Time Series HW11



B082040005 高念慈

2023-05-12



Consider the daily simple returns of IBM stock, CRSP value-weighted index, CRSP equal-weighted index, and the S&P composite index from January 1980 to December 2008. The index returns include dividend distributions. The data file is d-ibm3dxwkdays8008.txt, which has 12 columns. The columns are (year, month, day, IBM, VW, EW, SP, M, T, W, H, F), where M, T, W, R, and F denotes indicator variables for Monday to Friday, respectively.

the daily simple returns

```
data4 = read.table("https://faculty.chicagobooth.edu/-/media/faculty/ruey-s-tsay/teaching/fts
3/d-ibm3dxwkdays8008.txt", header=T)
head(data4)
```

```
##
                   ibm
                                            sp M T W R F
   year mom day
                            VW
                                    ew
## 1 1980
         1 2 -0.029126 -0.020089 -0.011686 -0.020196 0 0 1 0 0
         1 3 0.016000 -0.006510 -0.011628 -0.005106 0 0 0 1 0
## 2 1980
## 4 1980
        1 7 -0.003945 0.004368 0.007013 0.002722 1 0 0 0 0
## 5 1980
         1 8 0.067327 0.019340 0.014152 0.020036 0 1 0 0 0
## 6 1980
            9 -0.029685 0.001714 0.007452 0.000918 0 0 1 0 0
```

a. Use a regression model

to study the effects of trading days on the equal-weighted index returns. What is the fitted model? Are the weekday effects significant in the returns at the 5% level?

What is the fitted model?

```
ml = lm(ew ~ M+T+W+R+F, data=data4)
summary(ml)
```

```
##
## Call:
## lm(formula = ew \sim M + T + W + R + F, data = data4)
## Residuals:
               1Q
      Min
                   Median
                              3Q
                                     Max
## -0.102962 -0.003094 0.000533 0.003795 0.108319
##
## Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0022386 0.0002155 10.389 < 2e-16 ***
           ## M
## T
           ## W
           ## R
## F
                 NA
                         NA
                                NA
                                       NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008234 on 7314 degrees of freedom
## Multiple R-squared: 0.01618, Adjusted R-squared: 0.01564
## F-statistic: 30.06 on 4 and 7314 DF, p-value: < 2.2e-16
```

step(ml)

```
## Start: AIC=-70250.08
## ew \sim M + T + W + R + F
##
##
## Step: AIC=-70250.08
## ew \sim M + T + W + R
##
         Df Sum of Sq
##
                        RSS AIC
## <none>
                      0.49586 -70250
## - W
        1 0.0007675 0.49663 -70241
## - R
          1 0.0007762 0.49664 -70241
## - T
        1 0.0028923 0.49876 -70210
## - M
         1 0.0071735 0.50304 -70147
```

```
ml2 = lm(ew ~ M+T+W+F+R, data=data4)
summary(ml2)
```

```
##
## Call:
## lm(formula = ew \sim M + T + W + F + R, data = data4)
## Residuals:
                  1Q
        Min
                       Median
                                     3Q
                                             Max
## -0.102962 -0.003094 0.000533 0.003795 0.108319
##
## Coefficients: (1 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0012092 0.0002148 5.631 1.86e-08 ***
             -0.0021440 0.0003080 -6.961 3.67e-12 ***
## M
## T
            0.0000109 0.0003022 0.036 0.971239
## W
              0.0010294 0.0003042 3.384 0.000719 ***
## F
## R
                     NA
                               NA
                                      NA
                                              NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008234 on 7314 degrees of freedom
## Multiple R-squared: 0.01618, Adjusted R-squared: 0.01564
## F-statistic: 30.06 on 4 and 7314 DF, p-value: < 2.2e-16
```

