Ch 7 : Regression in Time Series

Contents

April 7, 2017

1 Regressing Stationary TS

[ToC]

1.1 OLS and GLS

[ToC]

OLS

- From model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- Minimizes sum of squares, $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$,

-

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

GLS

- Minimizes weighted sum of squares,

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Gamma}_n^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

-

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Gamma}_n^{-1}\boldsymbol{Y}.$$

- If $\boldsymbol{X}$ is non-random, we have $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and

$$V(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

- For OLS, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$, and

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

- For GLS, $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_n$, and

$$V(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X})^{-1}.$$

  It can be shown that GLS estimator is the best linear unbiased estimator of $\boldsymbol{\beta}$. Therefore, GLS is better than OLS, but we need $\boldsymbol{\Gamma}_n$.

Iterative Regression Scheme

1. Compute OLS. Get residuals.

2. Fit ARMA model to the residuals by Gaussian Maximum Likelihood.

3. Using the fitted model, calculate $\mathbf{\Gamma}_n$. Go back to the original series, compute GLS.

4. Repeat from 2).

Example: OLS and GLS

```
D  <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/gtemp.txt")
X  <- ts(D, start=c(1880), freq=1)
plot(X, type='o')

Reg1 <- lm(X~ t);  summary(Reg1)
plot(Reg1$residuals)

library(forecast)
Fit1 <- auto.arima(Reg1$residuals)
Fit1$coef
```

```
library(nlme)
cs1 <- corARMA(Fit1$coef, p=1,q=2)*Fit1$sigma
Reg2 <- gls(X~t, correlation=cs1)
Reg2
Reg1
Fit2 <- auto.arima(Reg2$residuals)
Fit2
Fit1


cs2 <- corARMA(Fit2$coef, p=0,q=2)*Fit2$sigma
Reg3 <- gls(X~t, correlation=cs2)
Reg3
Reg1
Fit3 <- auto.arima(Reg3$residuals)
Fit3

Fit1

Randomness.test(Fit3$residuals)
```

1.2 Regressing TS to TS

[ToC]

Sometimes, you want to model your TS lniarly dependaent on other time series.

- Sales depends on Global Economy
- Utility Demand depends on Temparature
- Level of toxin depends on local industory production

- In that case, your Model is that the time series of interest X_t, has linear relstionship with your independent series B_t, plus stationary noise. In the vector notation,

$$\begin{aligned} X_t &= \boldsymbol{\beta} \boldsymbol{B}_t + Y_t \\ Y_t &\sim ARMA \end{aligned}$$

where x_t contains the "explanatory" TS.

- In scalar notation,

$$X_t = a + bB_t + Y_t \qquad Y_t \sim ARMA$$

where x_t contains the "explanatory" TS.

Example: Sales Data and Indicator

```
#- Indicator and Sales -
D  <- read.table("http://gozips.uakron.edu/~nmimoto/pages/datasets/LS2.txt", header=T)
A <- ts(D$A, start=c(1,1), freq=1)
B <- ts(D$B, start=c(1,1), freq=1)


#- plot of Sales and index -
layout(matrix(1:2, 2, 1))
plot(A, type='o')
plot(B, type='o')


layout(1)
plot(A-A[1], type='o')
lines((B-B[1])*.1, col="red")     #- play around with *.1 part


#- Regress A on B
Reg1 <- lm(A~B)
summary(Reg1)

Reg1 <- lm(A~0+B)  #- no intercept
summary(Reg1)


plot(A, type="o")
lines(B*.0538-.537, col="red")

layout(matrix(c(1,1,2,3), 2,2, byrow=TRUE))
plot(Reg1$residuals, type="o")
acf(Reg1$residuals)
pacf(Reg1$residuals)
```

```
#- Model the residual

library(forecast)
source("http://gozips.uakron.edu/~nmimoto/477/TS_R-90.txt")

Fit1 <- auto.arima(Reg1$residuals)
Fit1

Fit2 <- Arima(Reg1$residuals, order=c(2,0,1) )
Fit2

Fit3 <- Arima(Reg1$residuals, order=c(2,0,1), include.mean=FALSE )
Fit3

Fit4 <- Arima(Reg1$residuals, order=c(2,0,1), include.mean=FALSE, fix=c(0,NA,NA) )
Fit4

Randomness.tests(Fit4$residuals)



#--- You can do the same thing using Arima( xreg=B ) ---

Fit5 <- Arima(A, xreg=B, order=c(2,0,1))
Fit5

Fit6 <- Arima(A, xreg=B, order=c(2,0,1), fix=c(0,NA,NA,NA,NA) )
Fit6

Fit7 <- Arima(A, xreg=B, order=c(2,0,1), fix=c(0,NA,NA,0,NA) )
Fit7

ts.plot(A,Reg1$fitted, col=c("black","red"))
```

