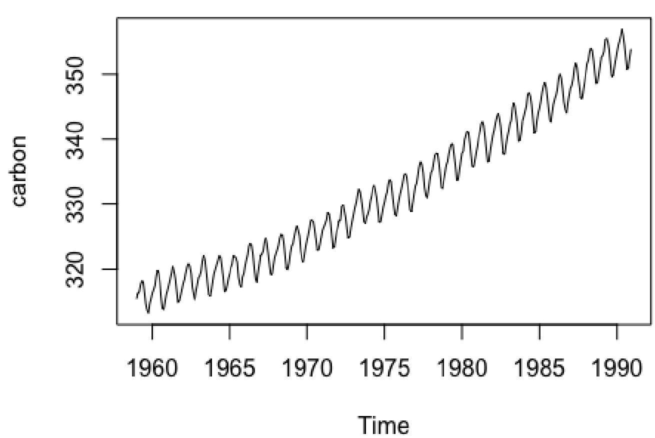
R graphs and codes

Shuoyang Shi

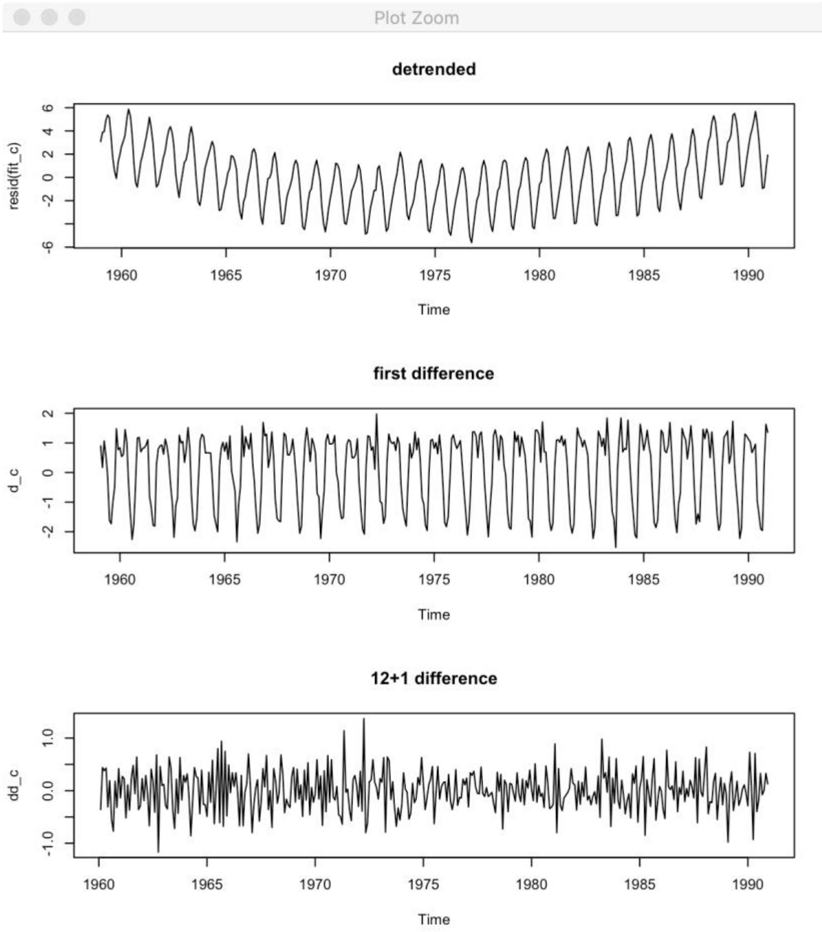
12/22/2020

Data 1: (univariate) "Carbon dioxide emissions in Hawaii.csv"

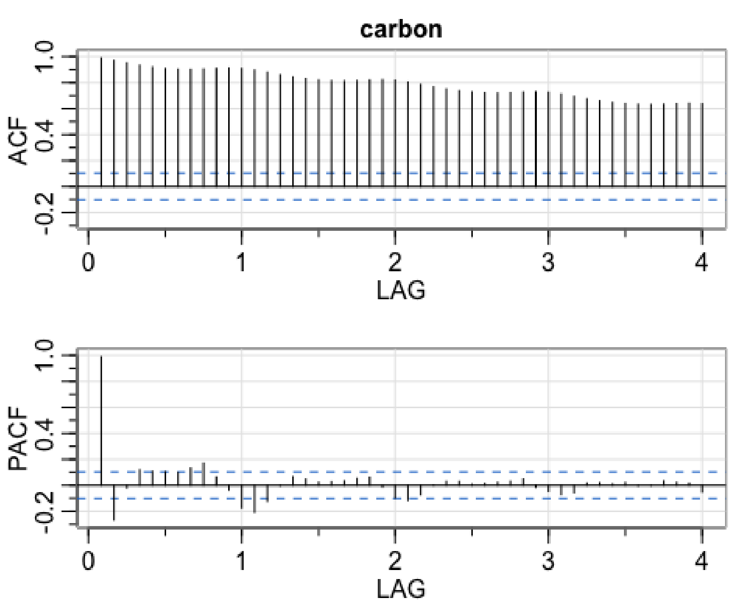
```
data_carbon <- read.csv("Carbon dioxide emissions in Hawaii.csv")
carbon <- data_carbon$Carbondioxide
carbon <- ts(carbon, start=c(1959,1), frequency=12)
ts.plot(carbon)
```



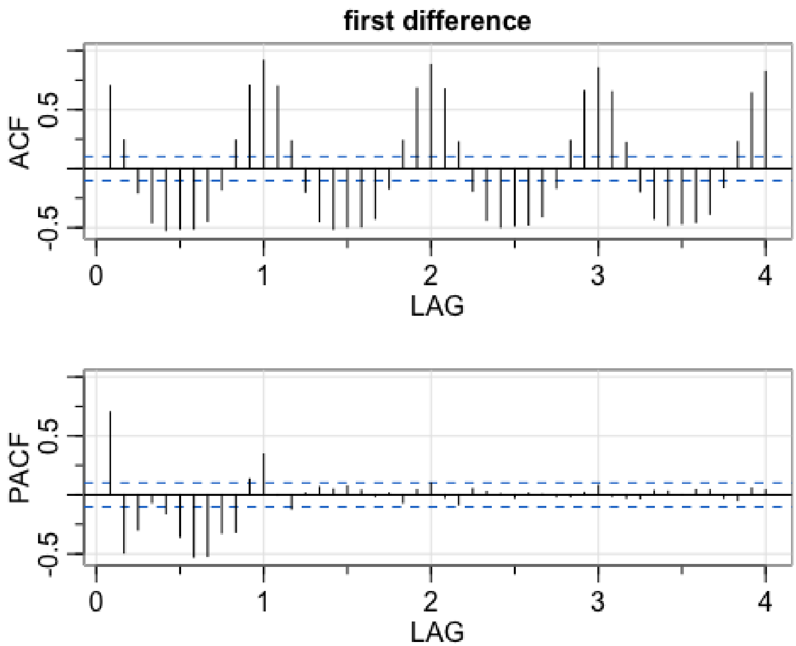

```
### (1) use an (S)ARIMA approach
fit_c <- lm(carbon ~ time(carbon), na.action=NULL)
d_c <- diff(carbon)
dd_c <- diff(diff(carbon, 12))
par(mfrow=c(3,1))
plot(resid(fit_c), main="detrended")
plot(d_c, main="first difference")
plot(dd_c, main="12+1 difference")
```



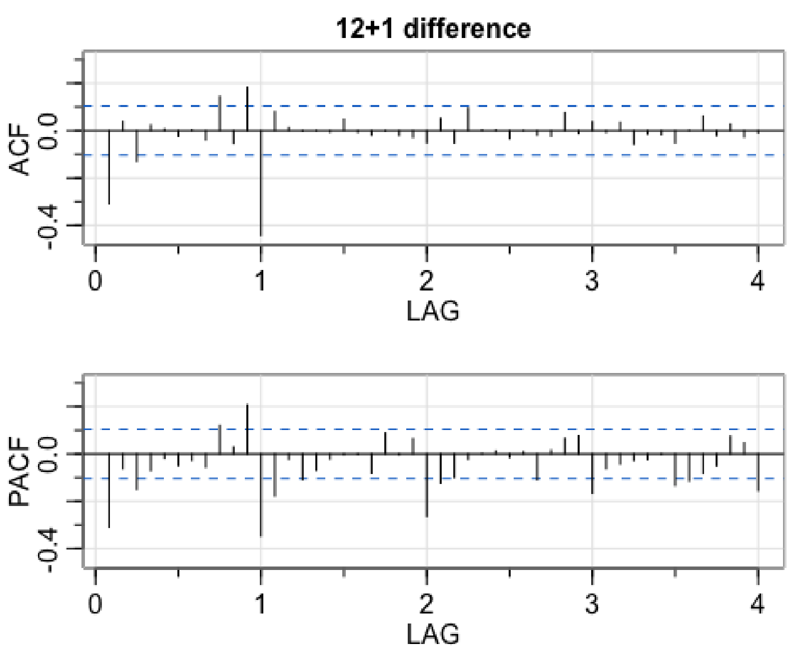

```r
# We go for differencing
# plot ACFs
acf2(carbon, 48, main="carbon")
```



```r
acf2(d_c, 48, main="first difference")
```

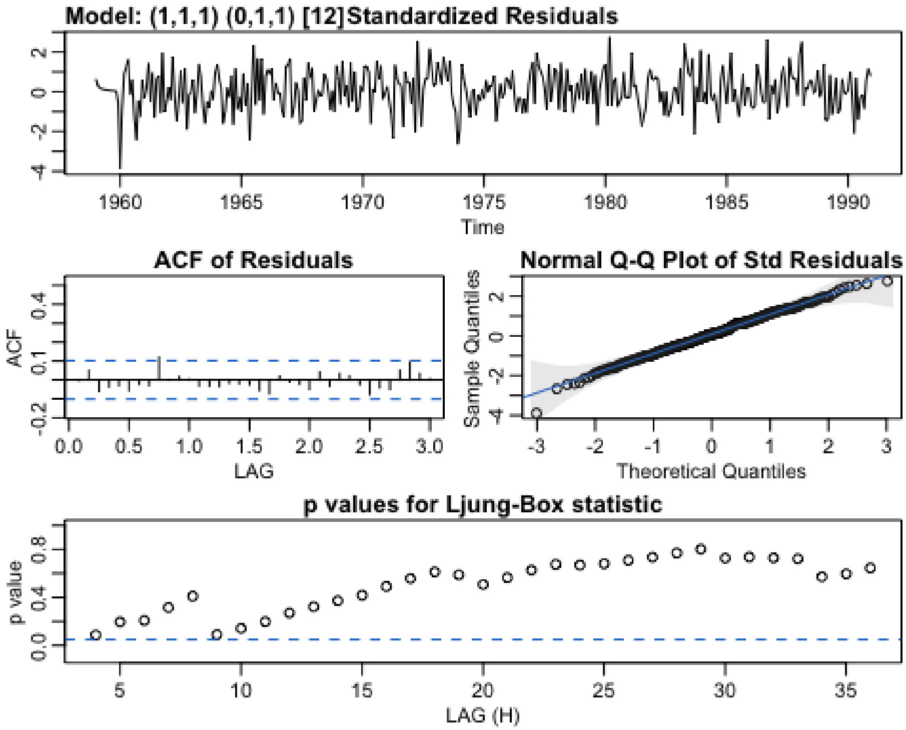



```r
acf2(dd_c, 48, main="12+1 difference")
```




```r
c_model1 <- sarima(carbon, 1, 1, 1, 1, 1, 1, 12)

c_model2 <- sarima(carbon, 1, 1, 1, 1, 1, 0, 12)

c_model3 <- sarima(carbon, 1, 1, 1, 0, 1, 1, 12)

##      Estimate     SE  t.value p.value
```

```
## ar1     0.2526 0.1503   1.6808  0.0936
## ma1    -0.5953 0.1286  -4.6299  0.0000
##sma1  -0.8568 0.0312 -27.4526  0.0000
```

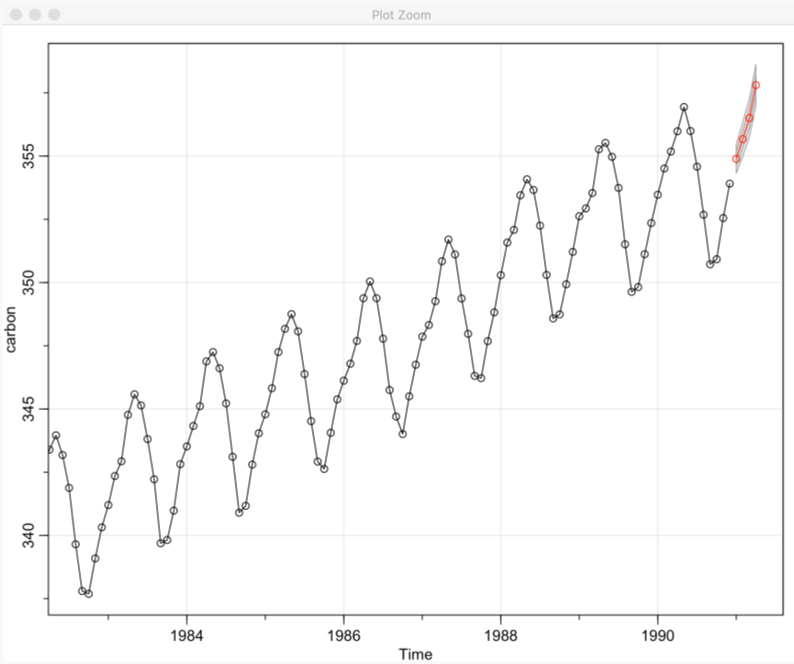



```
c_model4 <- sarima(carbon, 0, 1, 1, 0, 1, 1, 12)
c_model5 <- sarima(carbon, 1, 1, 0, 0, 1, 1, 12)
c_model1$AIC; c_model2$AIC; c_model3$AIC; c_model4$AIC; c_model5$AIC
## [1] 0.3698142
## [1] 0.6188375
## [1] 0.364627
## [1] 0.3647343
## [1] 0.3796195
c_model1$BIC; c_model2$BIC; c_model3$BIC; c_model4$BIC; c_model5$BIC
## [1] 0.4210734
## [1] 0.6598448
## [1] 0.4056344
## [1] 0.3954898
```

```
## [1] 0.410375
```

```r
sarima.for(carbon, 4, 1, 1, 1, 0, 1, 1, 12)
```

```
## $pred
##           Jan      Feb      Mar      Apr
## 1991 354.8905 355.6698 356.5030 357.8046
##
## $se
##            Jan       Feb       Mar       Apr
## 1991 0.2826180 0.3382069 0.3747055 0.4055480
```



