　為一個分水嶺
   2010 整年　出生人數都處於低迷的情況
   推測：2008 2009經濟蕭條
   ➡隔一年出生人數下降

文獻回顧——經濟蕭條與否跟出生率的關係

劉錦添、陳妍倩、王齡懋(2013)以台灣關廠資料，利用 probit model 檢視丈夫經歷關廠所導致的失業對婚姻的影響，發現丈夫失業會提高離婚機率約 9%，尤其是失業之後的三年以及失業持續期間超過一年的的影響更強烈，此篇文獻也指出年齡層較大的丈夫失業以及教育程度較低的家庭失業對離婚的負向效果越強，而妻子失業則對離婚沒有顯著的影響。

在生育方面，如果人們認為蕭條將長期持續，那麼對將來預期財富是下降的，可是養育子女的成本下降的幅度較小，所以養育小孩總財富負增長較多，因此在大蕭條時期，所得及替代效果將使得大蕭條期的生育率下降；同樣的，二次大戰期間，資源轉移至軍事上，同樣隱含財富下降，因此同樣降低了生育率；同時養育小孩的成本相對於工資成長較多，因此會造成女性勞動參予率上升，加上男性從軍及較高的實質貼現率，這些結果都將暫時性的增加子與養育的邊際成本，造成生育率下降。