1.3 ARMAX model

[ToC]

ARMAX(2,1) is the model

$$Y_t = \boldsymbol{\beta x}_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t + \theta_1 e_{t-1}$$

Don't confuse this with additive model

$$\begin{aligned} K_t &= \boldsymbol{\beta x}_t + Y_t \\ Y_t &\sim ARMA(2,1) \end{aligned}$$

If we write ARMAX using backwards operator,

$$
\begin{aligned}
Y_t &= \boldsymbol{\beta x}_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t + \theta_1 e_{t-1} \\
\Phi(B) Y_t &= \boldsymbol{\beta x}_t + \Theta(B) e_t \\
Y_t &= \frac{\boldsymbol{\beta x}_t}{\Phi(B)} + \frac{\Theta(B)}{\Phi(B)} e_t
\end{aligned}
$$

which is very hard to interpret.

Currently, there is no package that directly deal with ARMAX model.

2

Cointegration and Spurious Regression

[ToC]

2.1 When Two Stationary TS are Regressed

[ToC]

- What happens when two stationary time series that has nothing to do with each other are regressed?

- Regression estimates comes directry from sample correlation. What does correlation say?

- Simulate with two independent ARMA(1,1)

```
#-- X1 and X2 are both ARMA(1,1), but independent  --
X1  <- arima.sim(n = 250, list(ar = c(0.8), ma = c(-.24) ))
X2  <- arima.sim(n = 250, list(ar = c(0.8), ma = c(-.24) ))


#-- They have nothing to do with each other, and correlation says so --
plot(X1, type="l", main=paste("cor=", round(cor(X1,X2), 4) ))
lines(X2, col="red")
cor(X1,X2)
```

- To see the overall behavior, repleat 1000 times and plot histogram of correlations

```
#-- Repeat above in lopp to see overall behavior --
COR1 <- 0
for (i in 1:1000) {

  n=250
  X1  <- arima.sim(n = n, list(ar = c(0.8), ma = c(-.24) ))
  X2  <- arima.sim(n = n, list(ar = c(0.8), ma = c(-.24) ))

  COR1[i] <- cor(X1,X2)

}

hist(COR1, main=paste("mean = ", round(mean(COR1), 4) ), xlim=c(-1,1) )

sort(COR1)[c(50,950)]   #- Simulated 90% empirical CI for Correlation of 0.
```

Two dependent stationary TS 1

```
library(MASS)
e <- mvrnorm(n = 250, mu=c(0,0), Sigma=matrix(c(1,.7,.7,1), 2,2))   #- generate e from biv normal
e
plot(e[,1], e[,2])

#-- Generate X1 and X2 with correlated errors
X1  <- arima.sim(n = 250, list(ar = c(0.8), ma = c(-.24)), innov=e[,1])
X2  <- arima.sim(n = 250, list(ar = c(0.8), ma = c(-.24)), innov=e[,2])

#-- They seem to be correlated (because they are) --
plot(X1, type="l", main=paste("cor=", round(cor(X1,X2), 4) ));  lines(X2, col="red")
cor(X1,X2)

#-- Monte Carlo simulation shows --
COR2 <- 0
for (i in 1:1000) {

  e <- mvrnorm(n = 250, mu=c(0,0), Sigma=S1)
  X1  <- arima.sim(n = 250, list(ar = c(0.8), ma = c(-.24)), innov=e[,1])
  X2  <- arima.sim(n = 250, list(ar = c(0.8), ma = c(-.24)), innov=e[,2])
  COR2[i] <- cor(X1,X2)

}
hist(COR2, main=paste("mean = ", round(mean(COR2), 4) ), xlim=c(-1,1) )
```

2.2 When Two Non-stationary TS are regressed

[ToC]

- What happens when two non-stationary time series are regressed?
- Simulate with two independent ARIMA(1,1,1) ?

```
#-- X1 and X2 are both ARMA(1,1), but independent  --
X1  <- arima.sim(n = 250, list(order = c(1,1,1), ar = c(0.8), ma = c(-.24) ))
X2  <- arima.sim(n = 250, list(order = c(1,1,1), ar = c(0.8), ma = c(-.24) ))


#-- They have nothing to do with each other, and correlation says so --
plot(X1, type="l", ylim=c(min(c(X1,X2)), max(c(X1,X2))), main=paste("cor=", round(cor(X1,X2), 4) ))
lines(X2, col="red")
cor(X1,X2)
```