step(ml2)

```
## Start: AIC=-70250.08
## ew \sim M + T + W + F + R
##
##
## Step: AIC=-70250.08
## ew \sim M + T + W + F
##
##
         Df Sum of Sq
                          RSS
## - W
         1 0.0000001 0.49586 -70252
## <none>
                      0.49586 -70250
## - T
         1 0.0006673 0.49653 -70242
## - F
         1 0.0007762 0.49664 -70241
## - M
          1 0.0032852 0.49915 -70204
##
## Step: AIC=-70252.08
## ew \sim M + T + F
##
##
         Df Sum of Sq
                          RSS
                                 AIC
## <none>
                      0.49586 -70252
## - T
         1 0.0009061 0.49677 -70241
## - F
          1 0.0010262 0.49689 -70239
## - M
          1 0.0043768 0.50024 -70190
```

```
ml3 = lm(ew ~ M+T+F+R+W, data=data4)
summary(ml3)
```

```
##
## Call:
## lm(formula = ew \sim M + T + F + R + W, data = data4)
## Residuals:
##
        Min
                  1Q
                       Median
                                    3Q
                                            Max
## -0.102962 -0.003094 0.000533 0.003795 0.108319
## Coefficients: (1 not defined because of singularities)
##
               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0012201 0.0002126 5.739 9.91e-09 ***
             ## M
## T
             -0.0009593 0.0003008 -3.190 0.00143 **
             0.0010185 0.0003027 3.365 0.00077 ***
## F
             -0.0000109 0.0003022 -0.036 0.97124
## R
## W
                    NA
                              NA
                                   NA
                                              NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 0.008234 on 7314 degrees of freedom
## Multiple R-squared: 0.01618, Adjusted R-squared: 0.01564
## F-statistic: 30.06 on 4 and 7314 DF, p-value: < 2.2e-16
```

```
step(ml3)
```

```
## Start: AIC=-70250.08
## ew \sim M + T + F + R + W
##
##
## Step: AIC=-70250.08
## ew \sim M + T + F + R
##
      Df Sum of Sq RSS AIC
##
       1 0.0000001 0.49586 -70252
## - R
## <none>
         0.49586 -70250
## - M 1 0.0033513 0.49922 -70203
##
## Step: AIC=-70252.08
## ew \sim M + T + F
##
    Df Sum of Sq RSS AIC
##
## <none>
                 0.49586 -70252
## - T 1 0.0009061 0.49677 -70241
## - F
       1 0.0010262 0.49689 -70239
## - M 1 0.0043768 0.50024 -70190
```

```
ml4 = lm(ew ~ M+W+F+R+T, data=data4)
summary(ml4)
```

```
##
## Call:
## lm(formula = ew \sim M + W + F + R + T, data = data4)
## Residuals:
                 1Q Median
       Min
                                    3Q
                                            Max
## -0.102962 -0.003094 0.000533 0.003795 0.108319
##
## Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0002608 0.0002127 1.226 0.22028
            ## M
## W
            0.0009593 0.0003008 3.190 0.00143 **
             0.0019778 0.0003028 6.532 6.94e-11 ***
## F
              0.0009484 0.0003023 3.137 0.00171 **
## R
## T
                    NA
                              NA
                                     NA
                                             NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008234 on 7314 degrees of freedom
## Multiple R-squared: 0.01618, Adjusted R-squared: 0.01564
## F-statistic: 30.06 on 4 and 7314 DF, p-value: < 2.2e-16
```