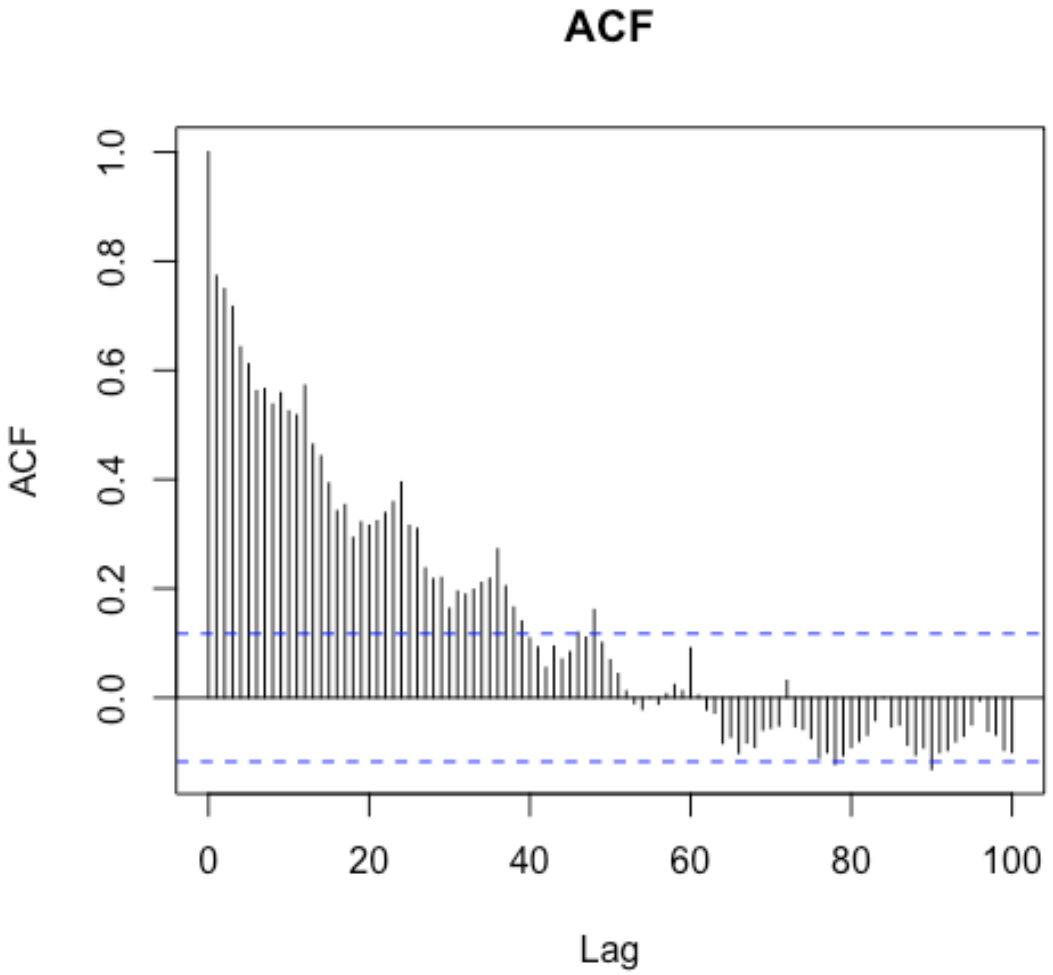
王英傑。2015，中央大學
《景氣循環對婚育行為之影響》，p.8

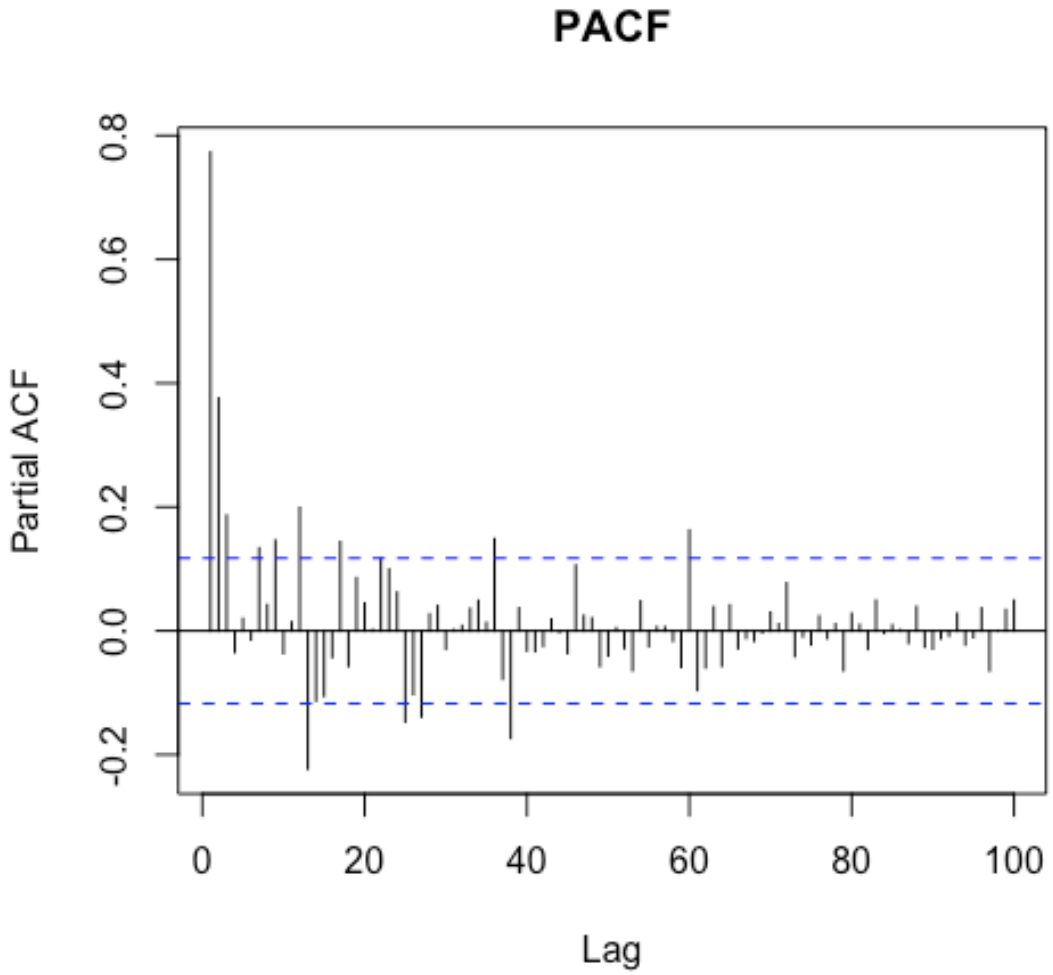
文獻回顧——經濟蕭條發生時間點

房屋市場活絡，加上金融中介者操作槓桿投資，使貨幣流通速度自 2003 年開始持續上升，市場逐漸揮別 2000 年初的網路泡沫和 911 攻擊事件。貨幣流通速度在 2007 年至 2008 年第三季達到最高，之後卻急劇下降，象徵市場蕭條再度來臨。金融中介者鬆綁制度，而金融中介者又利用制度鬆綁的漏洞極大化個人利益。

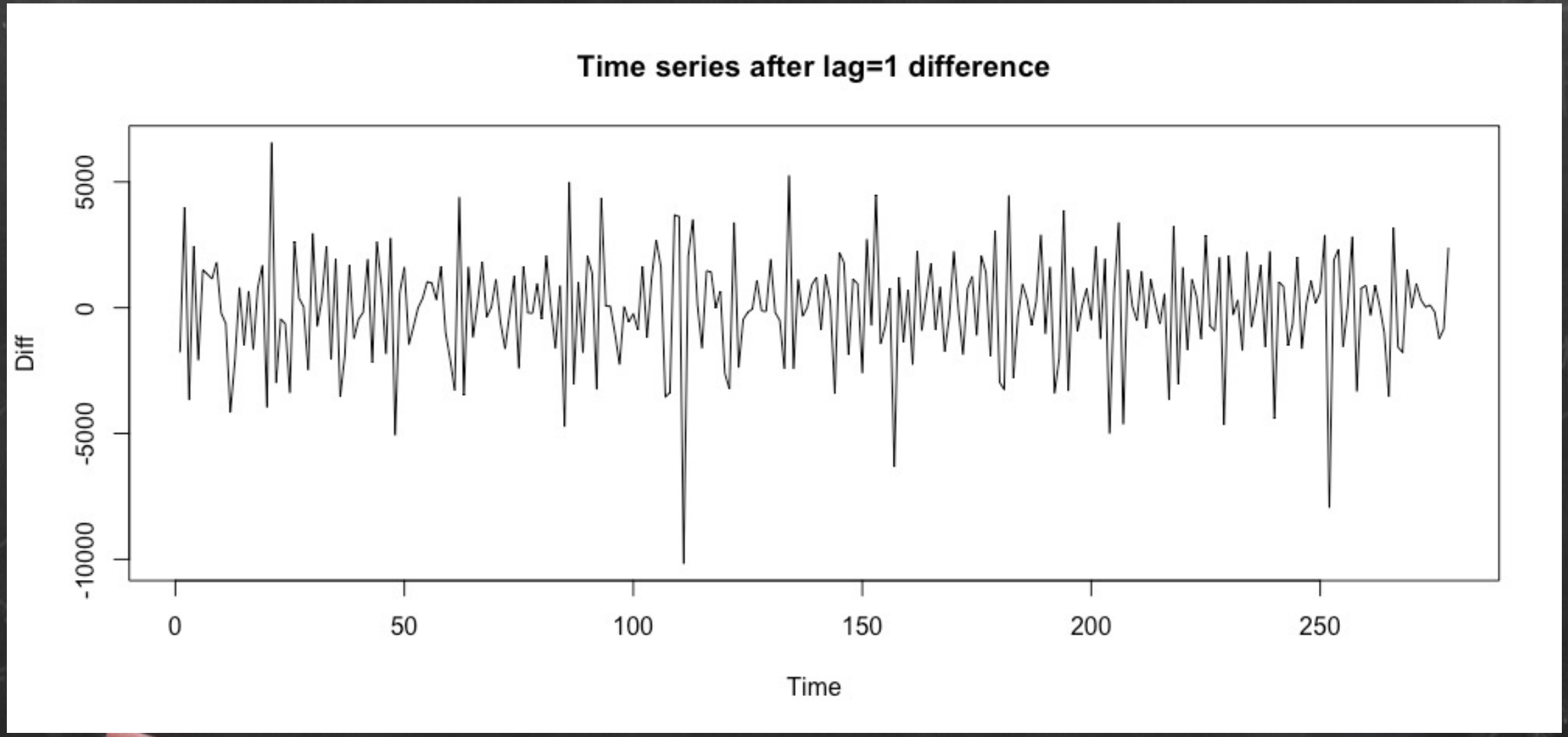
戴庭玉。2011，陽明交通大學
《美國2007年至2009年引發環球金融風暴的制度因素》，p.7