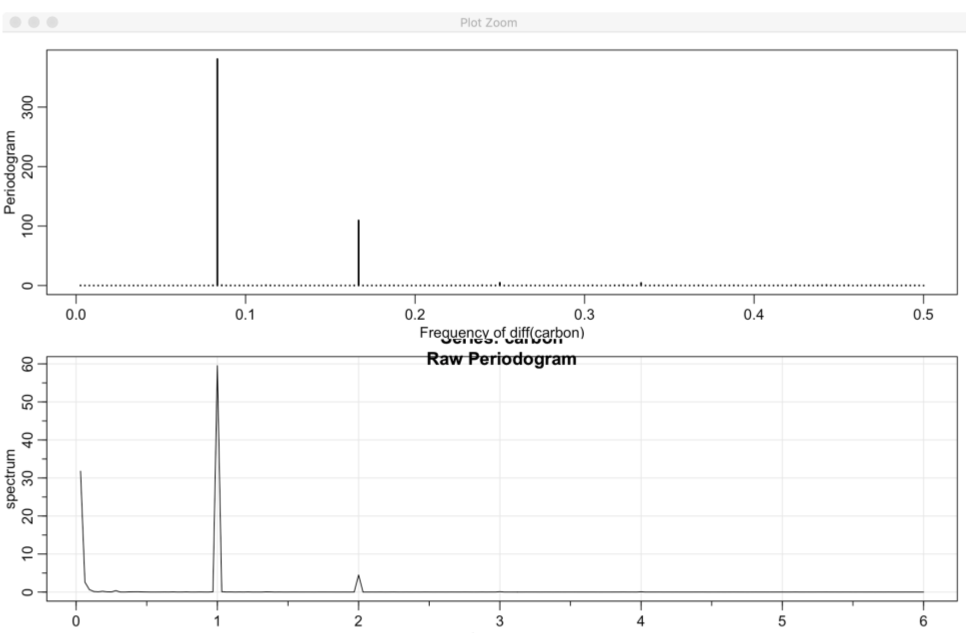

```r
### (2) unit root test

adf.test(resid(fit_c), k=0) # DF test
```

```
## Warning in adf.test(resid(fit_c), k = 0): p-value smaller than printed p-v
alue

##
##  Augmented Dickey-Fuller Test
##
## data:  resid(fit_c)
## Dickey-Fuller = -4.7364, Lag order = 0, p-value = 0.01
## alternative hypothesis: stationary
```

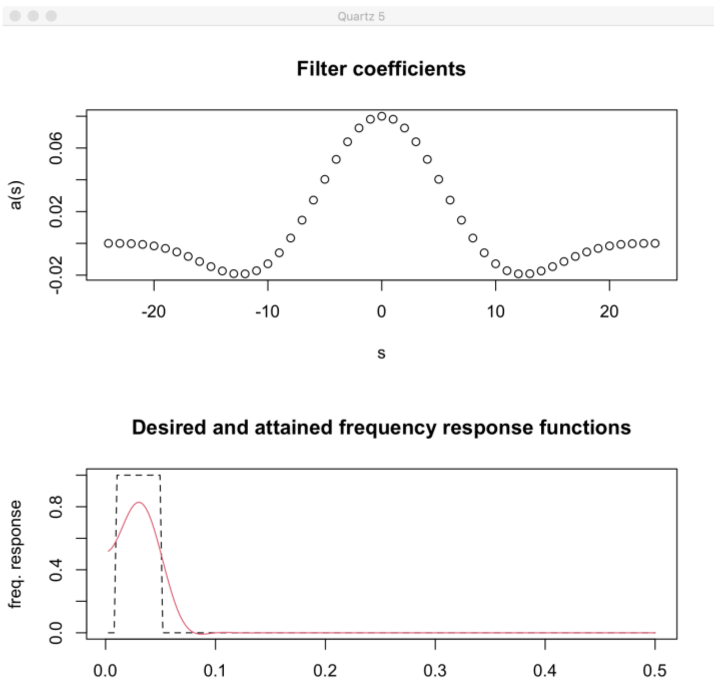
```r
adf.test(resid(fit_c)) # ADF test
```

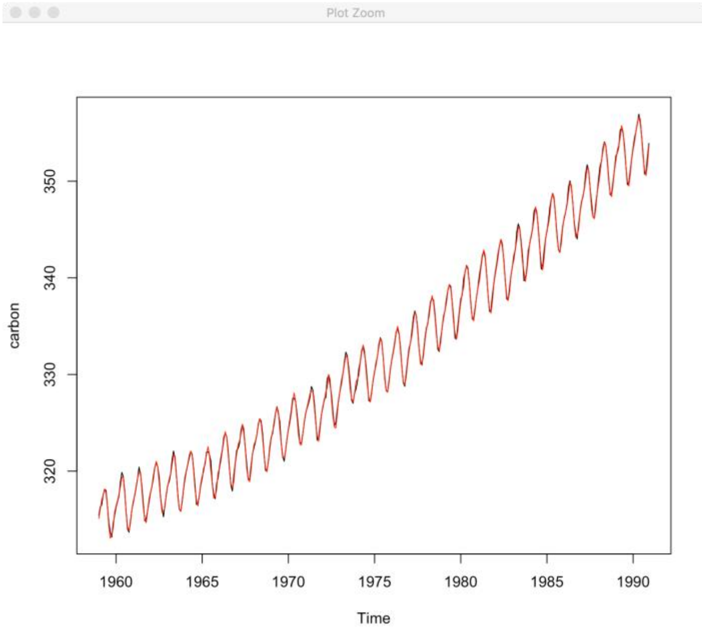
```
##
##  Augmented Dickey-Fuller Test
##
```

```
## data:  resid(fit_c)
## Dickey-Fuller = -2.428, Lag order = 7, p-value = 0.3964
## alternative hypothesis: stationary

pp.test(resid(fit_c)) # PP test

## Warning in pp.test(resid(fit_c)): p-value smaller than printed p-value

##
##  Phillips-Perron Unit Root Test
##
## data:  resid(fit_c)
## Dickey-Fuller Z(alpha) = -77.673, Truncation lag parameter = 5, p-value
## = 0.01
## alternative hypothesis: stationary


### (3) spectral analysis

# method 1
n <- length(carbon)
periodogram(d_c, xlab="Frequency of diff(carbon)")
omega <- order(periodogram(d_c, plot=FALSE)$spec, decreasing=T)[1:2]/n
(1/omega)/12 # years

## [1] 1.0 0.5



# method 2
mvspec(carbon, log='n')
```



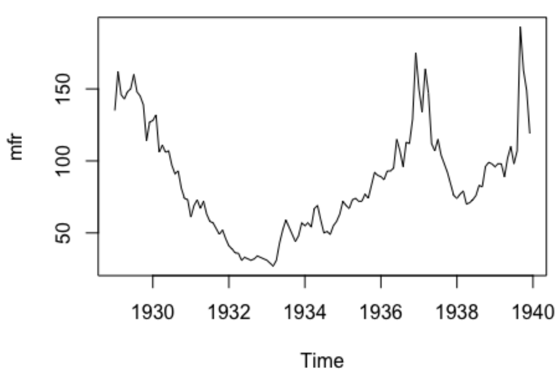
```
wave=cbind(cos((1:n)*2*pi*omega[1]),sin((1:n)*2*pi*omega[1]),
           cos((1:n)*2*pi*omega[2]),sin((1:n)*2*pi*omega[2]))
fit1=lm(carbon~wave[,1]+wave[,2]+wave[,3]+wave[,4])
summary(fit1)

plot(ts(fit1$residuals))

fitted_lm=fit1$fitted.values
fit0=sarima(carbon,1,1,1,xreg=ts(wave))

fitted_xreg=carbon-fit0$fit$residuals
fitted_signal=SigExtract(carbon)

plot(carbon)
lines(fitted_xreg,col="red")
```



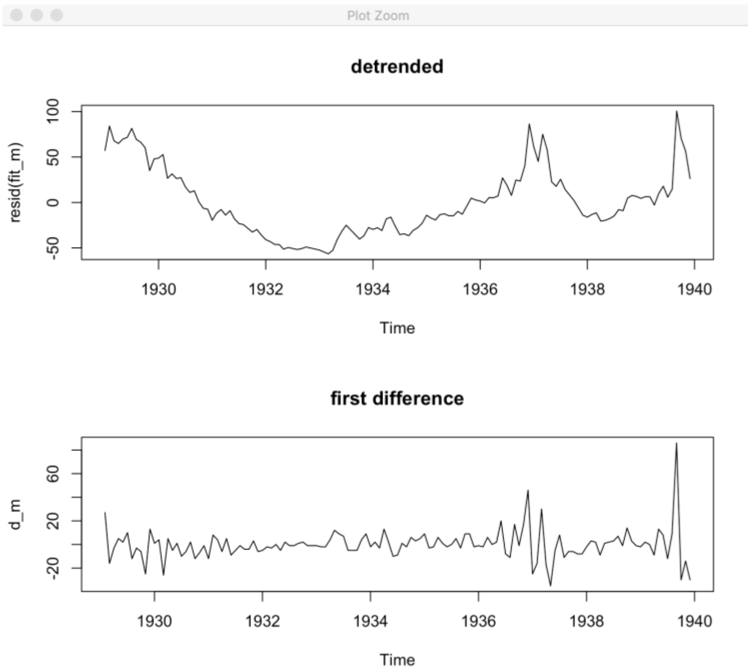


Data 2: (univariate) "manufacturers-index-of-new-order.xlsx"

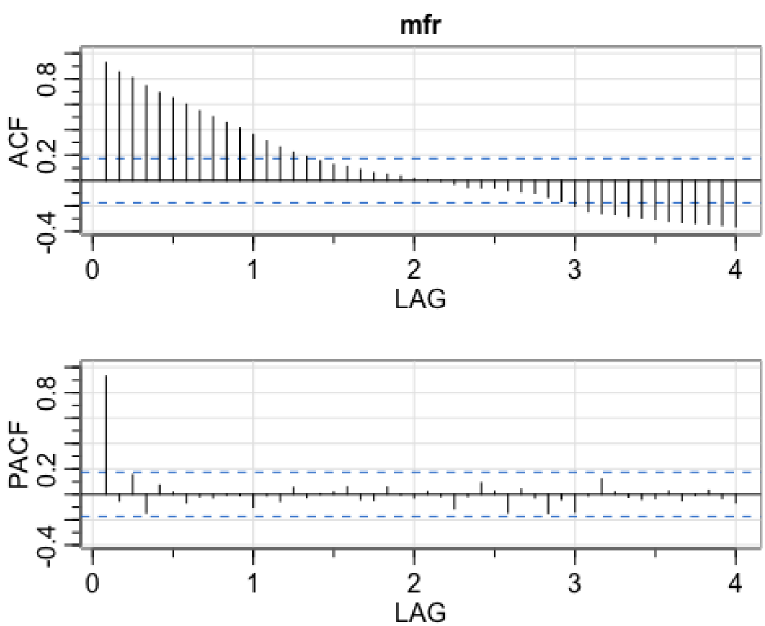
```
data_mfr <- read_excel("manufacturers-index-of-new-order.xlsx")
mfr <- data_mfr$`Manufacturers' Index of New Orders of Durable Goods for Unit
ed States`
mfr <- ts(mfr, start=c(1929,1), frequency=12)
ts.plot(mfr)
```



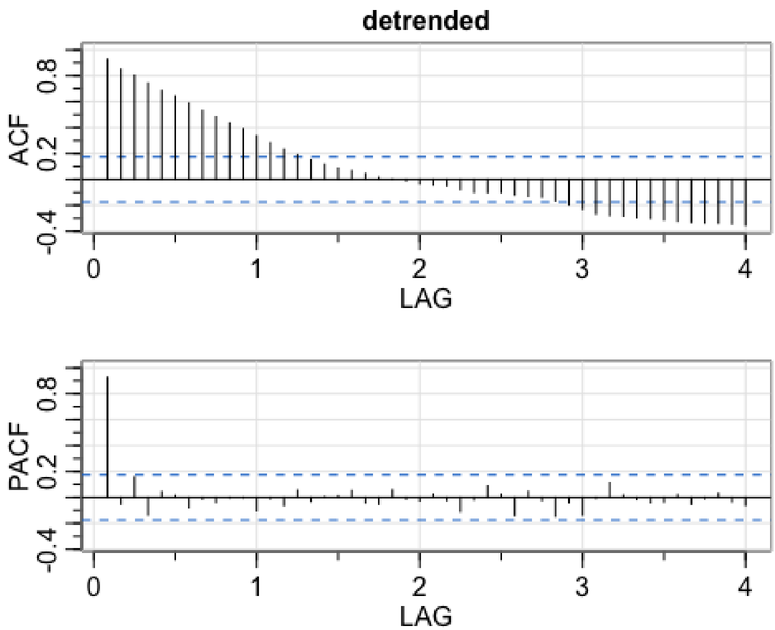

```
### (1) use an (S)ARIMA approach
fit_m <- lm(mfr ~ time(mfr), na.action=NULL)
d_m <- diff(mfr)
par(mfrow=c(2,1))
plot(resid(fit_m), main="detrended")
plot(d_m, main="first difference")
```