step(ml4)

```
## Start: AIC=-70250.08
## ew \sim M + W + F + R + T
##
##
## Step: AIC=-70250.08
## ew \sim M + W + F + R
##
         Df Sum of Sq
##
                        RSS AIC
## <none>
                      0.49586 -70250
## - R
        1 0.00066734 0.49653 -70242
## - W
          1 0.00068972 0.49655 -70242
## - M 1 0.00103096 0.49689 -70237
## - F
        1 0.00289230 0.49876 -70210
```

```
ml5 = lm(ew ~ T+W+F+R+M, data=data4)
summary(ml5)
```

```
##
## Call:
## lm(formula = ew \sim T + W + F + R + M, data = data4)
##
## Residuals:
                 1Q
##
        Min
                       Median
                                    3Q
                                            Max
## -0.102962 -0.003094 0.000533 0.003795 0.108319
##
## Coefficients: (1 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
0.0011956 0.0003066 3.900 9.72e-05 ***
## T
## W
            0.0021549 0.0003065 7.031 2.24e-12 ***
              0.0031734 0.0003085 10.286 < 2e-16 ***
## F
## R
              0.0021440 0.0003080 6.961 3.67e-12 ***
## M
                    NA
                              NA
                                      NA
                                              NA
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008234 on 7314 degrees of freedom
## Multiple R-squared: 0.01618, Adjusted R-squared: 0.01564
## F-statistic: 30.06 on 4 and 7314 DF, p-value: < 2.2e-16
```

step(ml5)

```
## Start: AIC=-70250.08
## ew \sim T + W + F + R + M
##
##
## Step: AIC=-70250.08
## ew \sim T + W + F + R
##
##
         Df Sum of Sq
                        RSS AIC
## <none>
                      0.49586 -70250
## - T
          1 0.0010310 0.49689 -70237
## - R
          1 0.0032852 0.49915 -70204
## - W
         1 0.0033513 0.49922 -70203
## - F
         1 0.0071735 0.50304 -70147
```

• AIC 最低的模型:Im(formula = ew ~ M + T + F, data = data4)

```
best = lm(formula = ew ~ M + T + F, data = data4)
summary(best)
```

```
##
## Call:
## lm(formula = ew \sim M + T + F, data = data4)
## Residuals:
                1Q
##
       Min
                     Median
                                 3Q
                                         Max
## -0.102962 -0.003094 0.000533 0.003792 0.108319
##
## Coefficients:
             Estimate Std. Error t value Pr(>|t|)
##
## (Intercept) 0.0012147 0.0001511 8.040 1.04e-15 ***
            ## M
            ## F
            0.0010239 0.0002632 3.891 0.000101 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 0.008233 on 7315 degrees of freedom
## Multiple R-squared: 0.01617, Adjusted R-squared: 0.01577
## F-statistic: 40.09 on 3 and 7315 DF, p-value: < 2.2e-16
```

Are the weekday effects significant in the returns at the 5% level?

- 星期一、二、五在上面每個模型中, p-value 均小於0.05, 故這幾天顯著影響報酬率
- 但其實大部分的模型每天對報酬率都有顯著影響·AIC間也變化不大
- b. Use the HAC estimator of the covariance matrix to obtain the t ratio of regression estimates. Does the HAC estimator change the conclusion of weekday effects?

```
coeftest(best, vcov=vcovHAC(best))
```

- By using the HAC estimator, it doesn't change the conclusion of weekday effects.
- 仍是 M、T、F 影響顯著

2

Now consider similar questions of the previous exercise for the IBM stock returns.(d-ibm3dxwkdays8008.txt)

c. Refine the above model by using the technique of regression model with time series errors. In there a significant weekday effect based on the refined model?