Raw data 的 ACF、PACF

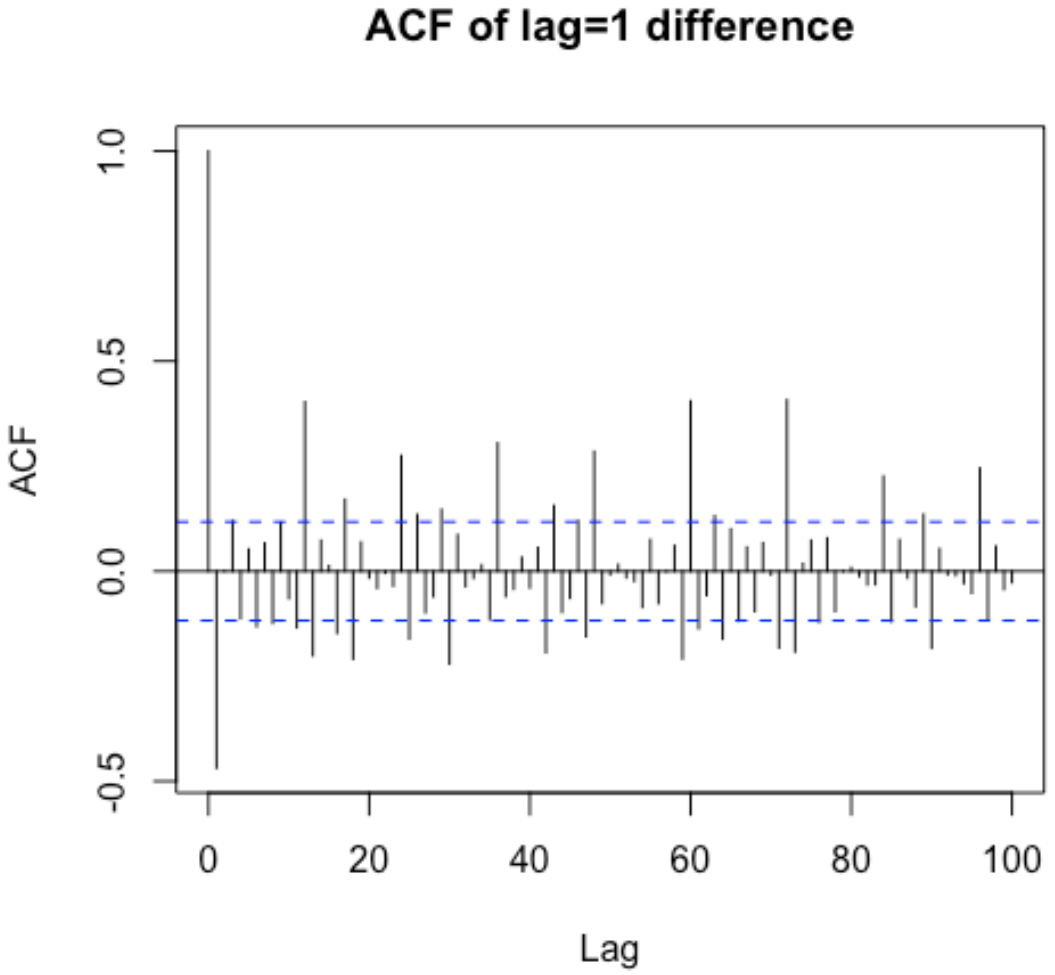


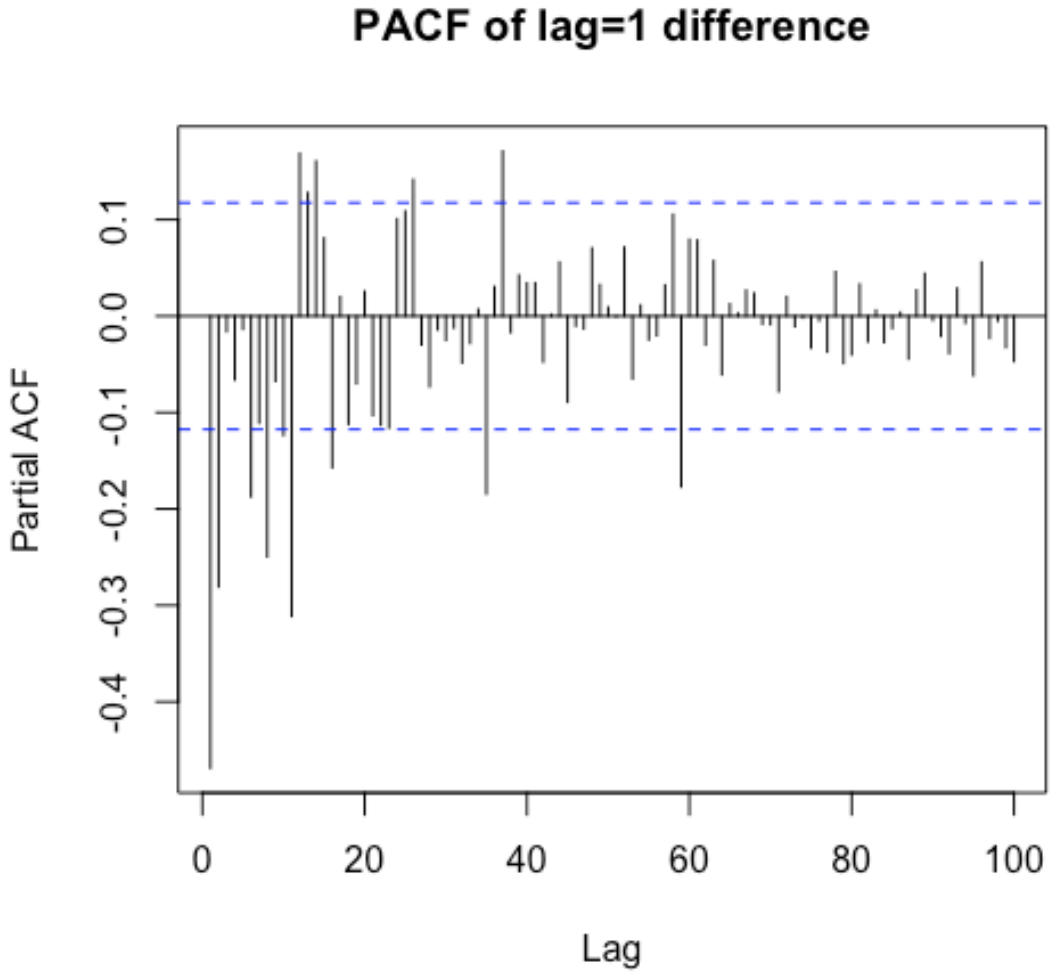


一階差分後的 time series (lag=1)