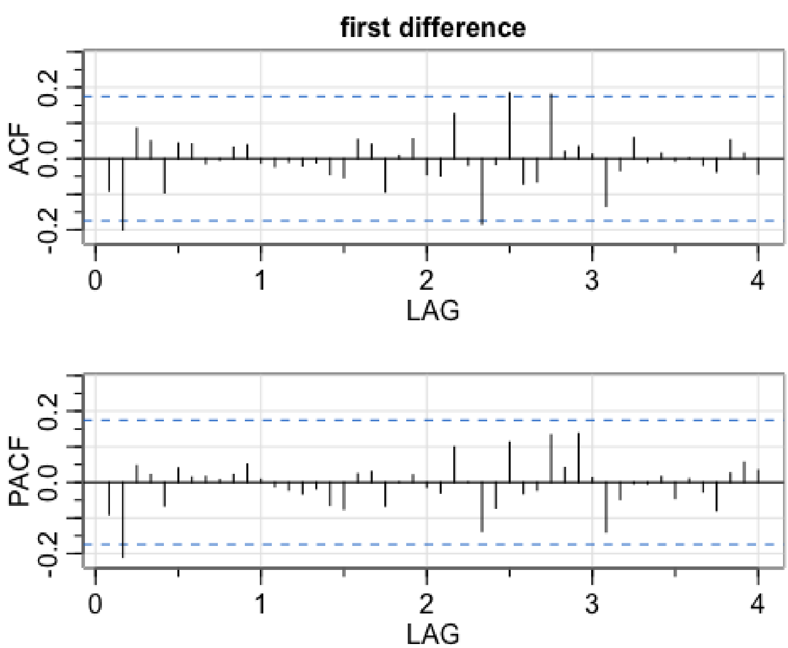

```r
# plot ACFs
acf2(mfr, main="mfr")
```



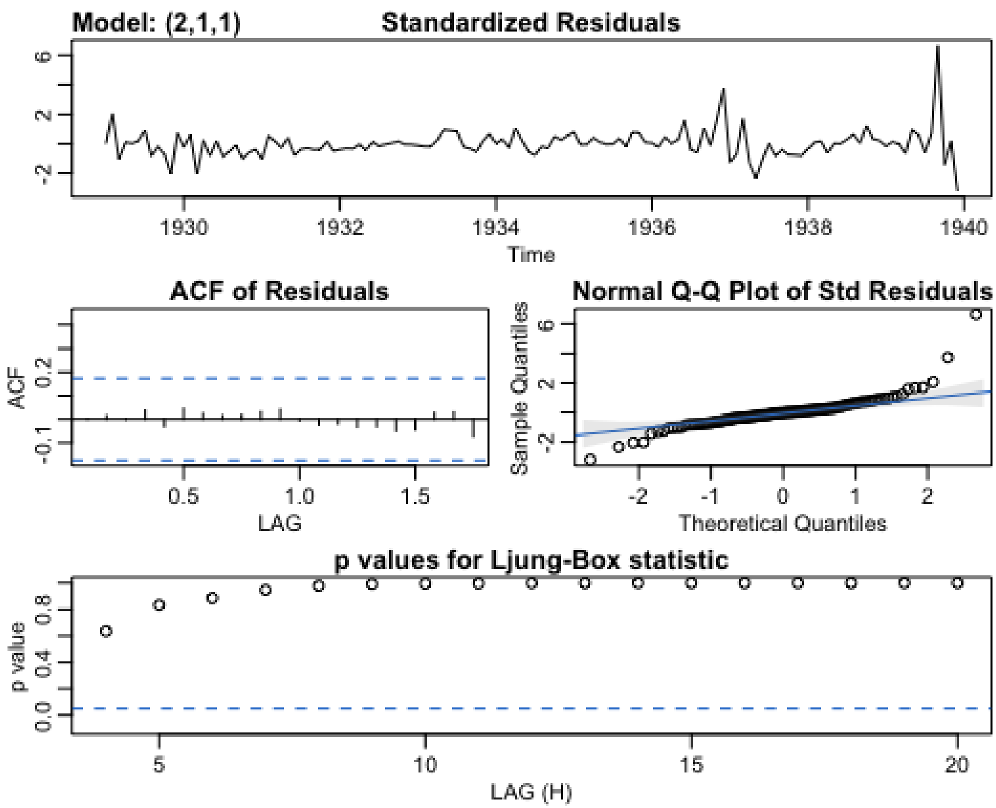

```r
acf2(resid(fit_m), main="detrended")
```




```r
acf2(d_m, main="first difference")
```

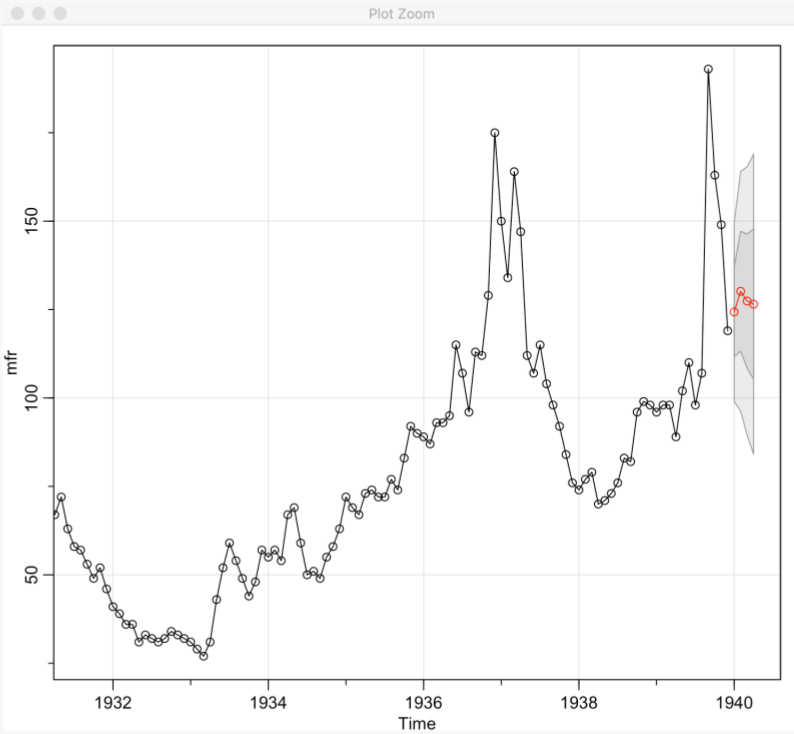



```r
m_model1 <- sarima(mfr, p=2, d=1, q=2)

m_model2 <- sarima(mfr, p=1, d=1, q=2)

m_model3 <- sarima(mfr, p=2, d=1, q=1)
```



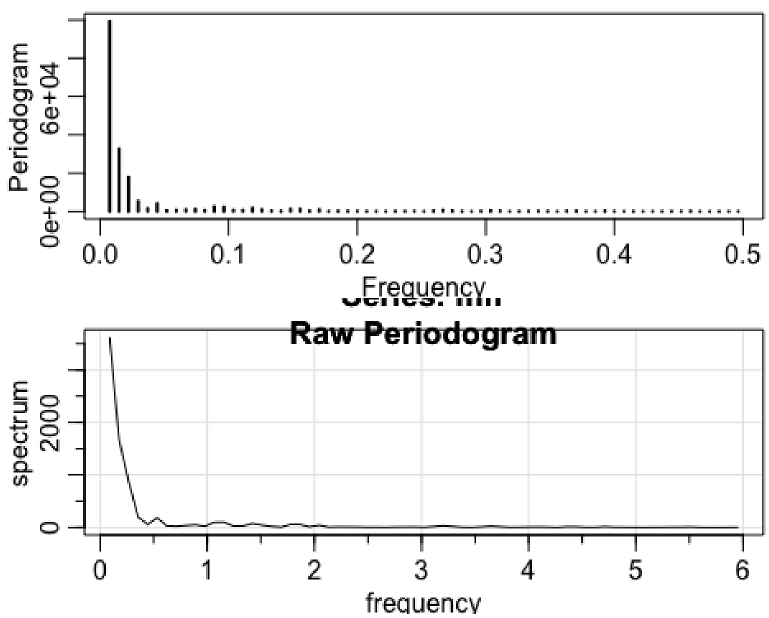

```
m_model4 <- sarima(mfr, p=1, d=1, q=1)

m_model1$AIC; m_model2$AIC; m_model3$AIC; m_model4$AIC

## [1] 8.005338

## [1] 7.999789

## [1] 7.992636

## [1] 8.00915

m_model1$BIC; m_model2$BIC; m_model3$BIC; m_model4$BIC

## [1] 8.137026

## [1] 8.109529

## [1] 8.102376

## [1] 8.096942

sarima.for(mfr, 4, p=2, d=1, q=1)

## $pred
##           Jan      Feb      Mar      Apr
## 1940 124.3165 130.1649 127.4162 126.5153
##
```

```
## $se
##            Jan      Feb      Mar      Apr
## 1940 12.66391 16.94342 18.97432 21.22082
```

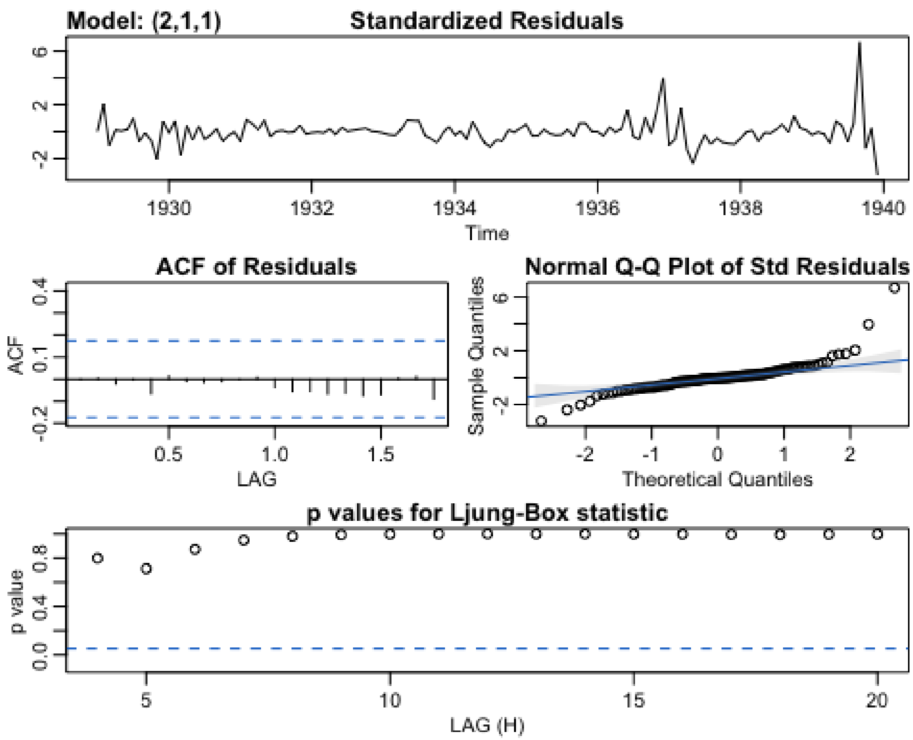



```
### (2) unit root test
adf.test(mfr, k=0) # DF test

##
##  Augmented Dickey-Fuller Test
##
## data:  mfr
## Dickey-Fuller = -2.2816, Lag order = 0, p-value = 0.4594
## alternative hypothesis: stationary

adf.test(mfr) # ADF test

##
##  Augmented Dickey-Fuller Test
##
## data:  mfr
## Dickey-Fuller = -2.2406, Lag order = 5, p-value = 0.4765
## alternative hypothesis: stationary

pp.test(mfr) # PP test

##
##  Phillips-Perron Unit Root Test
##
## data:  mfr
```

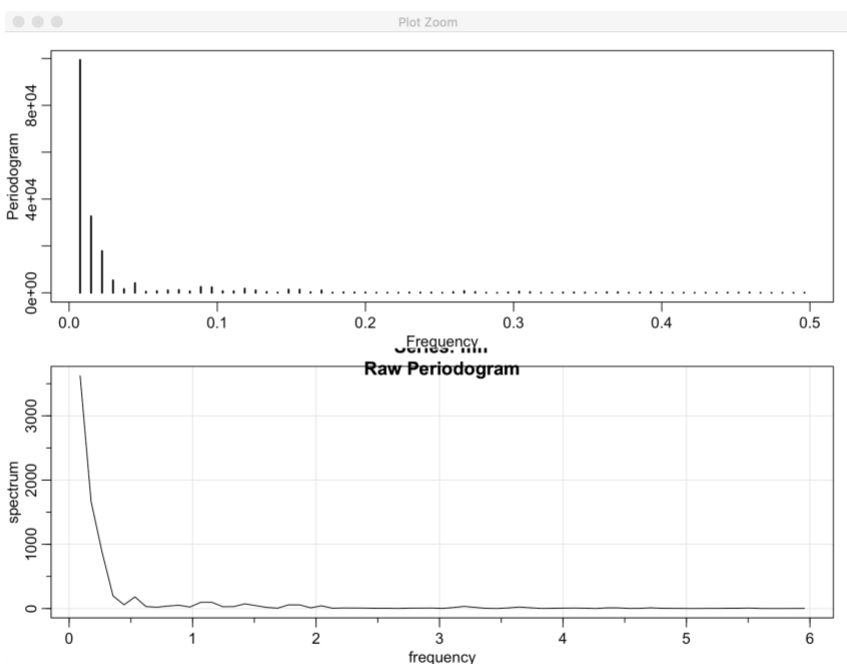
```
## Dickey-Fuller Z(alpha) = -7.2937, Truncation lag parameter = 4, p-value
## = 0.695
## alternative hypothesis: stationary


### (3) spectral analysis
# method 1
n <- length(mfr)
omega=order(periodogram(mfr)$spec,decreasing = T)[1:2]/n
(1/omega)/12 # years

## [1] 11.0  5.5

# method 2
mvspec(mfr, log='n')
```