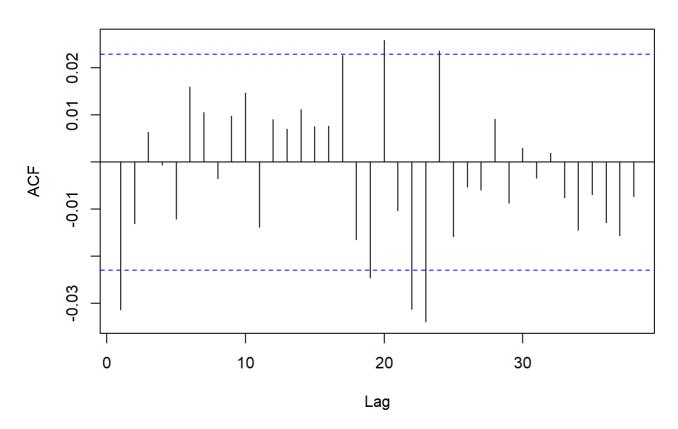
```
ibm = lm(ibm ~ M + T + W + R + F ,data=data4)
summary(ibm)
```

```
##
## Call:
## lm(formula = ibm \sim M + T + W + R + F, data = data4)
##
## Residuals:
##
        Min
                   1Q
                         Median
                                       3Q
                                               Max
## -0.231629 -0.009290 -0.000036 0.008840 0.131619
##
## Coefficients: (1 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
##
## (Intercept) -0.0005902 0.0004671 -1.264 0.206382
             0.0025896 0.0006687 3.873 0.000109 ***
## M
               0.0020296 0.0006563 3.092 0.001992 **
## T
## W
               0.0002289 0.0006561 0.349 0.727217
               0.0006073 0.0006594 0.921 0.357085
## R
## F
                      NA
                                 NA
                                        NA
                                                 NA
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01785 on 7314 degrees of freedom
## Multiple R-squared: 0.003269, Adjusted R-squared: 0.002723
## F-statistic: 5.996 on 4 and 7314 DF, p-value: 8.178e-05
```

- ullet model: $R_t = -0.0005902 + 0.0025896M + 0.0020296T + 0.0002289W + 0.0006073R$
- 星期一和星期二在不考慮星期五影響下 p-value < 0.05, 顯著影響報酬率

```
acf(ibm$residuals)
```

Series ibm\$residuals



serial correlations in the residuals

• Use Q(12) to perform the test

```
Box.test(ibm$residuals,lag=12,type='Ljung')
```

```
##
## Box-Ljung test
##
## data: ibm$residuals
## X-squared = 16.845, df = 12, p-value = 0.1555
```

• NO · 因為 p-value > 0.05 · 不拒絕 H0 (無序列相關)

```
eacf(ibm$residuals)
```

```
##
## Call:
## arima(x = data4$ibm, order = c(0, 0, 1), xreg = data4$M + data4$T)
##
## Coefficients:
## ma1 intercept xreg
## -0.0323 -3e-04 2e-03
## s.e. 0.0118 3e-04 4e-04
##
## sigma^2 estimated as 0.000318: log likelihood = 19086.27, aic = -38166.53
```

rbind(estmodel1\$coef-2*sqrt(diag(estmodel1\$var.coef)),estmodel1\$coef+2*sqrt(diag(estmodel1\$var.coef)))

```
## ma1 intercept xreg
## [1,] -0.055990572 -0.0008361281 0.001170257
## [2,] -0.008665408 0.0002139907 0.002869598
```

```
##
## Call:
## arima(x = data4$ibm, order = c(0, 0, 1), xreg = data4$M + data4$T, fixed = c(NA),
##
      0, NA))
##
## Coefficients:
##
            ma1 intercept xreg
##
        -0.0322
                         0 0.0017
## s.e. 0.0118
                         0 0.0003
##
## sigma^2 estimated as 0.0003181: log likelihood = 19085.57, aic = -38167.15
```

```
Box.test(estmodel2$residuals, lag=12, type="Ljung") fifoff - >
```

```
##
## Box-Ljung test
##
## data: estmodel2$residuals
## X-squared = 9.4467, df = 12, p-value = 0.6644
```

estmodel2\$coef

```
## ma1 intercept xreg
## -0.03224112 0.00000000 0.00169704
```

estmodel2\$var.coef

```
## ma1 xreg
## ma1 1.399566e-04 6.153852e-08
## xreg 6.153852e-08 1.068652e-07
```

$$H_0: eta=0$$

$$t = rac{\hat{eta} - 0}{stdev(\hat{eta})}$$

```
t_ibm = estmodel2$coef[3]/sqrt(estmodel2$var.coef[2,2])
t_ibm
```

```
## xreg
## 5.191273
```

```
p_value = 2*(1-pnorm(t_ibm))
cbind(test_statistic=t_ibm, p_value=p_value)
```

```
## test_statistic p_value
## xreg 5.191273 2.088611e-07
```

• There is a significant weekday effect based on the refined model since p-value < 0.05

3.