- Repleat 1000 times and plot histogram of correlations

```
#-- Repeat above in lopp to see overall behavior --
COR1 <- 0
for (i in 1:1000) {

  n=250
  X1  <- arima.sim(n = n, list(order=c(1,1,1), ar = c(0.8), ma = c(-.24) ))
  X2  <- arima.sim(n = n, list(order=c(1,1,1), ar = c(0.8), ma = c(-.24) ))

  COR1[i] <- cor(X1,X2)

}

hist(COR1, main=paste("mean = ", round(mean(COR1), 4) ), xlim=c(-1,1) )

sort(COR1)[c(50,950)]   #- Simulated 90% empirical CI for Correlation of 0.
```

Spurious Regression

- When you have two independent non-stationary time series, sample correlation tends to be high.
- When correlation is high, regression parameters will be deemed significant.z
- Can't trust the result of regression, when two time series are non-stationary.

```
#-- X1 and X2 are both ARMA(1,1), but independent  --
X1  <- arima.sim(n = 250, list(order = c(1,1,1), ar = c(0.8), ma = c(-.24) ))
X2  <- arima.sim(n = 250, list(order = c(1,1,1), ar = c(0.8), ma = c(-.24) ))

Reg1 <- lm(X1~X2)
summary(Reg1)

#-- They have nothing to do with each other, and correlation says so --
plot(X1, type="l", ylim=c(min(c(X1,X2)), max(c(X1,X2))), main=paste("cor=", round(cor(X1,X2), 4) ))
lines(X2, col="red")
cor(X1,X2)
```

I(d) notation (order of integration)

- Series is called I(d) of taking difference d times make it stationary.

- Notation:

$$
\begin{aligned}
X_t \sim I(1) \quad &:= \quad X_t \text{ is integrated series of order 1} \\
&= \quad \triangledown X_t \text{ is stationary}
\end{aligned}
$$

- e.g. If X_t is ARIMA(2,1,2), then $X_t \sim I(1)$.

What to do?

- When two non-stationary series have something to do with each other, they may look like this:

```
#- When two non-stationary series have relationship
X1  <- arima.sim(n = 250, list(order = c(1,1,1), ar = c(0.8), ma = c(-.24) ))
X2  <- 5 + .5*X1 + rnorm(251)

plot(X1, type="l", ylim=c(min(c(X1,X2)), max(c(X1,X2))), main=paste("cor=", round(cor(X1,X2), 4) ))
lines(X2, col="red")
cor(X1,X2)
```

2.3 Engle-Granger Method

- Let's check if the residual from regression is stationary or not.

```
source("http://gozips.uakron.edu/~nmimoto/477/TS_R-90.txt")

#- When two non-stationary series have relationship
X1  <- arima.sim(n = 250, list(order = c(1,1,1), ar = c(0.8), ma = c(-.24) ))
X2  <- 5 + .5*X1 + rnorm(251)

Reg1 <- lm(X1~X2)
St.test1 <- Stationarity.tests(Reg1$resid)
summary(Reg1)
St.test1
cor(X1,X2)



layout(matrix(1:2, 2, 1))

plot(X1, type="l", ylim=c(min(c(X1,X2)), max(c(X1,X2))), main=paste("cor=", round(cor(X1,X2), 4) ))
lines(X2, col="red")

plot(Reg1$resid, type="o")
```

E-G method on independent series

- If we use E-G method on two independent ARIMA(1,1,1), how many can we catch?

```
COR1  <- 0
ST.T1 <- 0
for (i in 1:1000) {

  X1  <- arima.sim(n = 250, list(order=c(1,1,1), ar = c(0.8), ma = c(-.24) ))
  X2  <- arima.sim(n = 250, list(order=c(1,1,1), ar = c(0.8), ma = c(-.24) ))

  Reg1    <- lm(X1~X2)
  COR1[i] <- cor(X1,X2)
  ST.T1[i] <- Stationarity.tests(Reg1$resid)[2]  #- use adf test p-val

}

hist(COR1, main=paste("mean = ", round(mean(COR1), 4) ), xlim=c(-1,1) )

cbind(COR1, ST.T1)

coint <- (ST.T1<.05)   #- Residual is stationary according to adf.test
cbind(COR1[coint], ST.T1[coint])

sum(coint)   #- number of spurious reg not caught by E-G method (out of 1000)
```

Cointegration

Many Economic Theory implies cointegrated relationship

- Money Demand Model
- Permanent Income Model
- Unbiased Forward Rates Hypothesis
- Fisher Equation