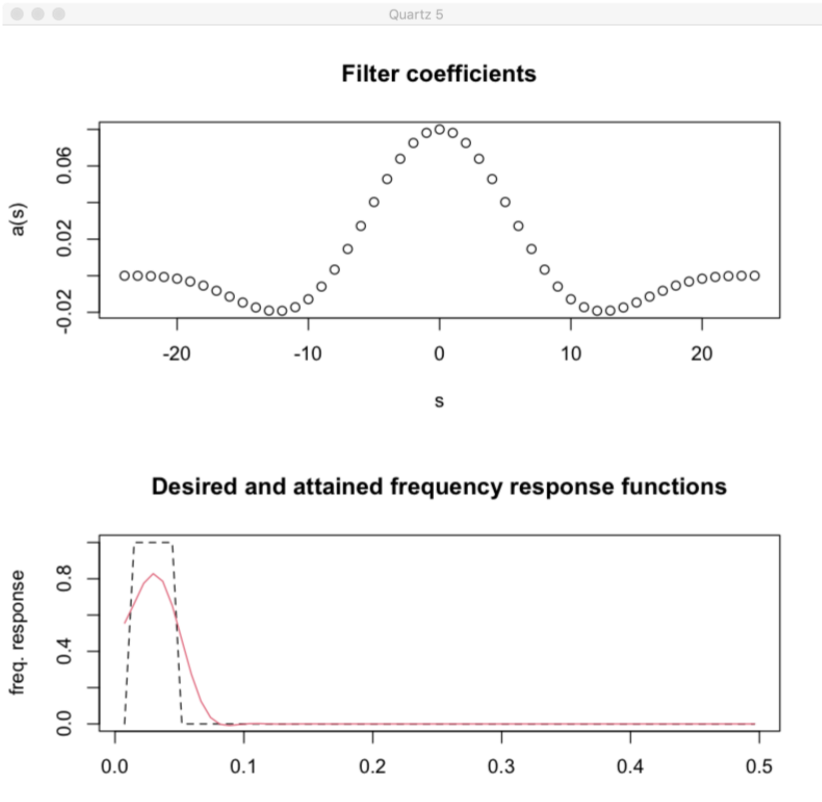



```
wave=cbind(cos((1:n)*2*pi*omega[1]),sin((1:n)*2*pi*omega[1]),
           cos((1:n)*2*pi*omega[2]),sin((1:n)*2*pi*omega[2]))
fit1=lm(mfr~wave[,1]+wave[,2]+wave[,3]+wave[,4])
summary(fit1)

plot(ts(fit1$residuals))

fitted_lm=fit1$fitted.values
fit0=sarima(mfr,2,1,1,xreg=ts(wave))
```

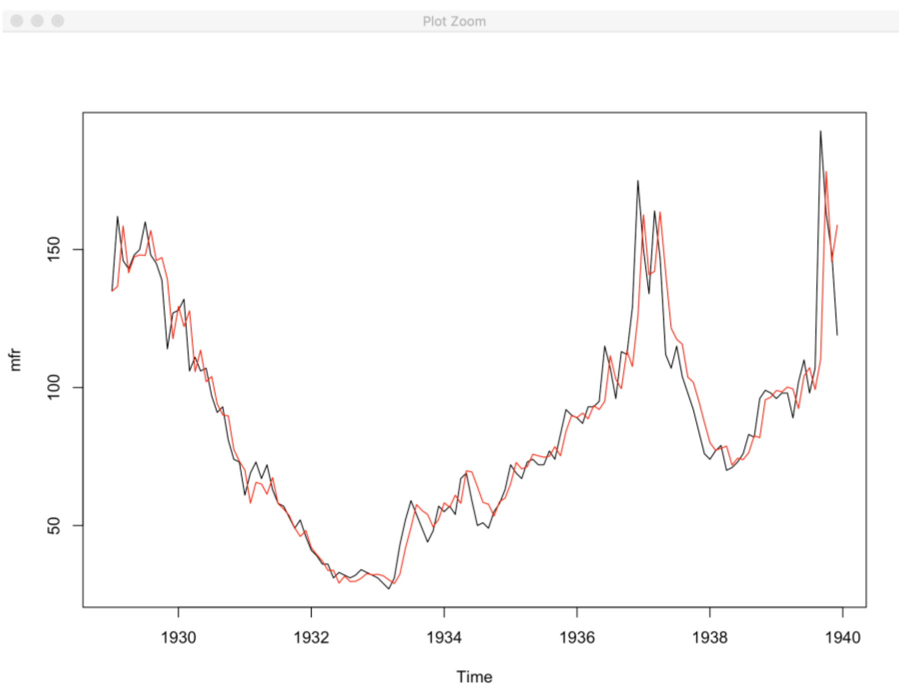



```
fitted_xreg=mfr-fit0$fit$residuals
fitted_signal=SigExtract(mfr)
```



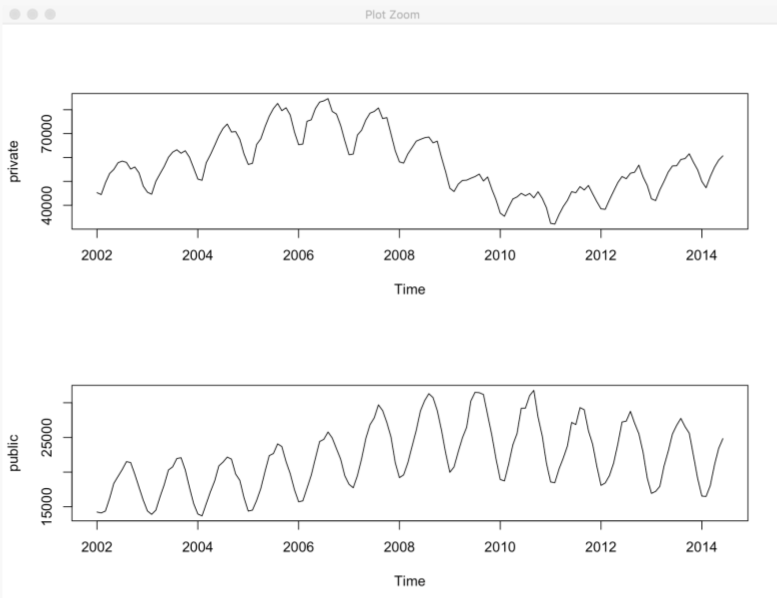



```
plot(mfr)
lines(fitted_xreg,col="red")
```

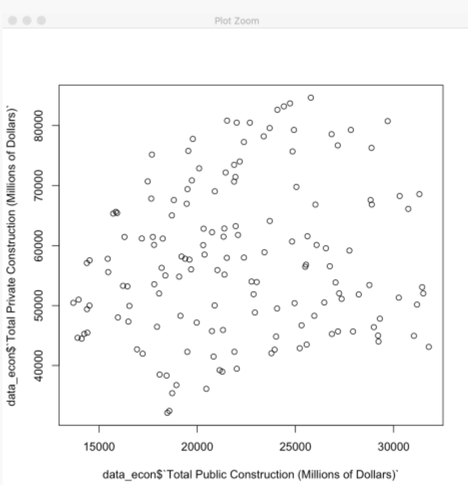


Data 3: **(multivariate) "economic-indicators-time-series.xlsx"**

```r
data_econ <- read_excel("economic-indicators-time-series.xlsx")
private <- data_econ$`Total Private Construction (Millions of Dollars)`
public <- data_econ$`Total Public Construction (Millions of Dollars)`
private <- ts(private, start=c(2002,1), frequency=12)
public <- ts(public, start=c(2002,1), frequency=12)
par(mfrow=c(2,1))
ts.plot(private)
ts.plot(public)
```




```r
### (1) linear regression with dependent errors

plot(x=data_econ$`Total Public Construction (Millions of Dollars)`,
     y=data_econ$`Total Private Construction (Millions of Dollars)`)
```




```r
trend <- time(public)
pub <- public-mean(public)
```

```r
pub2 <- pub^2
fit1 <- lm(private ~ trend + pub, na.action=NULL)
fit2 <- lm(private ~ trend + pub + pub2, na.action=NULL)
summary(aov(lm(private ~ cbind(trend, pub, pub2))))
```

```
##                           Df    Sum Sq   Mean Sq F value  Pr(>F)
## cbind(trend, pub, pub2)   3 8.879e+09 2.960e+09   28.99 9.2e-15 ***
## Residuals               146 1.491e+10 1.021e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
num <- length(private) # sample size
AIC(fit1)/num - log(2*pi) # AIC
```

```
## [1] 19.59591
```

```r
BIC(fit1)/num - log(2*pi) # BIC
```

```
## [1] 19.67619
```

```r
AIC(fit2)/num - log(2*pi) # AIC
```

```
## [1] 19.48114
```

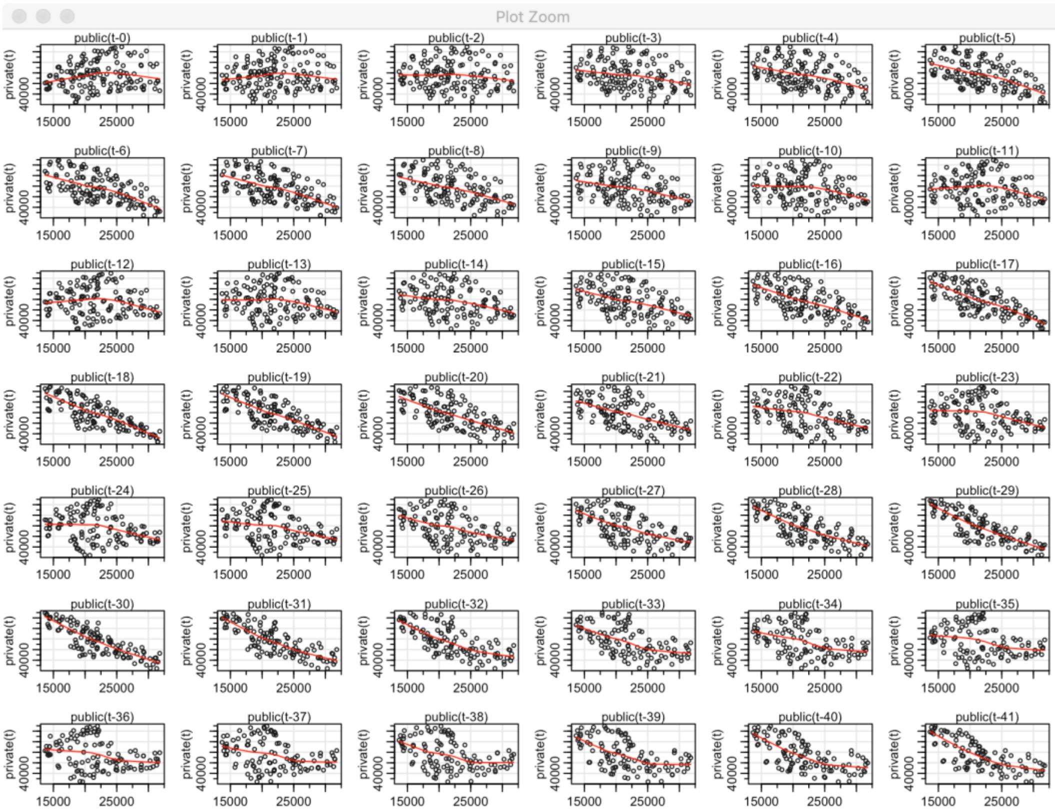
```r
BIC(fit2)/num - log(2*pi) # BIC
```

```
## [1] 19.5815
```

```r
summary(fit2)
```

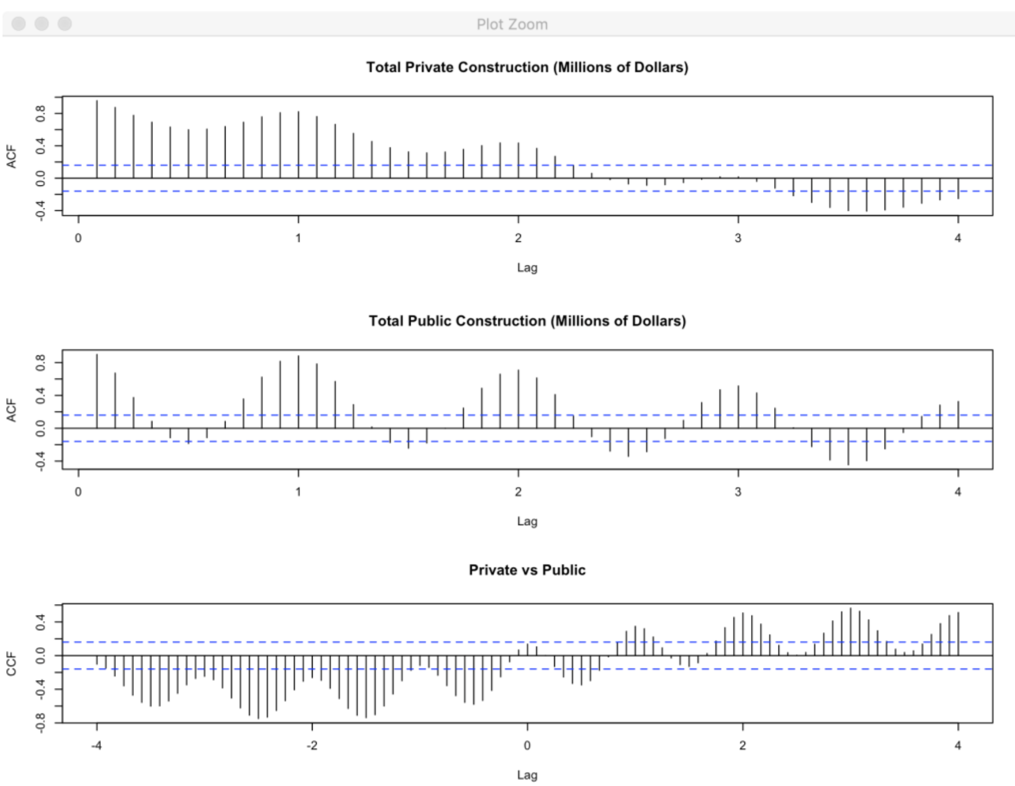
```
##
## Call:
## lm(formula = private ~ trend + pub + pub2, na.action = NULL)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -17257  -9211   1446   8568  18317
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.538e+06  5.201e+05   8.725 5.54e-15 ***
## trend       -2.229e+03  2.589e+02  -8.610 1.08e-14 ***
## pub          1.283e+00  2.036e-01   6.302 3.28e-09 ***
## pub2        -1.614e-04  3.614e-05  -4.467 1.58e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10100 on 146 degrees of freedom
## Multiple R-squared:  0.3733, Adjusted R-squared:  0.3604
## F-statistic: 28.99 on 3 and 146 DF,  p-value: 9.198e-15
```

```
### (2) lagged regression

# method 1:
lag2.plot(public, private, 41, corr=FALSE)
```