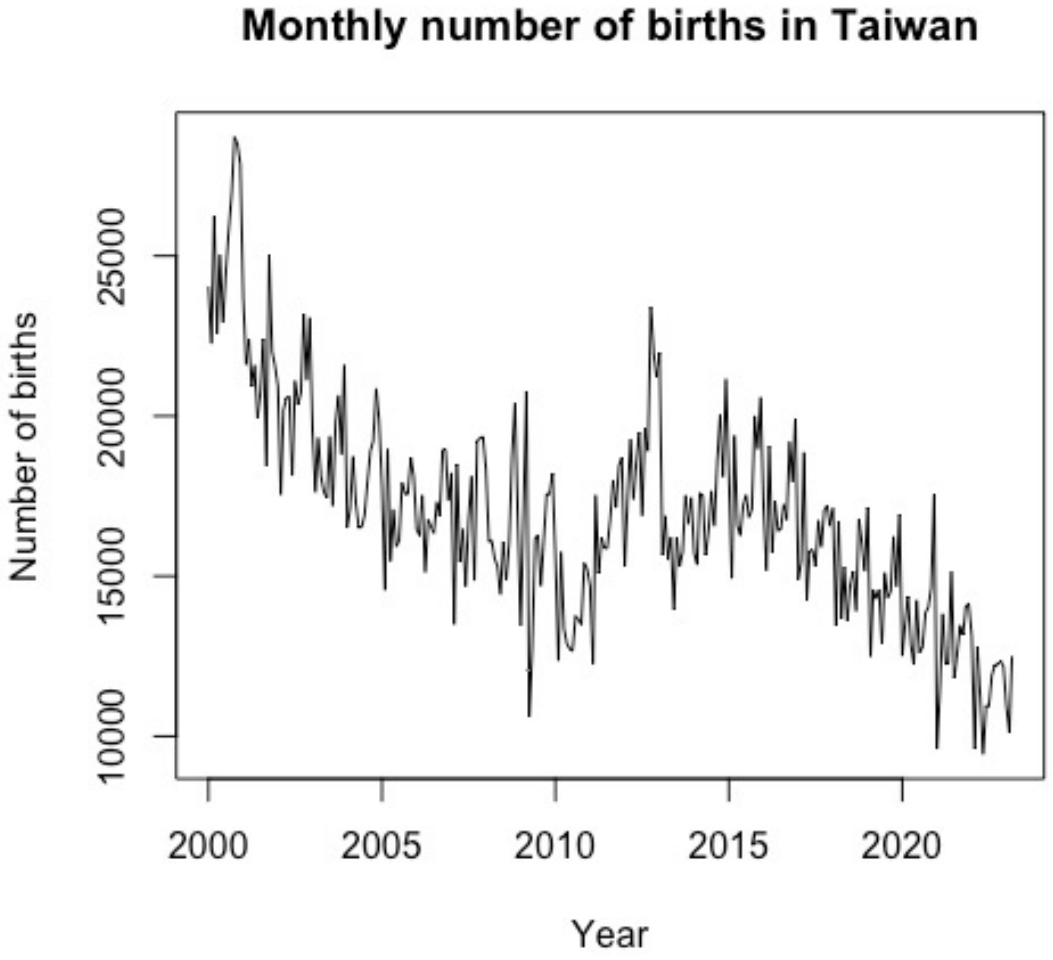


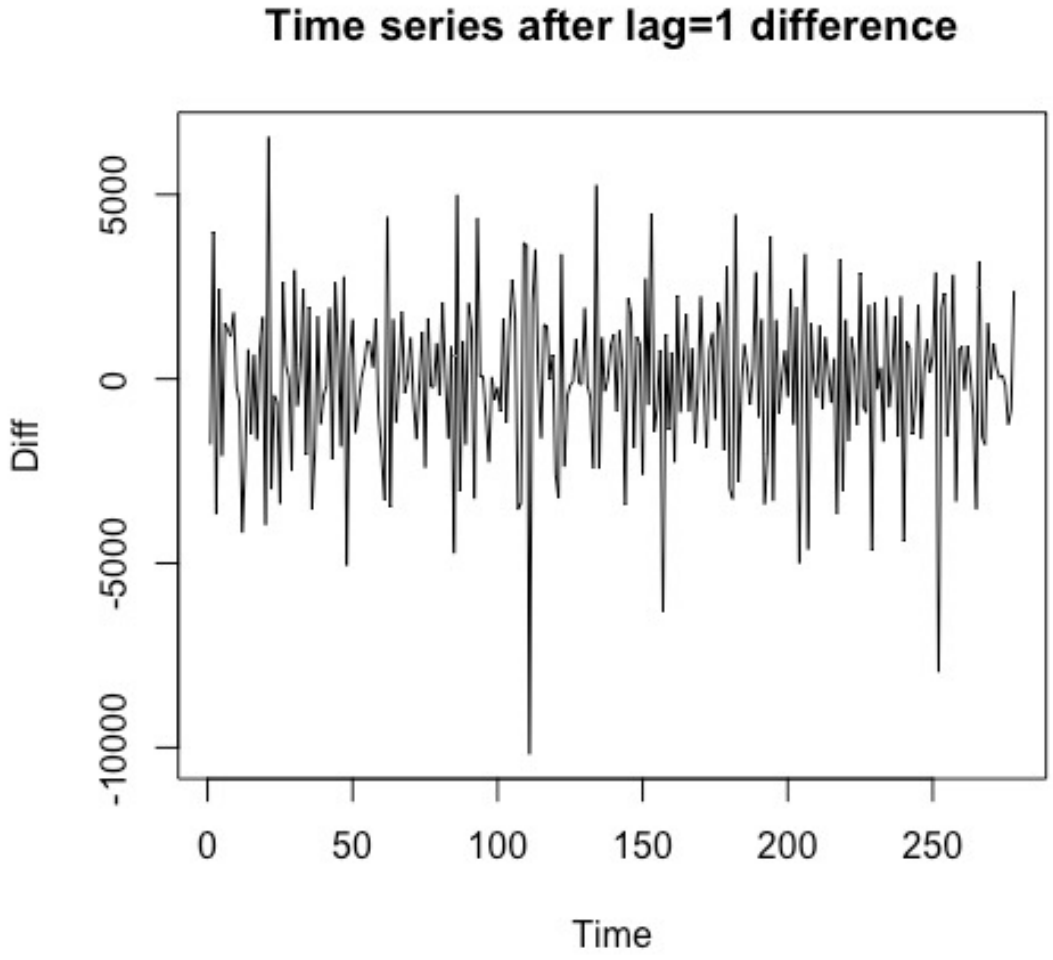
一階差分後的 ACF、PACF (lag=1)



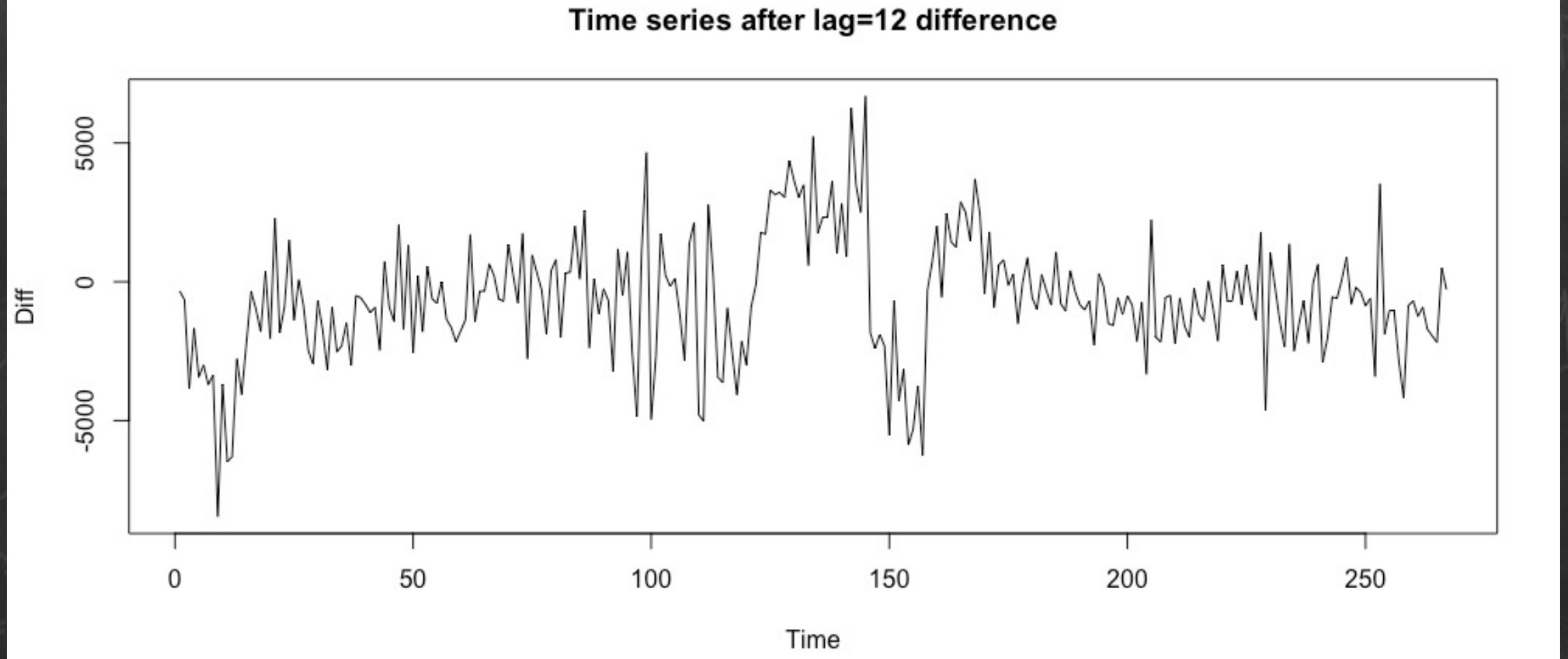


Time series 平穩比較 (原始 vs 一階差分；lag=1)