Again, consider the two bond yield series, that is, Aaa and Baa. What is the relationship between the two series? To answer this question, build a time series model using yields of Aaa bonds as the dependent variable and yields of Baa bonds as independent variable

```
Aaa = read.table("C:/Users/user/Desktop/time_series/HW/w-Aaa.txt",header = F)
head(Aaa)
```

```
## V1 V2 V3 V4
## 1 1962 1 5 4.43
## 2 1962 1 12 4.42
## 3 1962 1 19 4.42
## 4 1962 1 26 4.41
## 5 1962 2 2 4.42
## 6 1962 2 9 4.42
```

```
Baa = read.table("C:/Users/user/Desktop/time_series/HW/w-Baa.txt",header = F)
head(Baa)
```

```
## V1 V2 V3 V4

## 1 1962 1 5 5.11

## 2 1962 1 12 5.09

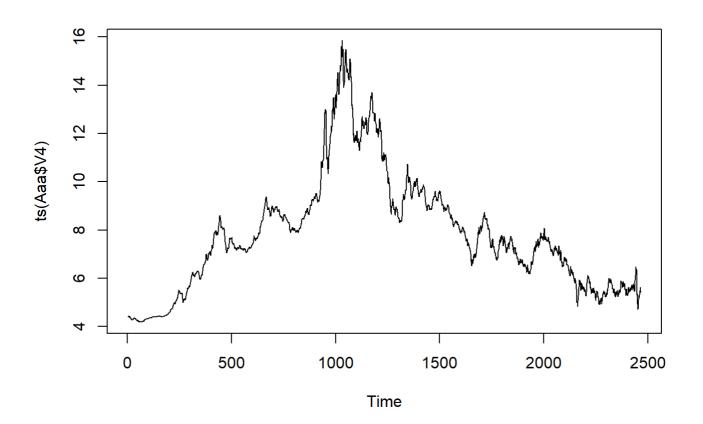
## 3 1962 1 19 5.08

## 4 1962 1 26 5.08

## 5 1962 2 2 5.07

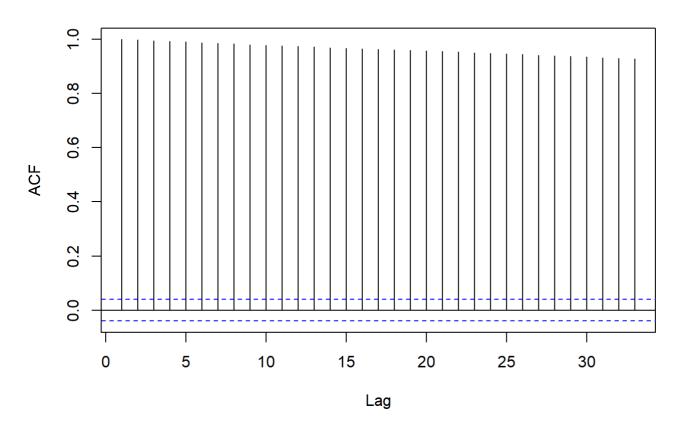
## 6 1962 2 9 5.08
```

```
plot(ts(Aaa$V4))
```