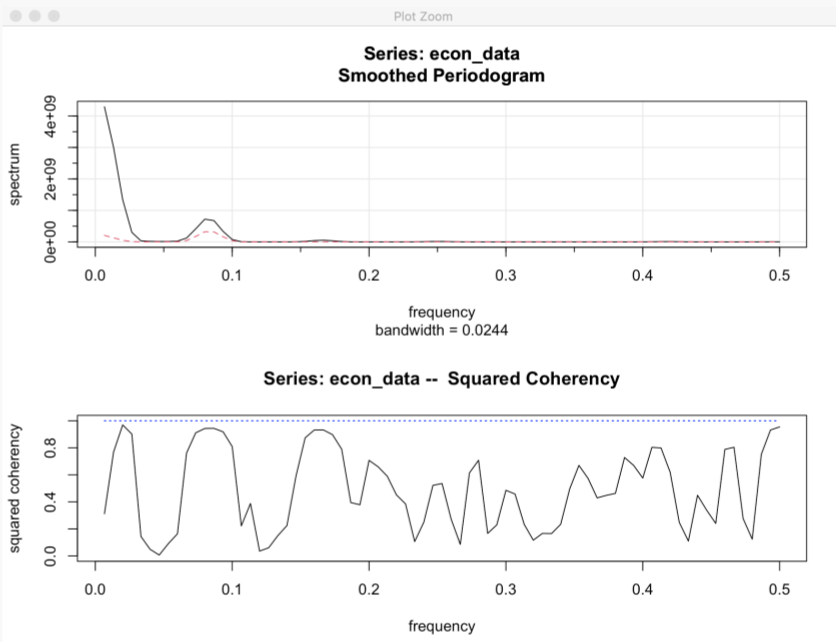

```
fit3 <- dynlm(private ~ L(public,18))
summary(fit3)

fit4 <- dynlm(private ~ L(public,19))
summary(fit4)

fit5 <- dynlm(private ~ L(public,30))
summary(fit5)

fit6 <- dynlm(private ~ L(public,31))
summary(fit6)

fit7 <- dynlm(private ~ L(public,18) + L(public,19) + L(public,30) + L(publi
c,31))
summary(fit7)

##
## Time series regression with "ts" data:
## Start = 2004(8), End = 2014(6)
##
```

```
## Call:
## dynlm(formula = private ~ L(public, 18) + L(public, 19) + L(public,
##     30) + L(public, 31))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -19160  -4669   1141   5573  16955
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.125e+05  3.540e+03  31.787   <2e-16 ***
## L(public, 18)  2.219e-01  1.141e+00   0.194   0.8462
## L(public, 19) -1.118e+00  1.147e+00  -0.975   0.3317
## L(public, 30) -2.079e+00  1.180e+00  -1.761   0.0809 .
## L(public, 31)  5.238e-01  1.183e+00   0.443   0.6589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7654 on 114 degrees of freedom
## Multiple R-squared:  0.6995, Adjusted R-squared:  0.689
## F-statistic: 66.36 on 4 and 114 DF,  p-value: < 2.2e-16
# method 2
par(mfrow=c(3,1))
acf(private, 48, main="Total Private Construction (Millions of Dollars)")
acf(public, 48, main="Total Public Construction (Millions of Dollars)")
ccf(public, private, 48, main="Private vs Public", ylab="CCF")
```



```r
fit8 <- dynlm(private ~ L(public,6) + L(public,18) + L(public,30) + L(public,
42))
summary(fit8)

fit9 <- dynlm(private ~ L(public,6) + L(public,18) + L(public,30) + L(public,
42) + L(public,19) + L(public,30) + L(public,31))
summary(fit9)

##
## Time series regression with "ts" data:
## Start = 2005(7), End = 2014(6)
##
## Call:
## dynlm(formula = private ~ L(public, 6) + L(public, 18) + L(public,
##     30) + L(public, 42) + +L(public, 19) + L(public, 30) + L(public,
##     31))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16735.3  -3611.5    417.8   4462.9  12367.9
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.011e+05  3.469e+03  29.142  < 2e-16 ***
## L(public, 6)   3.460e+00  5.228e-01   6.619 1.76e-09 ***
## L(public, 18) -3.426e+00  1.175e+00  -2.916  0.00437 **
## L(public, 30) -3.803e+00  1.186e+00  -3.206  0.00180 **
## L(public, 42)  2.471e+00  5.436e-01   4.546 1.52e-05 ***
## L(public, 19) -1.242e+00  9.848e-01  -1.261  0.21009
## L(public, 31)  6.106e-01  1.013e+00   0.603  0.54791
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6181 on 101 degrees of freedom
## Multiple R-squared:  0.8133, Adjusted R-squared:  0.8022
## F-statistic: 73.35 on 6 and 101 DF,  p-value: < 2.2e-16


# method 3
summary(LagReg(public, private, L=15, M=32, threshold=1))

##      lag s  beta(s)
## [1,]     0 1.188661

## The prediction equation is
## private(t) = alpha + sum_s[ beta(s)*public(t-s) ], where alpha = 31274.6
## MSE = 180066072

summary(LagReg(private, public, L=15, M=32, inverse=TRUE, threshold=0.1))
```

```
##       lag s   beta(s)
## [1,]     5 -0.122755

##
## The prediction equation is
## public(t) = alpha + sum_s[ beta(s)*private(t+s) ], where alpha = 29138.43
## MSE = 16561206

mse(fit9, as.data.frame(cbind(private, public))) # 271361832

## [1] 271361832


### (3) coherence analysis (cross periodogram)

econ_data<-as.data.frame(cbind(private, public))
econ=mvspec(econ_data,spans=c(3,3),taper=.5)
plot(econ,plot.type="coh",ci=-1)
```

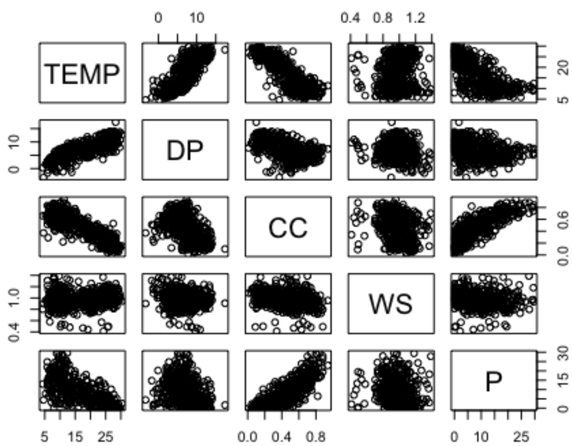



```
econ$df

f = qf(.999, 2, econ$df-2)

C = f/(18+f)
```

Data 4: (multivariate) "wheatherPr.xlsx"

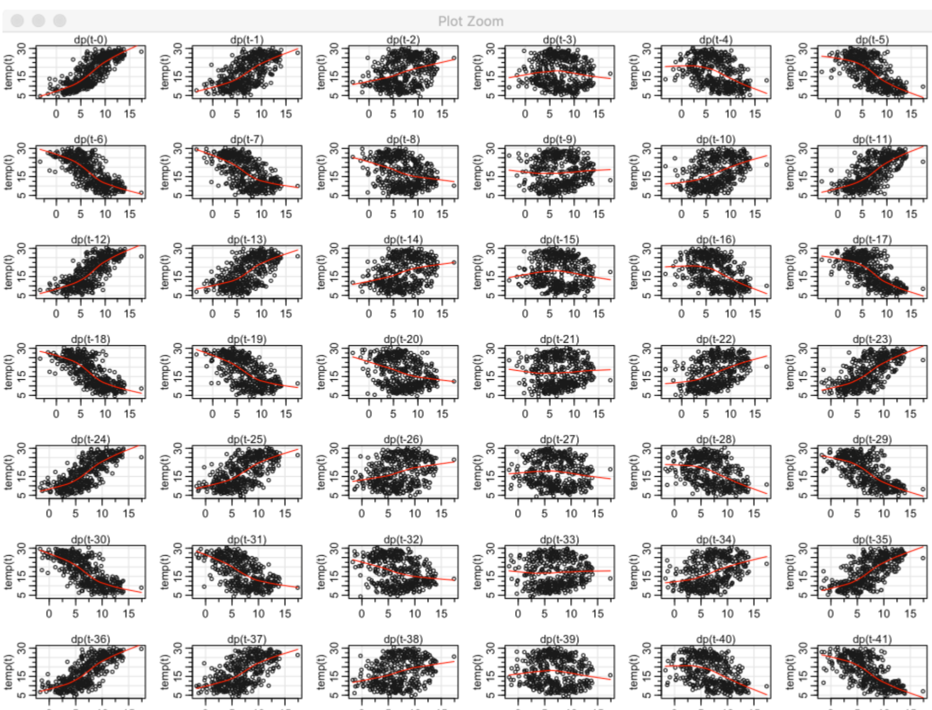
```
data_wheather <- read_excel("wheatherPr.xlsx")

temp <- data_wheather$Temp
dp <- data_wheather$DewPt
```

```r
cc <- data_wheather$CldCvr
ws <- data_wheather$WndSpd
pr <- data_wheather$Precip

### (1) linear regression with dependent errors

pairs(cbind(TEMP=temp, DP=dp, CC=cc, WS=ws, P=sqrt(pr)))
```

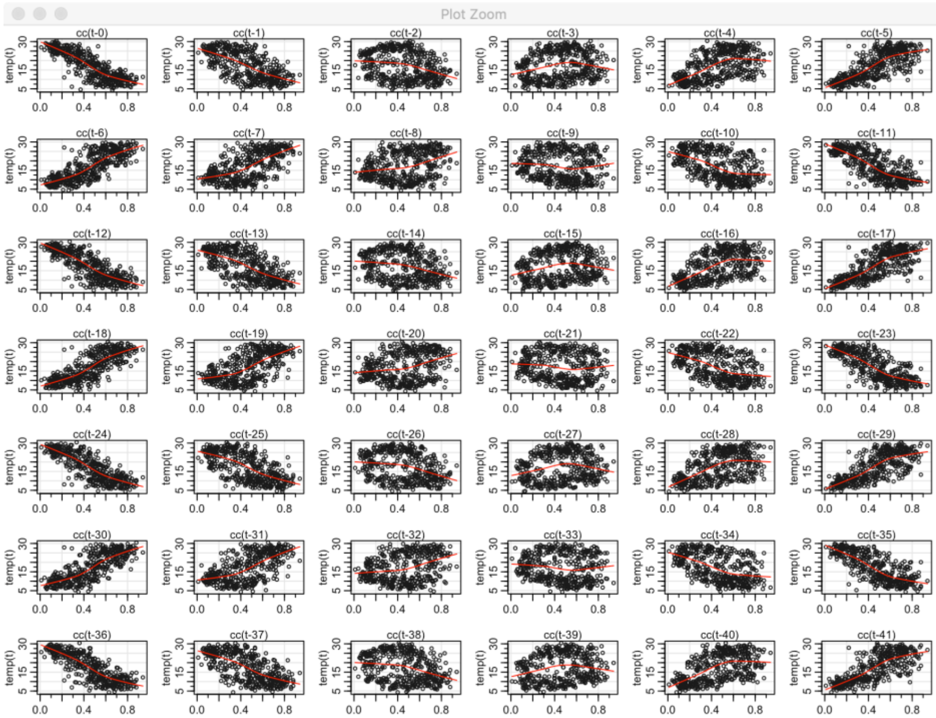



```r
trend <- time(temp)
P <- sqrt(pr)
fit1 <- lm(temp ~ trend + dp + cc + ws + pr, na.action=NULL)
fit2 <- lm(temp ~ trend + dp + cc + ws + pr + P, na.action=NULL)
summary(aov(lm(temp ~ cbind(trend, dp, cc, ws, pr, P))))

##                                  Df Sum Sq Mean Sq F value Pr(>F)
## cbind(trend, dp, cc, ws, pr, P)   6  22047    3675    1041 <2e-16 ***
## Residuals                       447   1578       4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

num <- length(temp) # sample size
AIC(fit1)/num - log(2*pi) # AIC

## [1] 2.450571

BIC(fit1)/num - log(2*pi) # BIC

## [1] 2.514066

AIC(fit2)/num - log(2*pi) # AIC

## [1] 2.280973

BIC(fit2)/num - log(2*pi) # BIC

## [1] 2.353538
```

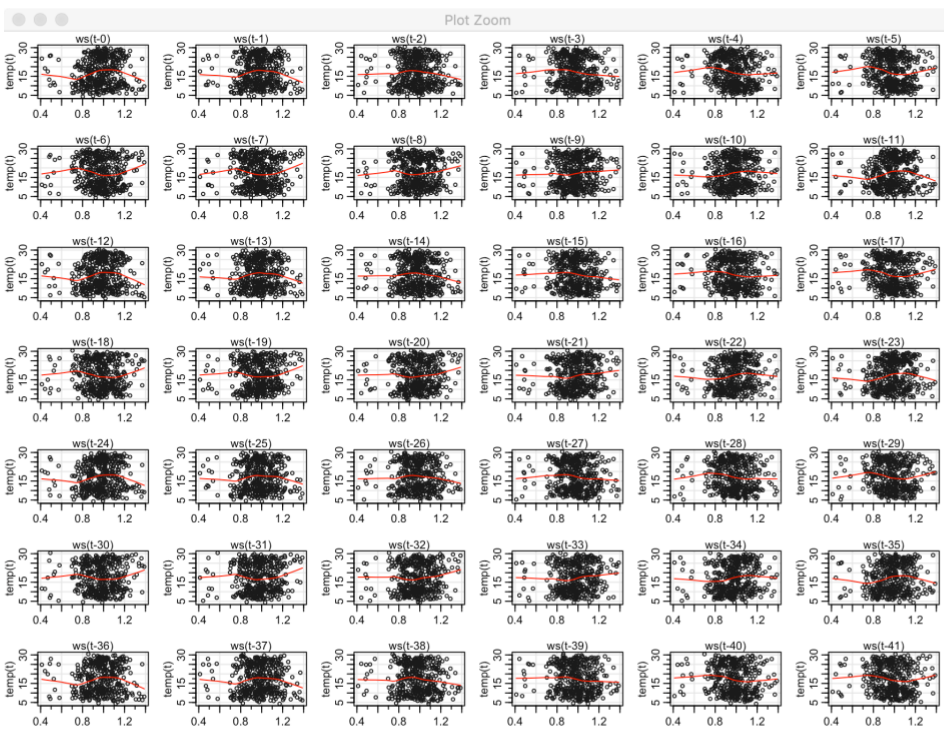
```r
summary(fit2)

##
## Call:
## lm(formula = temp ~ trend + dp + cc + ws + pr + P, na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5848 -1.1172  0.0092  1.0201  6.1681
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.929e+01  9.552e-01  20.197  < 2e-16 ***
## trend       -4.073e-03  7.149e-04  -5.697 2.22e-08 ***
## dp           1.111e+00  3.223e-02  34.472  < 2e-16 ***
## cc          -1.285e+01  8.740e-01 -14.697  < 2e-16 ***
## ws          -5.311e-01  6.703e-01  -0.792    0.429
## pr           1.015e-02  1.658e-03   6.123 2.02e-09 ***
## P           -4.362e-01  4.733e-02  -9.217  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.879 on 447 degrees of freedom
## Multiple R-squared:  0.9332, Adjusted R-squared:  0.9323
## F-statistic:  1041 on 6 and 447 DF,  p-value: < 2.2e-16
### (2) lagged regression
lag2.plot(dp, temp, 41, corr=FALSE)
```