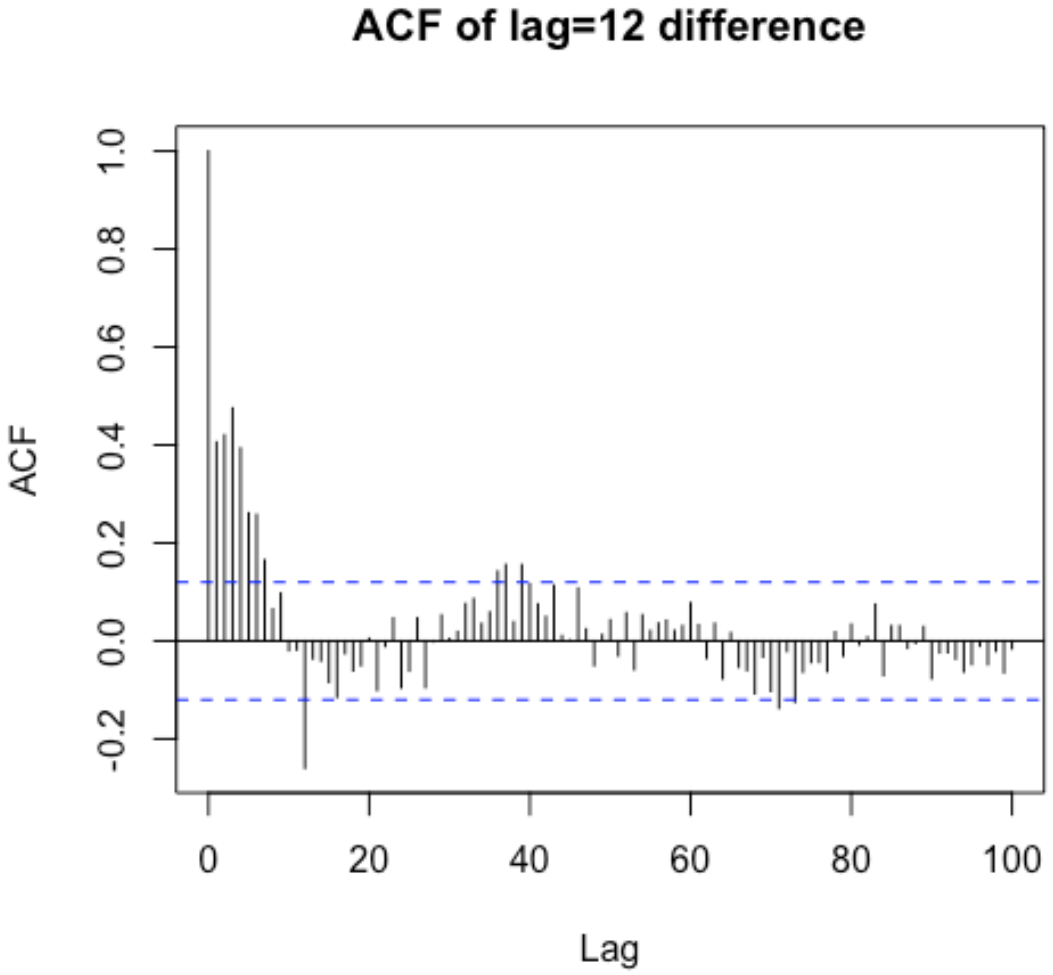


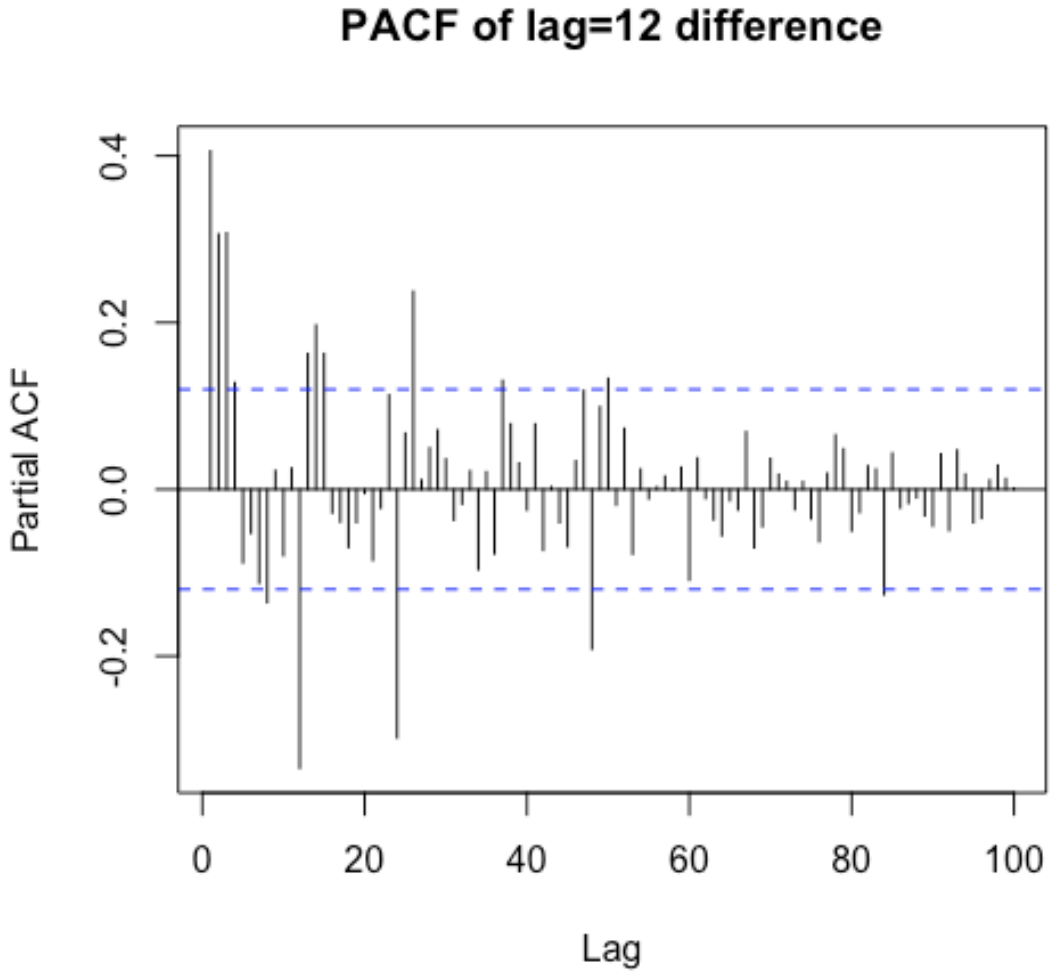


一階差分後的 time series (lag=12)