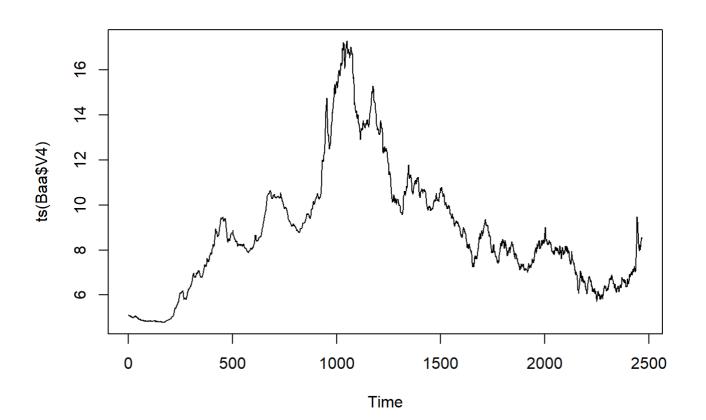


acf(Aaa\$V4)

Series Aaa\$V4

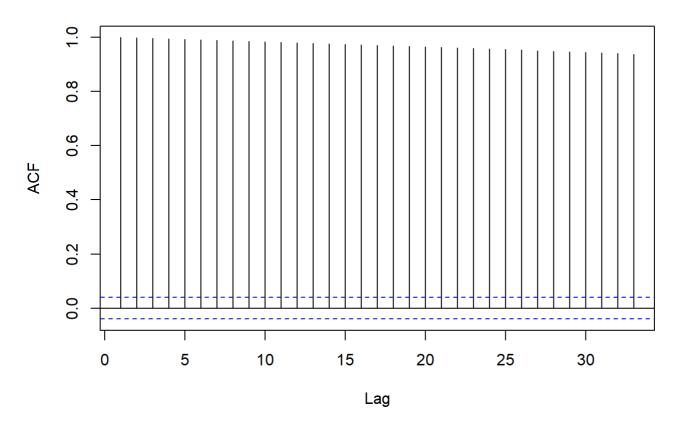


plot(ts(Baa\$V4))



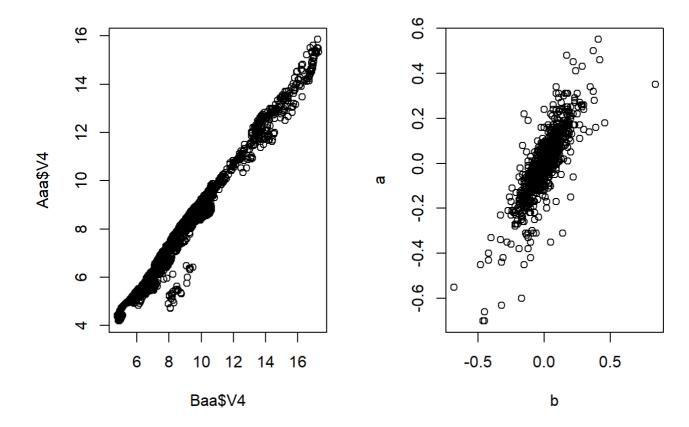
acf(Baa\$V4)

Series Baa\$V4



```
a = diff(Aaa$V4)
b = diff(Baa$V4)

par(mfrow=c(1,2))
plot(Baa$V4,Aaa$V4)
plot(b,a)
```