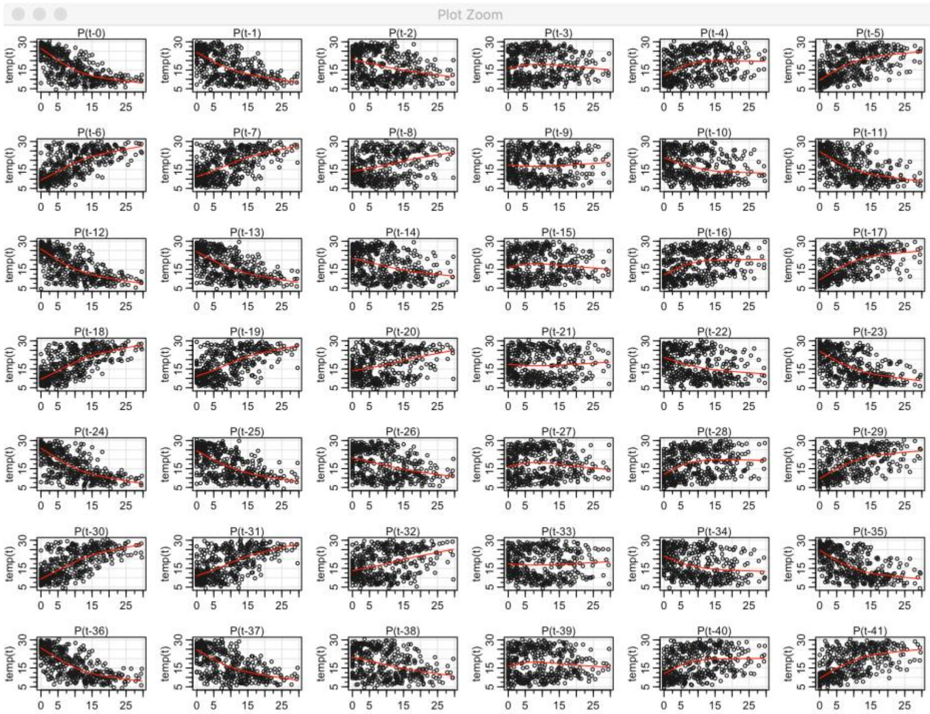
```r
lag2.plot(cc, temp, 41, corr=FALSE)
```



```
lag2.plot(ws, temp, 41, corr=FALSE)
```



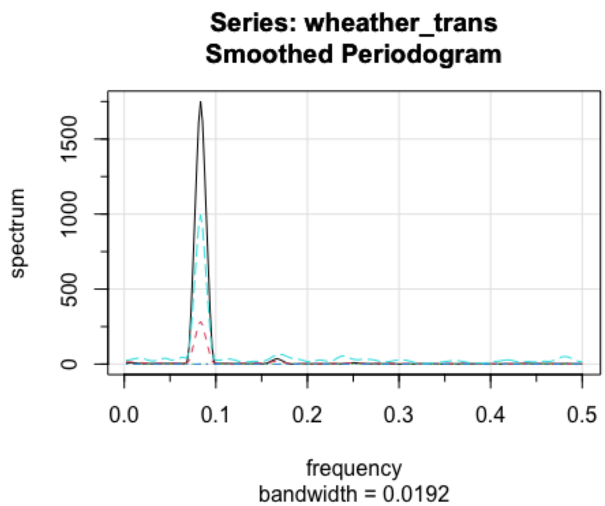
```
lag2.plot(P, temp, 41, corr=FALSE)
```



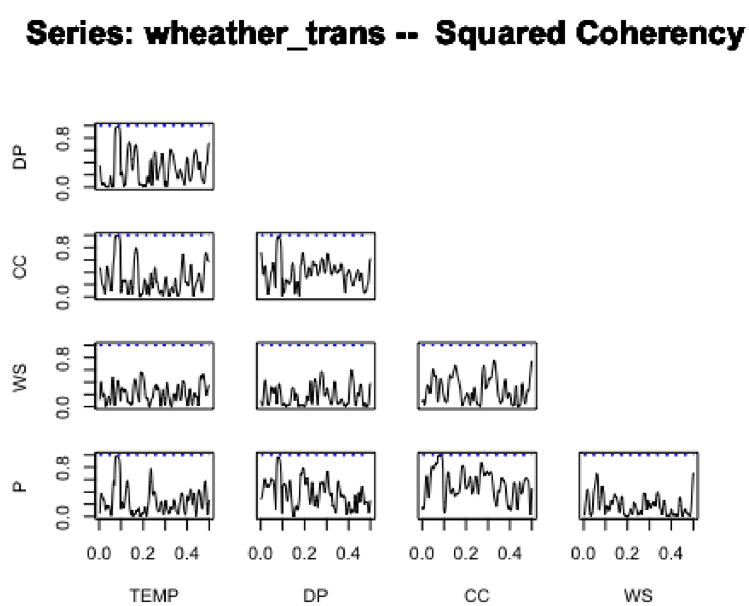

```r
fit3 <- dynlm(temp ~ L(dp,38) + L(cc,1) + L(ws,9) + L(P,8))
summary(fit3)
```

```r
fit4 <- dynlm(temp ~ L(dp,38) + L(cc,1) + L(P,8))
summary(fit4)

##
## Time series regression with "numeric" data:
## Start = 1, End = 454
##
## Call:
## dynlm(formula = temp ~ L(dp, 38) + L(cc, 1) + L(P, 8))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6654 -1.2703  0.0727  1.2726  6.3407
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.76701    0.42037   42.27  < 2e-16 ***
## L(dp, 38)     1.08722    0.03145   34.57  < 2e-16 ***
## L(cc, 1)    -14.70077    0.85106  -17.27  < 2e-16 ***
## L(P, 8)      -0.16402    0.02377   -6.90 1.78e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.018 on 450 degrees of freedom
## Multiple R-squared:  0.9224, Adjusted R-squared:  0.9219
## F-statistic:  1783 on 3 and 450 DF,  p-value: < 2.2e-16
```

```r
### (3) coherence analysis (cross periodogram)

wheather_trans = rename(as.data.frame(cbind(temp, dp, cc, ws, pr)),
                        TEMP=temp, DP=dp, CC=cc, WS=ws, P=pr)
wheather_trans[,"P"]=sqrt(wheather_trans[,"P"])

weather=mvspec(wheather_trans,spans=c(6,6),taper=.5)
```

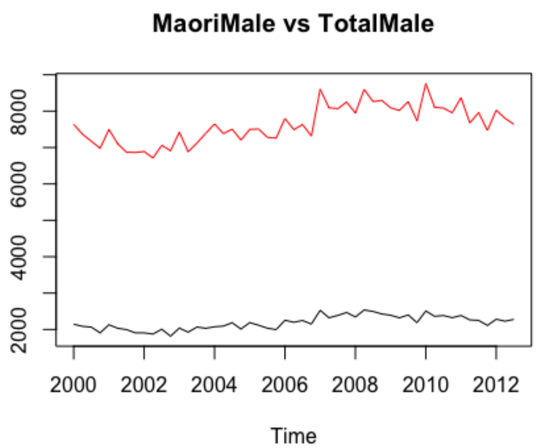



```r
weather$df

f = qf(.999, 2, weather$df-2)

C = f/(18+f)

plot(weather,plot.type="coh",ci=-1)
```

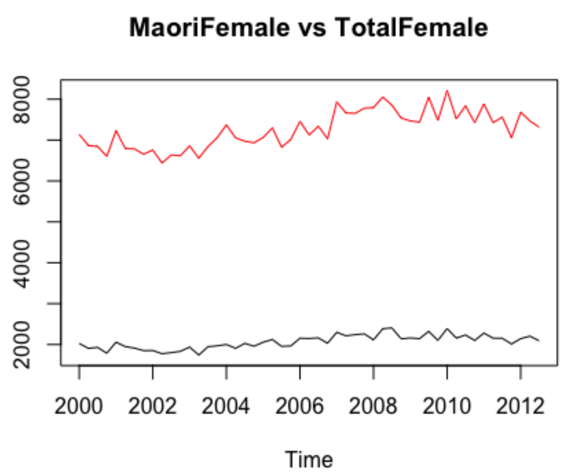


Data 5: **(multivariate) "NZBirths.csv"**

```
data_nzb <- read.csv("NZBirths.csv")
# I would study two variables
mm <- ts(data_nzb$MaoriMale, start=c(2000,1), frequency=4)
tm <- ts(data_nzb$TotalMale, start=c(2000,1), frequency=4)
mf <- ts(data_nzb$MaoriFemale, start=c(2000,1), frequency=4)
tf <- ts(data_nzb$TotalFemale, start=c(2000,1), frequency=4)
ts.plot(mm, tm, gpars=list(col=c("black","red")), main="MaoriMale vs TotalMal
e")
```

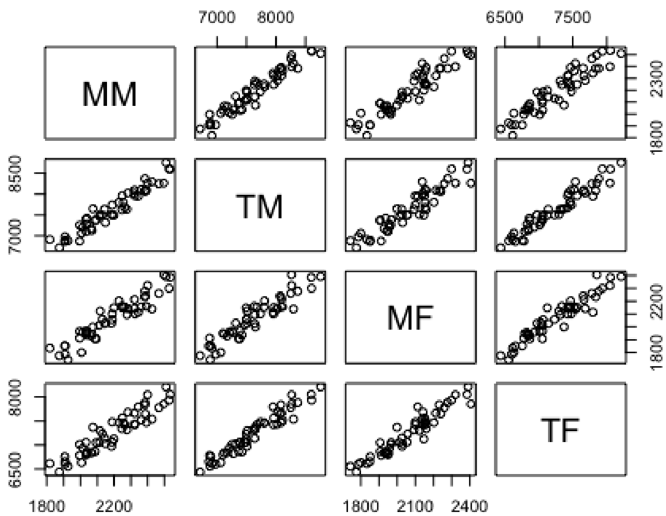



```
ts.plot(mf, tf, gpars=list(col=c("black","red")), main="MaoriFemale vs TotalF
emale")
```




```
### (1) linear regression with dependent errors
```

```r
pairs(cbind(MM=data_nzb$MaoriMale, TM=data_nzb$TotalMale, MF=data_nzb$MaoriFe
male, TF=data_nzb$TotalFemale))
```




```r
fit1 <- lm(tm ~ mm + mf + tf, na.action=NULL)
summary(fit1)
```

```
##
## Call:
## lm(formula = tm ~ mm + mf + tf, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -255.126  -53.857   -2.107   75.631  224.388
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 438.6923   291.9047   1.503    0.140
## mm            1.4966     0.2493   6.003 2.66e-07 ***
## mf           -0.4230     0.3296  -1.283    0.206
## tf            0.6606     0.1172   5.638 9.46e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 104 on 47 degrees of freedom
## Multiple R-squared:  0.9611, Adjusted R-squared:  0.9586
## F-statistic: 387.3 on 3 and 47 DF,  p-value: < 2.2e-16
```

```r
summary(aov(fit1))
```

```
##             Df   Sum Sq  Mean Sq  F value   Pr(>F)
## mm           1 12136840 12136840 1122.505  < 2e-16 ***
## mf           1    80805    80805    7.473   0.0088 **
```

```
## tf           1   343684   343684   31.786 9.46e-07 ***
## Residuals   47   508177    10812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### (2) lagged regression
summary(LagReg(mm, tm, L=15, M=32, threshold=0.2))
```

```
##      lag s    beta(s)
## [1,]     0  2.5524434
## [2,]     1 -0.4082032
## The prediction equation is
## tm(t) = alpha + sum_s[ beta(s)*mm(t-s) ], where alpha = 2941.714
## MSE = 164384.3
```

```
summary(LagReg(mf, tm, L=15, M=32, threshold=0.2))
```

```
##      lag s    beta(s)
## [1,]     0  2.4869839
## [2,]     1 -0.2792588
## [3,]     2  0.2615193
## The prediction equation is
## tm(t) = alpha + sum_s[ beta(s)*mf(t-s) ], where alpha = 2531.176
## MSE = 171474.5
```

```
summary(LagReg(tf, tm, L=15, M=32, threshold=0.2))
```

```
## [1,]     0 1.041174
## The prediction equation is
## tm(t) = alpha + sum_s[ beta(s)*tf(t-s) ], where alpha = 78.54232
## MSE = 150657.8
```

```
fit2 <- dynlm(tm ~ mm + L(mm,1) + mf + L(mf,1) + L(mf,2) + tf)
summary(fit2)
```

```
##
## Time series regression with "ts" data:
## Start = 2000(3), End = 2012(3)
##
## Call:
## dynlm(formula = tm ~ mm + L(mm, 1) + mf + L(mf, 1) + L(mf, 2) +
##     tf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -238.69  -66.61  -19.37   75.96  197.19
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 509.37542  319.39725   1.595    0.118
## mm            1.62727    0.28943   5.622 1.38e-06 ***
## L(mm, 1)     -0.01182    0.25385  -0.047    0.963
```

```
## mf            -0.44470    0.34843  -1.276    0.209
## L(mf, 1)      -0.08000    0.26793  -0.299    0.767
## L(mf, 2)      -0.12610    0.15768  -0.800    0.428
## tf             0.67922    0.12152   5.589 1.55e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 105.4 on 42 degrees of freedom
## Multiple R-squared:  0.9641, Adjusted R-squared:  0.959
## F-statistic: 187.9 on 6 and 42 DF,  p-value: < 2.2e-16
```

```
summary(LagReg(mm, tm, L=15, M=32, threshold=0.1))
```

```
##      lag s    beta(s)
## [1,]     0  2.5524434
## [2,]     1 -0.4082032
## [3,]     5 -0.1790251
## [4,]     7 -0.1712018
## [5,]     9 -0.1464236
## [6,]    11 -0.1460824
## [7,]    13 -0.1825546
## [8,]    15 -0.1458576
## The prediction equation is
## tm(t) = alpha + sum_s[ beta(s)*mm(t-s) ], where alpha = 5068.312
## MSE = 222556.3
```

