> #the file is name HKexchange

hk<-read.csv(file.choose(), header=TRUE)

attach(hk)

The following objects are masked from hk (pos = 3):

DATE, EXEUROUS, EXHKUS

plot(EXHKUS~DATE)



plot(EXHKUS~Time)



> model1=lm(EXHKUS ~ Time)

> summary(model1)

Call:

lm(formula = EXHKUS ~ Time)

Residuals:

Min 1Q Median 3Q Max

-0.024280 -0.014618 -0.003042 0.010123 0.037104

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.762e+00 1.518e-03 5113.540 <2e-16 \*\*\*

Time 4.702e-05 5.230e-06 8.991 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01698 on 500 degrees of freedom

Multiple R-squared: 0.1392, Adjusted R-squared: 0.1375

F-statistic: 80.84 on 1 and 500 DF, p-value: < 2.2e-16

#when we look at the plot,there is an obvious curved patttern

#and not a linear one,we need a sqr term here

> #In the summary output,we see that time significant since

> #the p-value is small, but because the R sqr are small

> #it is also not explaining a lot of the EXHKUS by itself

> #Time2 = Time\*Time

> Time2 = Time\*Time

> model2=lm(EXHKUS ~ Time + Time2)

> summary(model2)

Call:

lm(formula = EXHKUS ~ Time + Time2)

Residuals:

Min 1Q Median 3Q Max

-0.0232655 -0.0046462 0.0004868 0.0047363 0.0154038

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.797e+00 9.139e-04 8531.86 <2e-16 \*\*\*

Time -3.689e-04 8.391e-06 -43.97 <2e-16 \*\*\*

Time2 8.269e-07 1.615e-08 51.19 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.006798 on 499 degrees of freedom

Multiple R-squared: 0.8623, Adjusted R-squared: 0.8617

F-statistic: 1562 on 2 and 499 DF, p-value: < 2.2e-16

> # the Time sqr var is also significant since its p-value = 0 and

> #and the R-Sqr have increased dramatically

> #EXHKUS = 7.797 -.0003689(Time)+.00000008269(Time2)

> 7.797 -.0003689\*503 +.0000008269\*503\*503

[1] 7.820656

prescrip<-read.csv(file.choose(), header=TRUE)

> #this is the "Priscription Drug"

> attach(prescrip)

> plot(TIME,COST\_CLM)



> #the overall increase that looks like the linear trend shows that this is

> #not stationary in the mean since the values of COST\_CLM are increasing

> #with time

> #The overall spread in the points is also increasing so it is not stationary

> #in the variation, It has heterocedasticity

> COST\_CLM

[1] 14.54708 14.67209 14.80160 14.61254 14.29674 14.32503 14.62313 15.13620 15.21997

[10] 15.75915 16.07239 16.11256 16.18262 16.45369 15.98852 16.16458 16.67333 16.50048

[19] 16.84832 16.59351 17.43912 17.98944 17.80312 18.51412 18.81800 19.21166 18.96502

[28] 18.79720 19.16668 18.97395 18.59330 19.33125 20.15026 20.98675 21.18684 21.88781

[37] 21.85776 22.47405 22.16013 21.95031 21.99894 21.89219 22.23643 23.03491 23.46581

[46] 24.65132 24.73391 25.28700 25.32644 25.78771 25.19883 25.41002 25.28996 24.57147

[55] 24.29247 25.49047 26.16846 27.08572 27.47660 28.21639 27.99084 28.15354 27.53914

[64] 27.90920 26.82556 27.97149 29.47726 30.66242

> COST\_CLM1= COST\_CLM[-1]

> COST\_CLM1

[1] 14.67209 14.80160 14.61254 14.29674 14.32503 14.62313 15.13620 15.21997 15.75915

[10] 16.07239 16.11256 16.18262 16.45369 15.98852 16.16458 16.67333 16.50048 16.84832

[19] 16.59351 17.43912 17.98944 17.80312 18.51412 18.81800 19.21166 18.96502 18.79720

[28] 19.16668 18.97395 18.59330 19.33125 20.15026 20.98675 21.18684 21.88781 21.85776

[37] 22.47405 22.16013 21.95031 21.99894 21.89219 22.23643 23.03491 23.46581 24.65132

[46] 24.73391 25.28700 25.32644 25.78771 25.19883 25.41002 25.28996 24.57147 24.29247

[55] 25.49047 26.16846 27.08572 27.47660 28.21639 27.99084 28.15354 27.53914 27.90920

[64] 26.82556 27.97149 29.47726 30.66242

> #WE remove the 1st value,and shift each value back one

> #concern:linearity ?//Independent?//

> diff = COST\_CLM1-COST\_CLM[1:(length(COST\_CLM)-1)]

> #This will make the diff from period to the next in COST\_CLM.you

> #HAVE to shorten the lengths since in the last spot there is no item to

> #subtract in COST\_CLM1

#uSING THE diff instead of yja original values often helps us make

#the obs be independent instead of the dependemt originals,

> #we want independence since it is a regression assumption

> diff

[1] 0.12500668 0.12950897 -0.18905258 -0.31580258 0.02828694 0.29809951

[7] 0.51307488 0.08376884 0.53918076 0.31323814 0.04017067 0.07006455

[13] 0.27106094 -0.46516800 0.17606163 0.50875282 -0.17285537 0.34784126

[19] -0.25480843 0.84560776 0.55032540 -0.18632699 0.71100617 0.30387497

[25] 0.39366722 -0.24664497 -0.16781426 0.36947250 -0.19272804 -0.38064956

[31] 0.73795127 0.81900597 0.83649063 0.20009041 0.70097542 -0.03005409

[37] 0.61629104 -0.31392288 -0.20981407 0.04862213 -0.10674858 0.34424591

[43] 0.79847717 0.43090439 1.18550682 0.08259201 0.55308724 0.03943824

[49] 0.46127129 -0.58888245 0.21119309 -0.12005615 -0.71849060 -0.27900124

[55] 1.19799805 0.67799186 0.91725350 0.39088058 0.73979569 -0.22555542

[61] 0.16270446 -0.61439895 0.37005615 -1.08364296 1.14592934 1.50577545

[67] 1.18515778

> plot(TIME,diff)

Error in xy.coords(x, y, xlabel, ylabel, log) :

'x' and 'y' lengths differ

> #therr were 68 original time points,but only 67 diff

> plot(TIME[1:67],diff)



> #stationary in the mean (it all hovers around .1 or .2)

but not stationary in the variance

> #as the megaphone shape shows us the spread is

> #increasing as we move forward through time