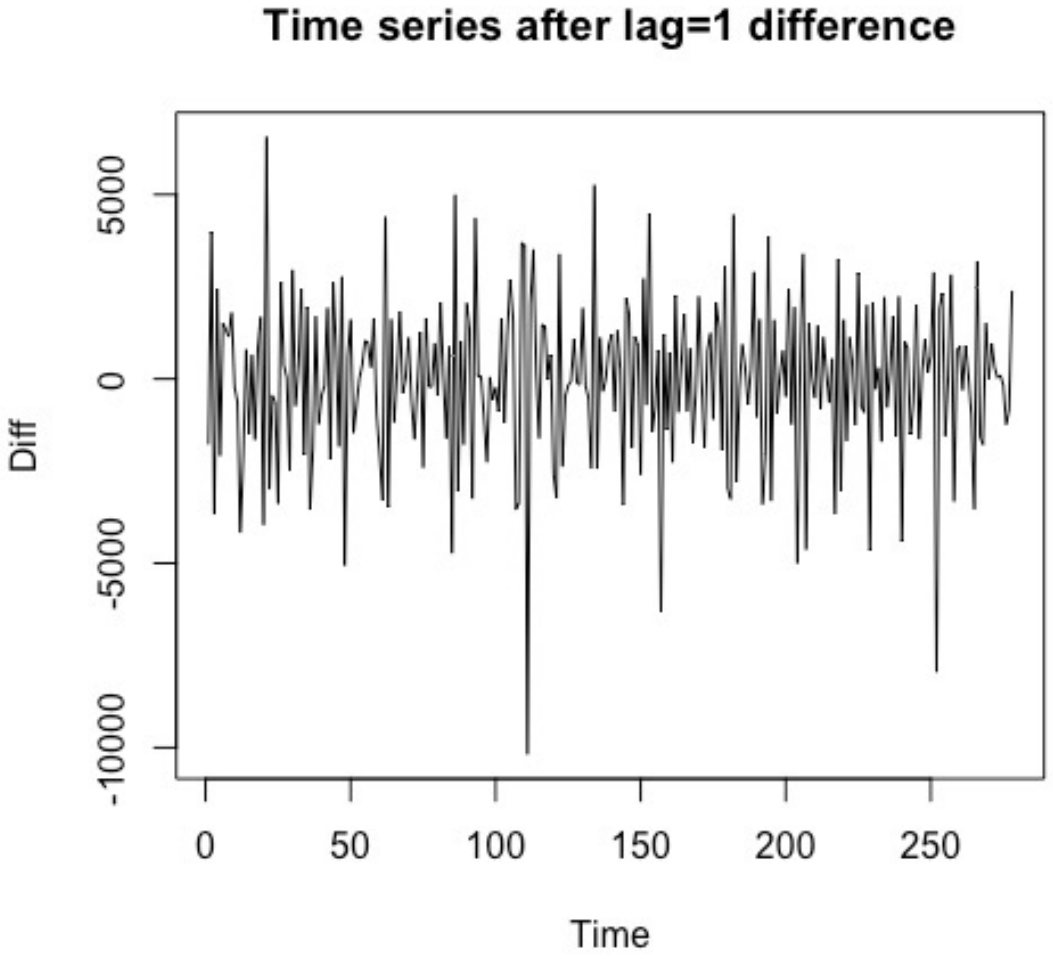


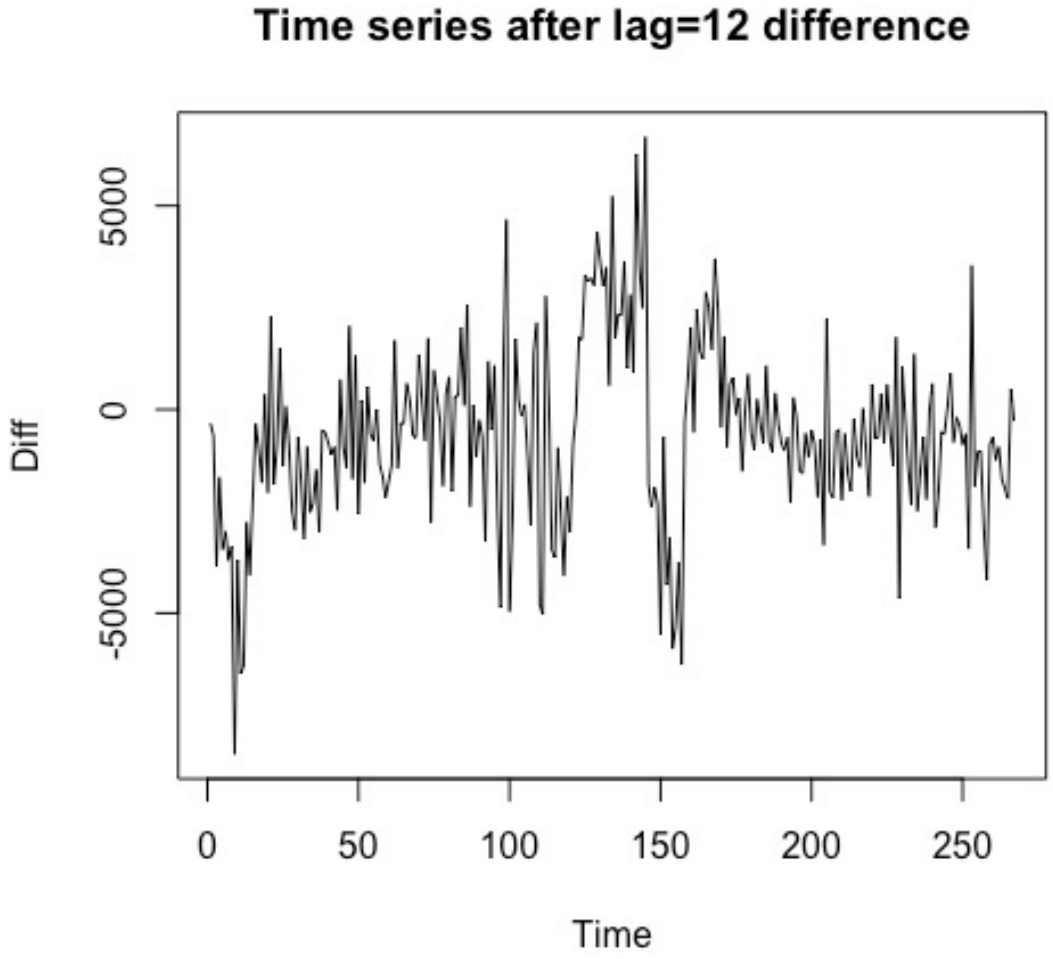
一階差分後的 ACF、PACF (lag=12)