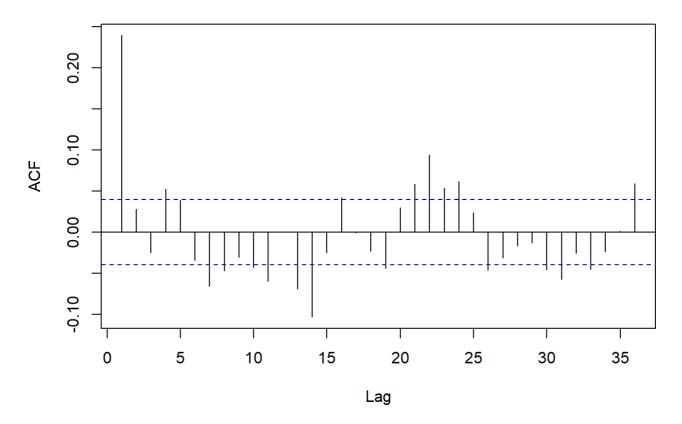


```
reg.fit = lm(a ~ -1+b)
summary(reg.fit)
```

```
##
## Call:
## lm(formula = a \sim -1 + b)
##
## Residuals:
##
       Min
                 1Q
                      Median
                                   3Q
                                           Max
## -0.44475 -0.01838 0.00000 0.01946 0.36192
##
## Coefficients:
    Estimate Std. Error t value Pr(>|t|)
## b 0.94613
              0.01264
                          74.87 <2e-16 ***
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 0.05341 on 2465 degrees of freedom
## Multiple R-squared: 0.6946, Adjusted R-squared: 0.6944
## F-statistic: 5605 on 1 and 2465 DF, p-value: < 2.2e-16
```

```
acf(reg.fit$residuals, lag=36)
```

Series reg.fit\$residuals



```
m1 = arima(a, order=c(0,0,1), xreg=b, include.mean=F)
m1
```

```
##
## Call:
## arima(x = a, order = c(0, 0, 1), xreg = b, include.mean = F)
##
## Coefficients:
## mal xreg
## 0.2335 0.9436
## s.e. 0.0185 0.0132
##
## sigma^2 estimated as 0.00269: log likelihood = 3797.81, aic = -7591.62
```

```
rbind(m1$coef-2*sqrt(diag(m1$var.coef)),
    m1$coef+2*sqrt(diag(m1$var.coef)))
```

```
## ma1 xreg
## [1,] 0.1964290 0.9171835
## [2,] 0.2705052 0.9699801
```

```
Box.test(m1$residuals, lag=12, type="Ljung")
```

```
##
## Box-Ljung test
##
## data: m1$residuals
## X-squared = 43.245, df = 12, p-value = 2.052e-05
```

• p value < 0.05 · 拒絕 H0 · 序列相關 · 該模型不夠 · (不好)

```
eacf(m1$residuals)
```

```
## AR/MA
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 0 0 x x 0 0 x 0 0 0 x 0 x 0 x x
## 1 x 0 0 x x 0 0 0 0 0 0 0 x 0
## 2 x 0 0 x 0 0 0 0 0 0 0 x 0
## 3 x 0 x x 0 0 0 0 0 0 0 0 x 0
## 4 x x x x 0 x 0 0 0 0 0 0 x x
## 5 x x x x 0 0 0 0 0 0 0 0 x
## 6 x x x 0 0 0 0 0 0 0 0 0 x
## 7 x x x 0 0 0 0 0 0 0 0 0 x
```

arima(2,0,4)

```
m2 = arima(a, order=c(2,0,4), xreg=b, include.mean=F)
m2
```

```
##
## Call:
## arima(x = a, order = c(2, 0, 4), xreg = b, include.mean = F)
##
## Coefficients:
##
           ar1
                    ar2
                             ma1
                                           ma3
                                    ma2
                                                   ma4
                                                          xreg
##
        0.9238 -0.9649 -0.6831 0.793 0.1817 0.0879 0.9467
## s.e. 0.0114 0.0101
                          0.0231 0.025 0.0236 0.0206 0.0130
##
## sigma^2 estimated as 0.002631: log likelihood = 3825.26, aic = -7636.52
```