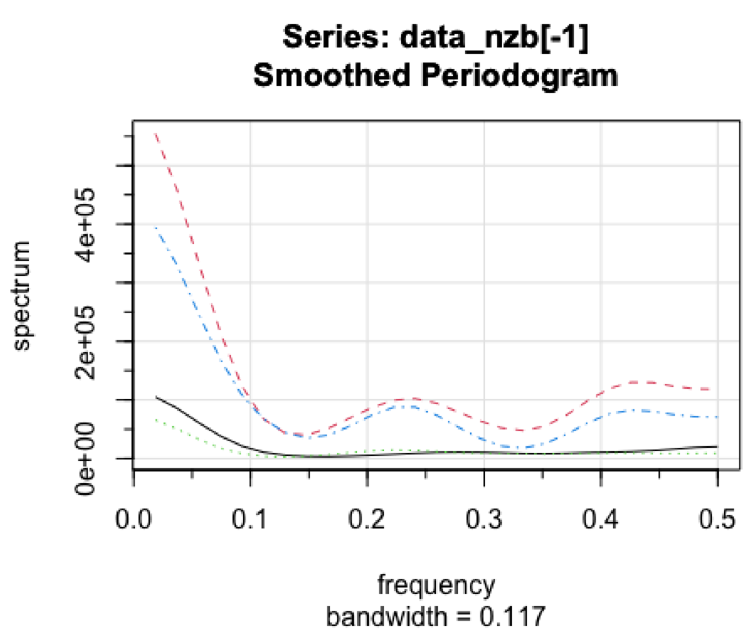
```
summary(LagReg(mf, tm, L=15, M=32, threshold=0.1))
```

```
##       lag s    beta(s)
##  [1,]     0  2.4869839
##  [2,]     1 -0.2792588
##  [3,]     2  0.2615193
##  [4,]     3 -0.1661519
##  [5,]     5 -0.1625314
##  [6,]     7 -0.1925313
##  [7,]     9 -0.1650006
##  [8,]    11 -0.1971650
##  [9,]    13 -0.1964090
## [10,]    15 -0.1656484

## The prediction equation is
## tm(t) = alpha + sum_s[ beta(s)*mf(t-s) ], where alpha = 5106.521
## MSE = 244360.5
```

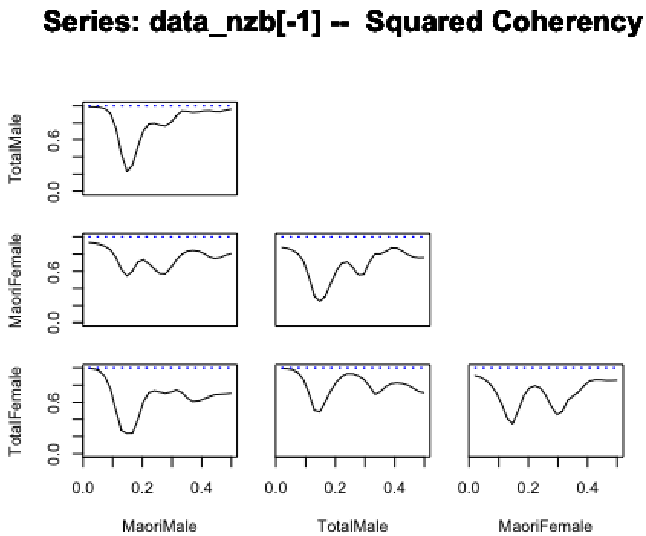
```
summary(LagReg(tf, tm, L=15, M=32, threshold=0.1))
```

```
##      lag s  beta(s)
## [1,]     0 1.041174
## The prediction equation is
## tm(t) = alpha + sum_s[ beta(s)*tf(t-s) ], where alpha = 78.54232
## MSE = 150657.8
```

```
fit3 <- dynlm(tm ~ mm + L(mm,1) + L(mm,5) + L(mm,7) + L(mm,9) + L(mm,11) + L
(mm,13) + L(mm,15) + mf + L(mf,1) + L(mf,2) + L(mf,3) + L(mf,5) + L(mf,7) + L
(mf,9) + L(mf,11) + L(mf,13) + L(mf,15) + tf)
summary(fit3)

fit4 <- dynlm(tm ~ mm + tf)
summary(fit4)
## Time series regression with "ts" data:
## Start = 2000(1), End = 2012(3)
##
## Call:
## dynlm(formula = tm ~ mm + tf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -231.370  -59.928   -4.556   84.342  220.694
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 551.65153  280.19179   1.969   0.0548 .
## mm            1.35044    0.22327   6.049 2.11e-07 ***
## tf            0.56866    0.09335   6.092 1.81e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 104.7 on 48 degrees of freedom
## Multiple R-squared:  0.9598, Adjusted R-squared:  0.9581
## F-statistic: 572.3 on 2 and 48 DF,  p-value: < 2.2e-16

### (3) coherence analysis (cross periodogram)
nzb=mvspec(data_nzb[-1],spans=c(5,5),taper=.5)
```

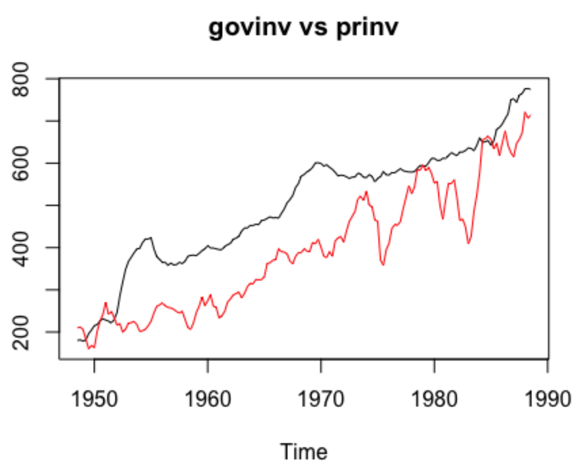


```
nzb$df
f = qf(.999, 2, nzb$df-2)
C = f/(18+f)
plot(nzb,plot.type="coh",ci=-1)
```

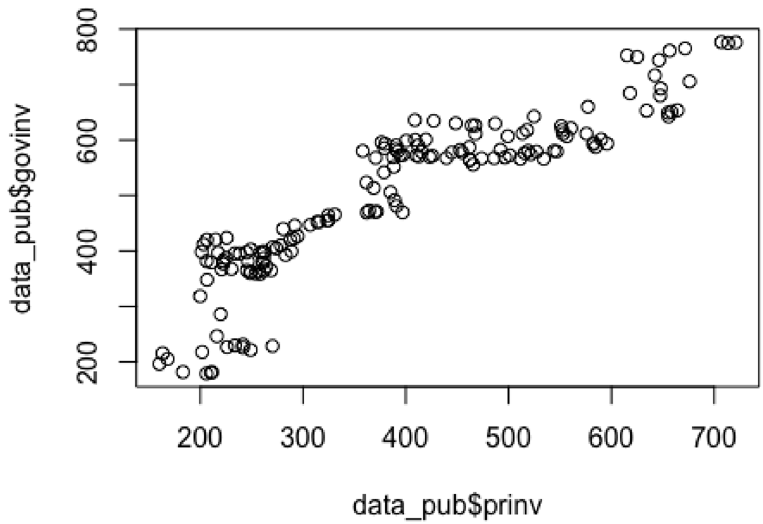


Data 6: **(multivariate) "pub-prinv.xlsx"**

```
data_pub <- read_excel("pub-prinv.xlsx")
govinv <- ts(data_pub$govinv, start=c(1948,3), frequency=4)
prinv <- ts(data_pub$prinv, start=c(1948,3), frequency=4)
ts.plot(govinv, prinv, gpars=list(col=c("black","red")), main="govinv vs prin
v")
```



```r
### (1) linear regression with dependent errors

plot(x=data_pub$prinv, y=data_pub$govinv)
```




```r
fit1 <- lm(govinv ~ prinv)
summary(fit1)

##
## Call:
## lm(formula = govinv ~ prinv)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -163.950  -28.794    9.021   38.140  122.231
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 155.57985   13.80731   11.27   <2e-16 ***
## prinv         0.87595    0.03288   26.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.86 on 159 degrees of freedom
## Multiple R-squared:  0.817,  Adjusted R-squared:  0.8158
## F-statistic: 709.7 on 1 and 159 DF,  p-value: < 2.2e-16

### (2) lagged regression and TFM

# method 1
summary(LagReg(prinv, govinv, L=15, M=32, threshold=0.02))
```

```
##      lag s     beta(s)
## [1,]     0 -0.03776259
## [2,]     1 -0.03684783
## [3,]     2  0.03691519
## [4,]     3 -0.06304850
## [5,]     5  0.05832572
## [6,]     8  0.03051177
## The prediction equation is
## govinv(t) = alpha + sum_s[ beta(s)*prinv(t-s) ], where alpha = 504.4026
## MSE = 16958.26
```

```r
summary(LagReg(govinv, prinv, L=15, M=32, inverse=TRUE, threshold=0.02))
```

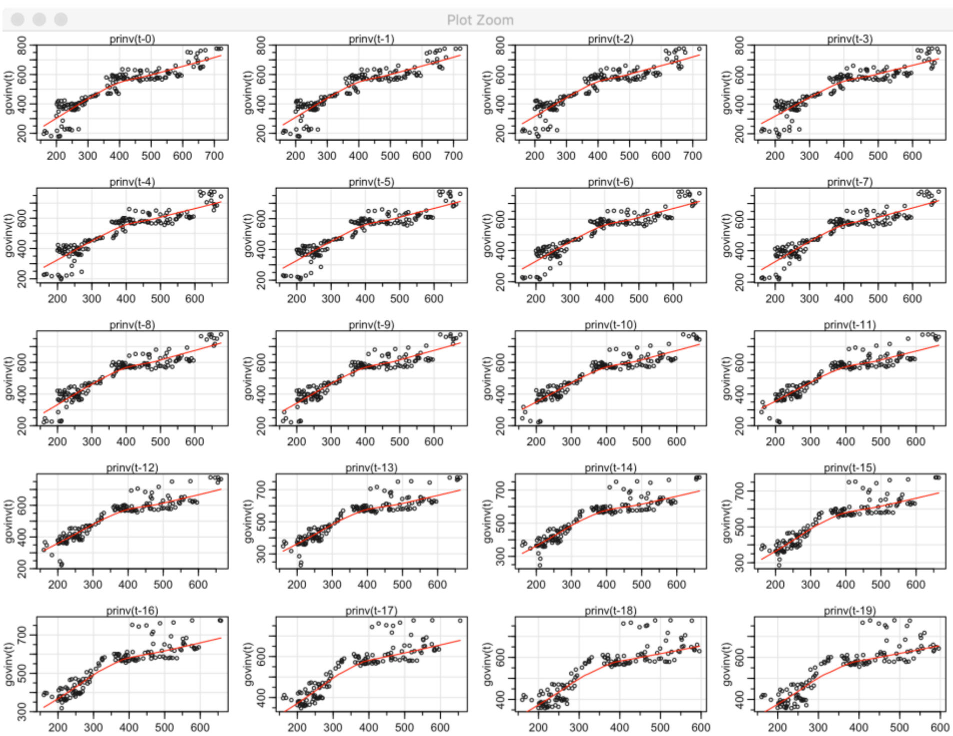
```
##       lag s     beta(s)
##  [1,]     1 -0.32082874
##  [2,]     2  0.45312647
##  [3,]     3 -0.60468558
##  [4,]     4 -0.05319550
##  [5,]     5  0.39065064
##  [6,]     6  0.10973269
##  [7,]     7 -0.22457037
##  [8,]     8  0.32045603
##  [9,]     9 -0.11558723
## [10,]    10  0.06702691
## [11,]    11 -0.16594176
## [12,]    12 -0.02368407
## [13,]    13  0.03883186
## [14,]    15 -0.02322145
## The prediction equation is
## prinv(t) = alpha + sum_s[ beta(s)*govinv(t+s) ], where alpha = 468.784
## MSE = 21785.41
```

```r
fit2 <- dynlm(govinv ~ prinv + L(prinv,1) +  L(prinv,2) +  L(prinv,3) +  L(pr
inv,5) +  L(prinv,8))
summary(fit2)
```

```r
fit3 <- dynlm(govinv ~ prinv + L(prinv,8))
summary(fit3)
```

```
##
## Time series regression with "ts" data:
## Start = 1950(3), End = 1988(3)
##
## Call:
## dynlm(formula = govinv ~ prinv + L(prinv, 8))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -148.179  -18.906    1.564   25.177   91.454
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 169.12614   10.69228  15.818  < 2e-16 ***
## prinv         0.37146    0.05251   7.074 5.36e-11 ***
## L(prinv, 8)   0.51964    0.05517   9.418  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.63 on 150 degrees of freedom
## Multiple R-squared:  0.8875, Adjusted R-squared:  0.886
## F-statistic: 591.9 on 2 and 150 DF,  p-value: < 2.2e-16

# method 2, to be continued
lag2.plot(prinv, govinv, 19, corr=FALSE)
```