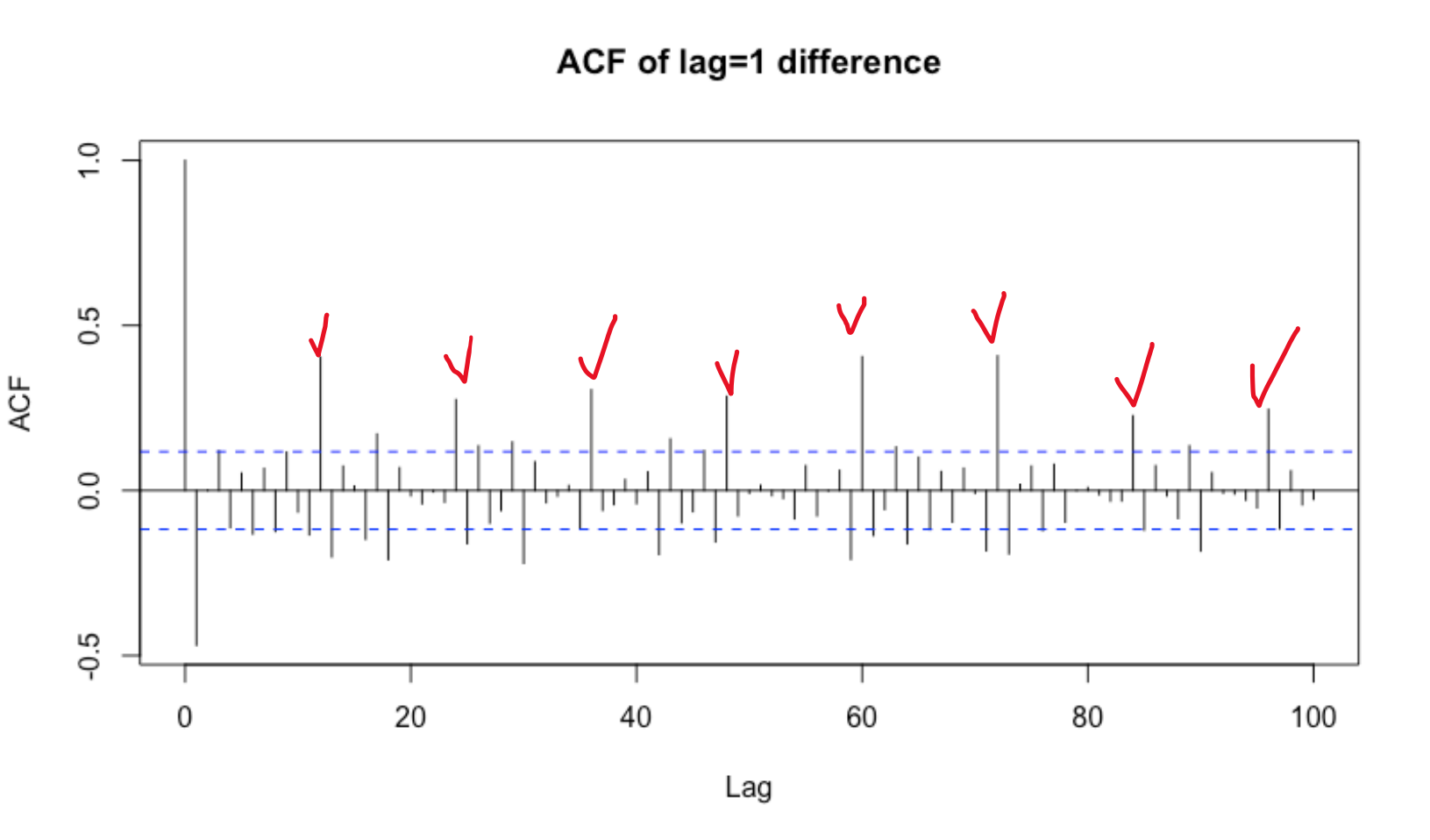


Time series 平穩比較 (一階差分；lag= 1、12)