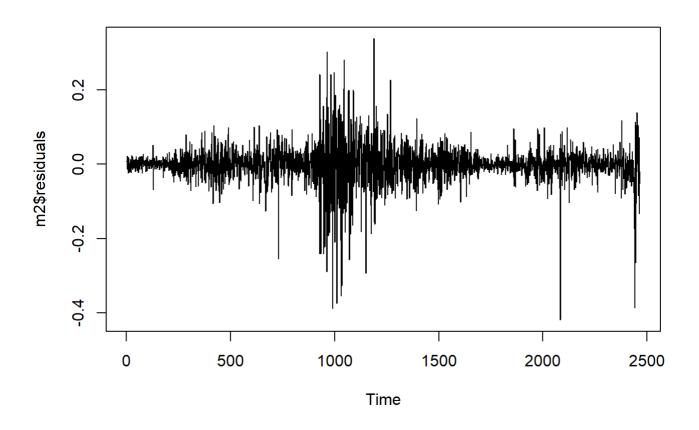
```
rbind(m2$coef-2*sqrt(diag(m2$var.coef)),
    m2$coef+2*sqrt(diag(m2$var.coef)))
```

```
## ar1 ar2 ma1 ma2 ma3 ma4 xreg
## [1,] 0.9009959 -0.9851984 -0.7293284 0.7430812 0.1344973 0.04675936 0.9206019
## [2,] 0.9465714 -0.9446953 -0.6368022 0.8429703 0.2289060 0.12906066 0.9727363
```

```
Box.test(m2$residuals, lag=12, type="Ljung")
```

```
##
## Box-Ljung test
##
## data: m2$residuals
## X-squared = 37.023, df = 12, p-value = 0.0002215
```

```
plot(m2$residuals)
```



adfTest(diff(b))

```
##
## Title:
##
   Augmented Dickey-Fuller Test
##
## Test Results:
##
    PARAMETER:
##
       Lag Order: 1
     STATISTIC:
##
       Dickey-Fuller: -53.0148
##
##
     P VALUE:
       0.01
##
##
## Description:
   Fri May 12 06:15:17 2023 by user: user
```

• p value<0.05 · 拒絕H0 · 序列相關 · 該模型不夠 · (不好) · 但有進步



$$H_0:eta=0$$

$$t = rac{\hat{eta} - 0}{stdev(\hat{eta})}$$

```
m2$coef
```

```
## ar1 ar2 ma1 ma2 ma3 ma4
## 0.92378369 -0.96494682 -0.68306530 0.79302577 0.18170167 0.08791001
## xreg
## 0.94666909
```

m2\$var.coef

```
##
                               ar2
                                            ma1
                                                          ma2
                                                                        ma3
        1.298204e-04 -2.631201e-05 -1.274029e-04 8.196282e-06
## ar1
                                                               3.809453e-05
## ar2 -2.631201e-05 1.025313e-04 3.223757e-05 -8.842491e-05 1.174010e-05
## ma1 -1.274029e-04 3.223757e-05 5.350685e-04 -2.836087e-04 2.435095e-04
        8.196282e-06 -8.842491e-05 -2.836087e-04 6.236151e-04 -5.428976e-04
## ma2
        3.809453e-05 1.174010e-05 2.435095e-04 -5.428976e-04 5.570630e-04
## ma3
## ma4 -3.603850e-05 3.100050e-05 1.566689e-04 2.498775e-04 -2.674465e-04
## xreg 2.077260e-06 1.153685e-06 -3.485335e-06 2.443752e-05 -2.310582e-05
##
                 ma4
                              xreg
## ar1 -3.603850e-05 2.077260e-06
        3.100050e-05 1.153685e-06
## ar2
## ma1 1.566689e-04 -3.485335e-06
        2.498775e-04 2.443752e-05
## ma2
## ma3 -2.674465e-04 -2.310582e-05
        4.233441e-04 2.372395e-05
## ma4
## xreg 2.372395e-05 1.698752e-04
```

```
t_ibm2 = m2$coef[7]/sqrt(m2$var.coef[7,7])
t_ibm2
```

```
## xreg
## 72.63286
```

```
p_value2 = 2*(1-pnorm(t_ibm2))
cbind(test_statistic=t_ibm2, p_value=p_value2)
```

```
## test_statistic p_value
## xreg 72.63286 0
```

• p-value為0, Aaa、Baa H(A)C 顯著相關