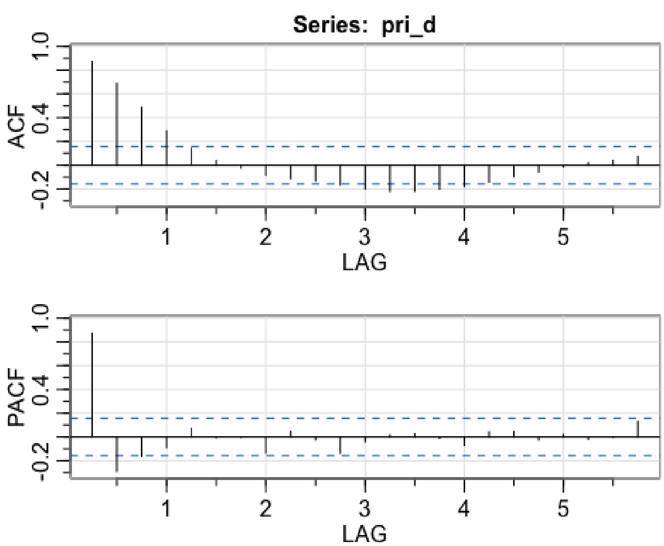

```
dummy = ifelse(prinv<400, 0, 1)
priL8 <- stats::lag(prinv,-8)
dL8 = stats::lag(dummy,-8)
inv = ts.intersect(govinv, prinv, priL8, dL8, dframe=TRUE)
fit4 <- lm(govinv ~ priL8*dL8, data=inv, na.action=NULL)
summary(fit4)

fit5 <- lm(govinv ~ prinv + priL8*dL8, data=inv, na.action=NULL)
summary(fit5)
## Call:
## lm(formula = govinv ~ prinv + priL8 * dL8, data = inv, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -119.388  -21.010    0.575   23.045   98.874
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  51.12047   15.93373   3.208  0.00164 **
## prinv         0.31625    0.04556   6.941 1.14e-10 ***
## priL8         1.02517    0.07348  13.951  < 2e-16 ***
## dL8         174.04024   36.14605   4.815 3.60e-06 ***
## priL8:dL8    -0.57520    0.08354  -6.886 1.53e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.52 on 148 degrees of freedom
## Multiple R-squared:  0.9264, Adjusted R-squared:  0.9245
## F-statistic:   466 on 4 and 148 DF,  p-value: < 2.2e-16
```

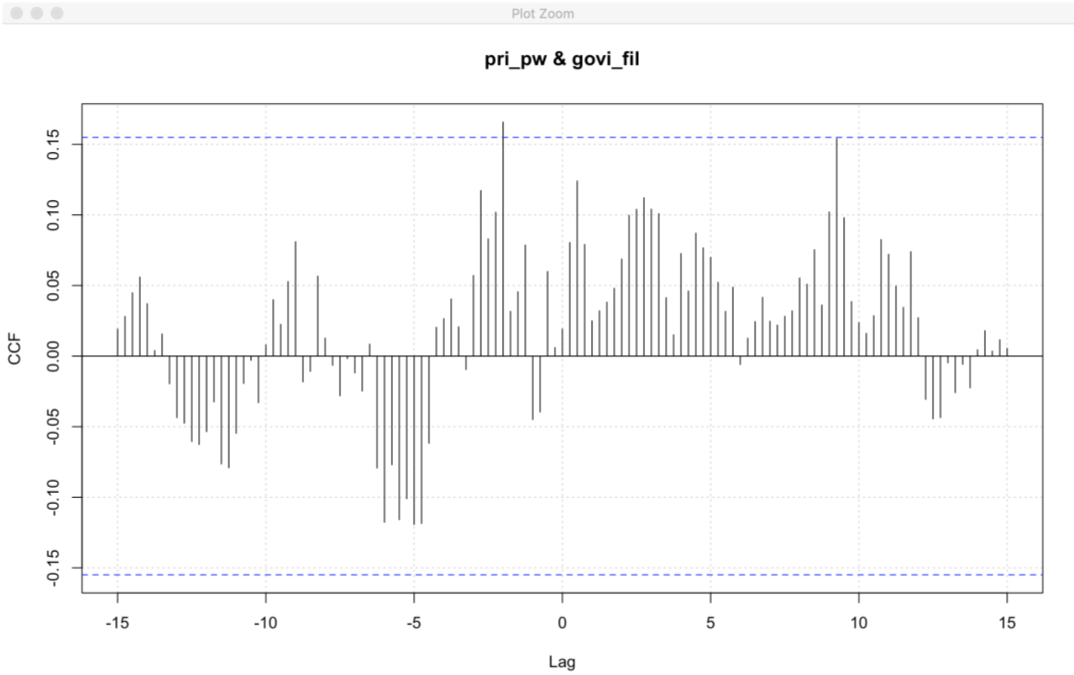
```
AIC(fit5)
```

```
## [1] 1533.578
```

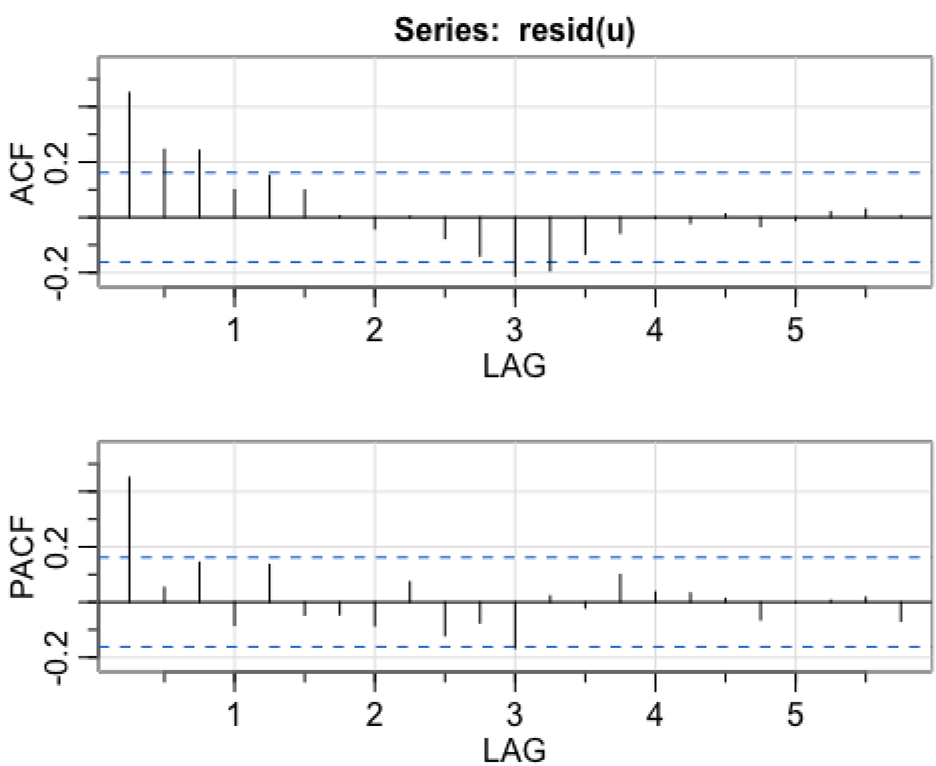
```
# TFM, input:prinv, output:govinv
pri_d = resid(lm(prinv~time(prinv), na.action=NULL)) # detrended prinv
acf2(pri_d)
```



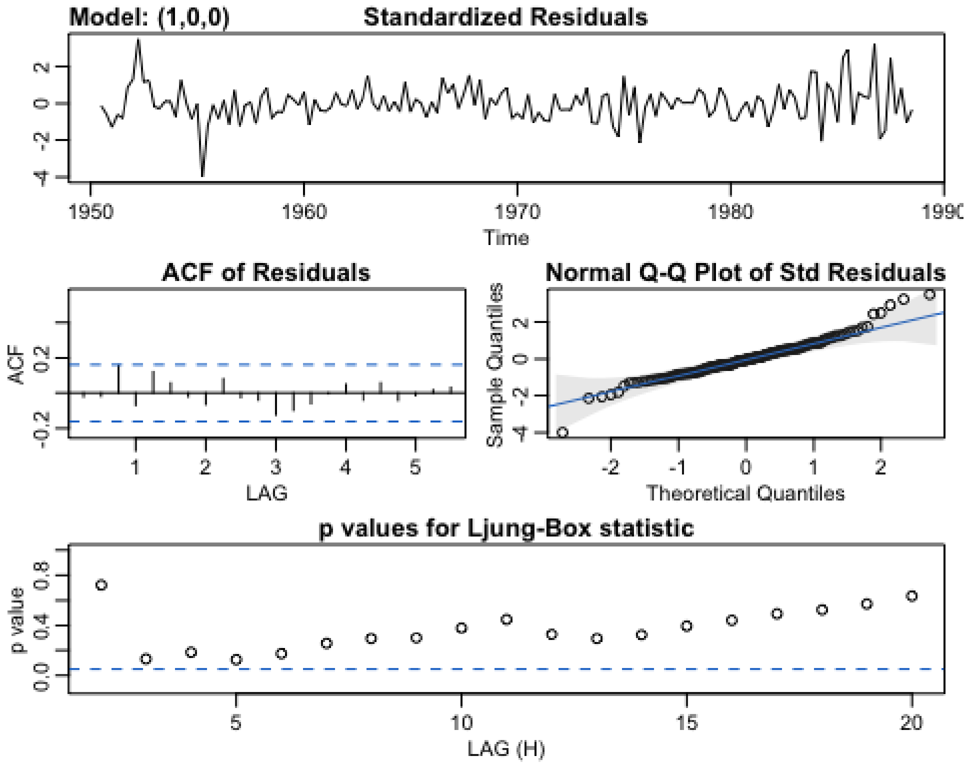

```
fit = arima(pri_d, order=c(1,0,0))
ar1 = as.numeric(coef(fit)[1]) # = 0.8912
pri_pw = resid(fit)
govi_fil = stats::filter(govinv, filter=c(1, -ar1), sides=1)
ccf(pri_pw, govi_fil, ylab="CCF", na.action=na.omit, panel.first=grid(), 60)
```




```
gov_pr = ts.intersect(govinv, govL1=stats::lag(govinv,-1), priL8=stats::lag(p
ri_d,-8))
u = lm(gov_pr[,1]~gov_pr[,2:3], na.action=NULL)
acf2(resid(u)) # suggests ar1
```

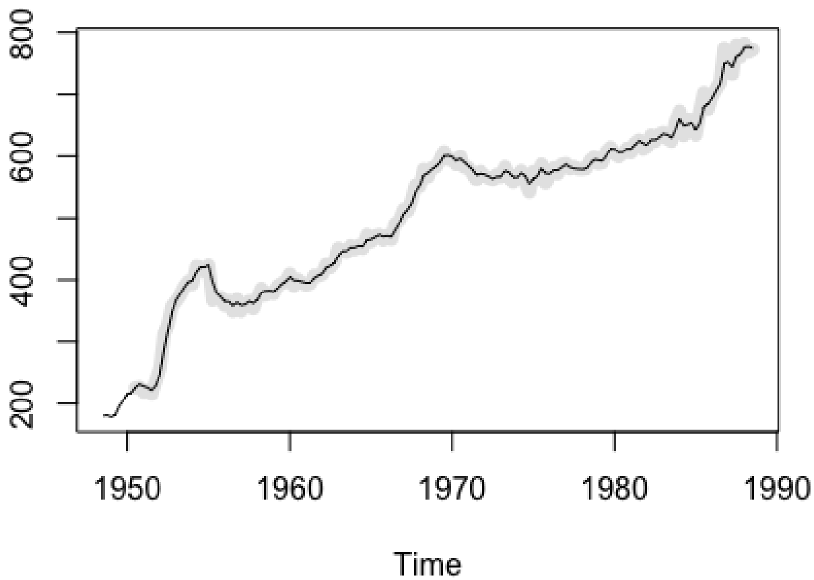



```
arx = sarima(gov_pr[,1], 1, 0, 0, xreg=gov_pr[,2:3]) # final model
```

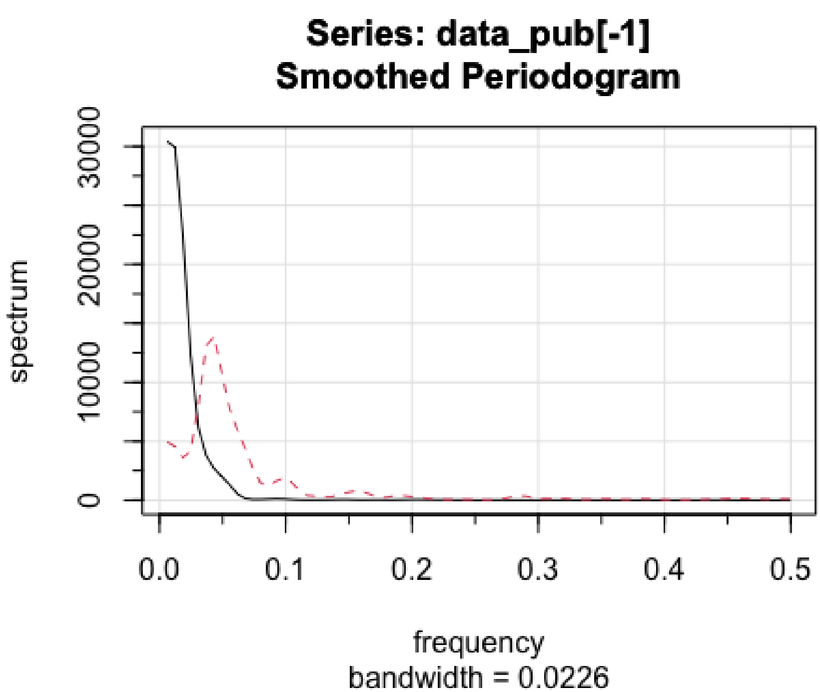



```r
arx$AIC

## [1] 6.962322

pred = govinv + resid(arx$fit) # 1-step-ahead predictions
ts.plot(pred, govinv, col=c('gray90',1), lwd=c(7,1))
```

```
Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace = trc,
    REPORT = 1, reltol = tol))

Coefficients:
         ar1  intercept   govL1   priL8
      0.4665     8.6165  0.9905  0.0591
s.e.  0.0753     4.6541  0.0088  0.0223

sigma^2 estimated as 57.83:  log likelihood = -527.62,  aic = 1065.24

$degrees_of_freedom
[1] 149

$ttable
          Estimate     SE  t.value p.value
ar1         0.4665 0.0753   6.1951  0.0000
intercept   8.6165 4.6541   1.8514  0.0661
govL1       0.9905 0.0088 112.3556  0.0000
priL8       0.0591 0.0223   2.6495  0.0089

$AIC
[1] 6.962322
```




```
### (3) coherence analysis (cross periodogram)

pub=mvspec(data_pub[-1],spans=c(3,3),taper=.5)
```




```
pub$df
```

```r
f = qf(.999, 2, pub$df-2)

C = f/(18+f)

plot(pub,plot.type="coh",ci=-1)
```