小結——配適模型選取



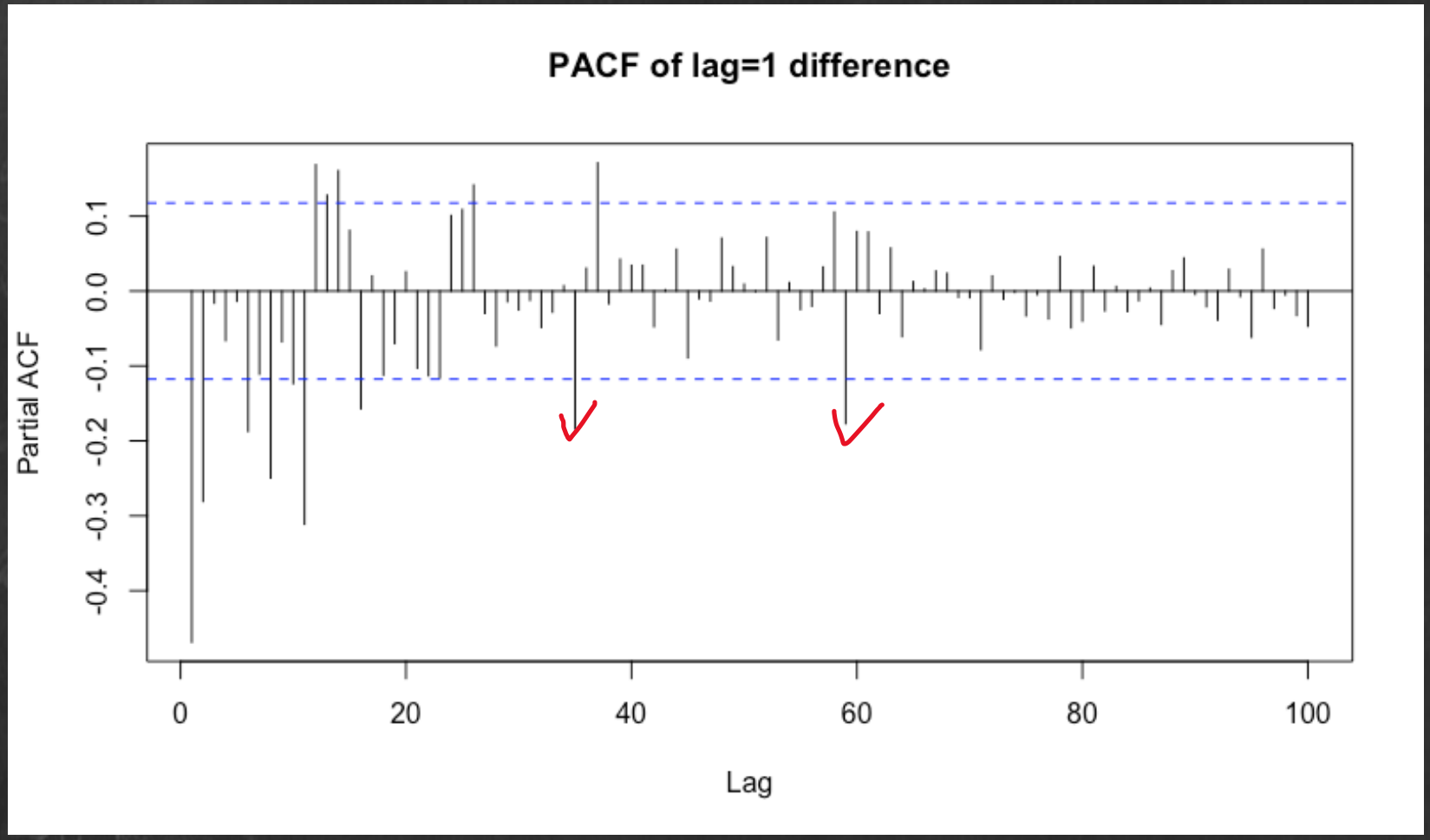
ACF 觀察：

I. 季節性Seasonal
   > s=12

II. Tail off at
   ks ;k=1,2,3...

III. 非季節性部分
   > 看起來有前後期相關
   MA(1)
   > 視為遞減
   AR(1)

小結——配適模型選取



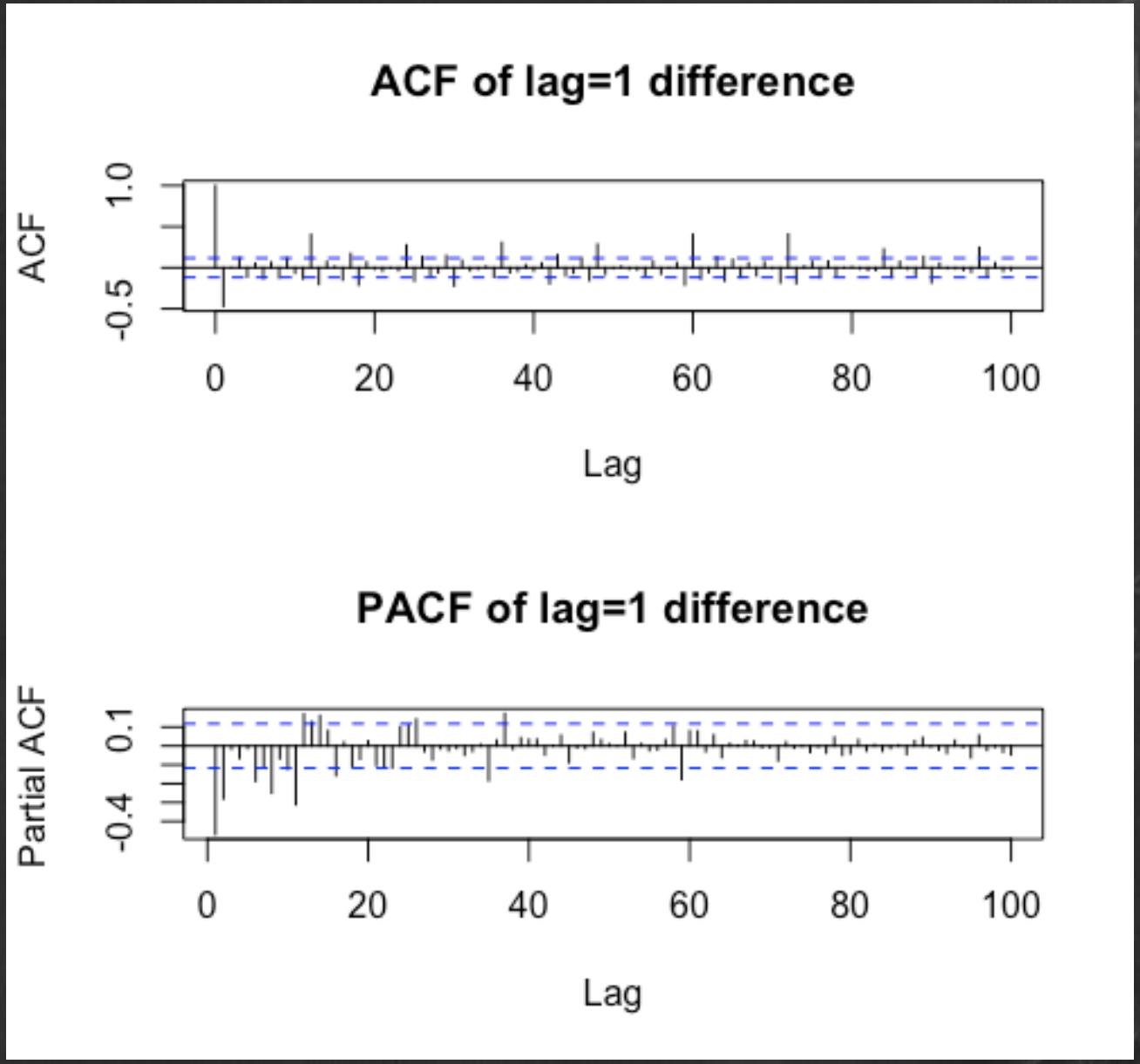
PACF 觀察：

I. cut off after
   Ps ;s=12
   P推測3 or 5

小結——配適模型選取

ARIMA
I. (4,1,3)

SARIMA：
I. $(0,1,1)x(3,0,0)_{12}$
II. $(1,1,0)x(3,0,0)_{12}$



感謝您的耐心聆聽
以上是我們的報告

附錄——R 程式碼

```
##### 時間數列分析 期中報告
##### 組員：陳葳芃 楊岳錩
##### 選用資料：2000-2023 台灣出生月人口數


library(readxl)
df_birth <- read_excel("Downloads/tw_birth(2000-2023)(1).xlsx")
str(df_birth)

plot(df_birth$time, df_birth$amount, main="Monthly number of births in Taiwan",type='l',xlab = "Year", ylab = "Number of births")

library(tidyverse)
df_birth$month <- as.factor(format(df_birth$time, "%m"))
month_colors <- c("black","black","blue","black","black", "black","black","black","black","purple","orange","red")
ggplot(df_birth, aes(x = time, y = amount)) +
  geom_line() +
  # 加入點的顯示
  geom_point(aes(color = month), size = 1, shape = 21) +
  # 在點上加上月份
  geom_text(aes(label = month), hjust = -0.1, color = month_colors[df_birth$month],size = 5)+
  scale_color_manual(values = month_colors) +
  labs(x = "Year", y = "Number of births", title = "Monthly number of births in Taiwan")
```

附錄——R 程式碼

```
par(mfrow=c(1,2))
acf(df_birth$amount, main="ACF",lag.max = 100)
pacf(df_birth$amount, main="PACF",lag.max = 100)

###1期 difference
data<-df_birth$amount
data_diff1<-diff(data)
ts.plot(data_diff1, main="Time series after lag=1 difference",xlab = "Time", ylab = "Diff")
acf(data_diff1, main="ACF of lag=1 difference", lag.max = 100)
pacf(data_diff1, main="PACF of lag=1 difference",lag.max = 100)


###12期 difference
data_diff12<-diff(data, lag = 12)
ts.plot(data_diff12, main="Time series after lag=12 difference",xlab = "Time", ylab = "Diff")
par(mfrow=c(2,1))
acf(data_diff12, main="ACF of lag=12 difference", lag.max = 100)
pacf(data_diff12, main="PACF of lag=12 difference", lag.max = 100)
```