E392: Problem Set 6

Estimating the Coefficient of Relative Risk Aversion in the US

Spring 2018

Due: April 4th 2018

Please work on the following questions and hand in your solutions in groups of at most 2 students. You are asked to answer all questions, but we will only select 2 (sub)questions randomly to grade.

The problem is estimation of the Cofficient of Relative Risk Aversion (RRA) in the US. This is a very important parameter in economics, that is often obtained as a regression of the real return on an asset on consumption growth. We use quaterly data from Campbell (2003) for the US, which has been posted in Canvas (USAQ.txt). We use the real interest rate (rf) as the dependent variable in the analysis and consumption growth as predictor (dc). Load (use `read.delim("USAQ.txt")`) and summarize the data.

Question 1: Load, prepare and summarize the data.

```r
setwd("C:/Users/jescanci/Dropbox/teaching/2017-2018/e390-bigdata/Rcode")
# You will need to change this to your address.
usa=read.delim("USAQ.txt")
attach(usa)
summary(usa)
```

```
##       DATE            r                    dp                rf
##  Min.   :1947   Min.   :-0.277729   Min.   :-4.171   Min.   :0.0008696
##  1st Qu.:1960   1st Qu.:-0.009654   1st Qu.:-3.501   1st Qu.:0.0065634
##  Median :1973   Median : 0.036969   Median :-3.305   Median :0.0117161
##  Mean   :1973   Mean   : 0.029730   Mean   :-3.296   Mean   :0.0119036
##  3rd Qu.:1986   3rd Qu.: 0.078906   3rd Qu.:-3.083   3rd Qu.:0.0158045
##  Max.   :1998   Max.   : 0.216497   Max.   :-2.624   Max.   :0.0374114
##
##       inf                  dc                   rr
##  Min.   :-0.017866   Min.   :-0.012533   Min.   :-0.30989
##  1st Qu.: 0.004144   1st Qu.: 0.002062   1st Qu.:-0.02174
##  Median : 0.007885   Median : 0.005167   Median : 0.02975
##  Mean   : 0.009778   Mean   : 0.004944   Mean   : 0.01995
##  3rd Qu.: 0.014130   3rd Qu.: 0.008283   3rd Qu.: 0.06769
##  Max.   : 0.043000   Max.   : 0.021462   Max.   : 0.20117
##
```

```
##       rrf                          z1                 z2                  z3
##  Min.   :-0.042120   -3.52907621:  2   0.01448459:  3   0.00000000 : 13
##  1st Qu.:-0.001211   .          :  2   .         :  2   .          :  2
##  Median : 0.002551   -2.62443082:  1   0.00705505:  2   -0.00123424:  1
##  Mean   : 0.002126   -2.70298804:  1   0.00706498:  2   -0.00220785:  1
##  3rd Qu.: 0.006769   -2.72112024:  1   0.00746209:  2   -0.00247969:  1
##  Max.   : 0.029566   -2.72116584:  1   0.01426773:  2   -0.00282692:  1
##                      (Other)    :200   (Other)   :195   (Other)    :189
##           z4
##  .          :  2
##  0.00865570 :  2
##  -0.00002277:  1
##  -0.00047509:  1
##  -0.00053455:  1
##  -0.00074666:  1
##  (Other)    :200
```

```
y<-usa$rf
x<-usa$dc
```

Question 2: Linear Regression Analysis

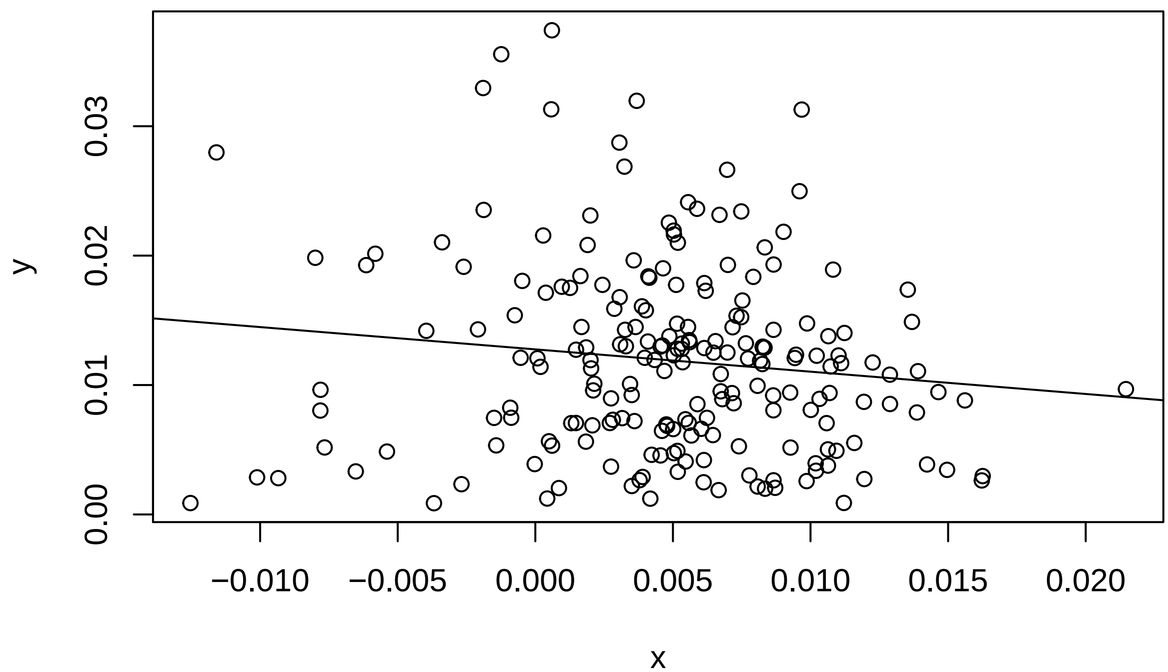
Run a linear regression and interpret the coefficients. The slope of the regression is the Coefficient of RRA. Are people risk lovers according to the data? Plot the regression line with the scatter plot.

```
fit1<-lm(y~x)
summary(fit1)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0140360 -0.0055123  0.0001318  0.0038900  0.0247594
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0127557  0.0006879  18.544   <2e-16 ***
## x           -0.1723330  0.0943185  -1.827   0.0691 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007292 on 206 degrees of freedom
```

```
## Multiple R-squared:  0.01595,    Adjusted R-squared:  0.01117
## F-statistic: 3.338 on 1 and 206 DF,  p-value: 0.06913
```

```r
plot(x,y)
abline(fit1)
```



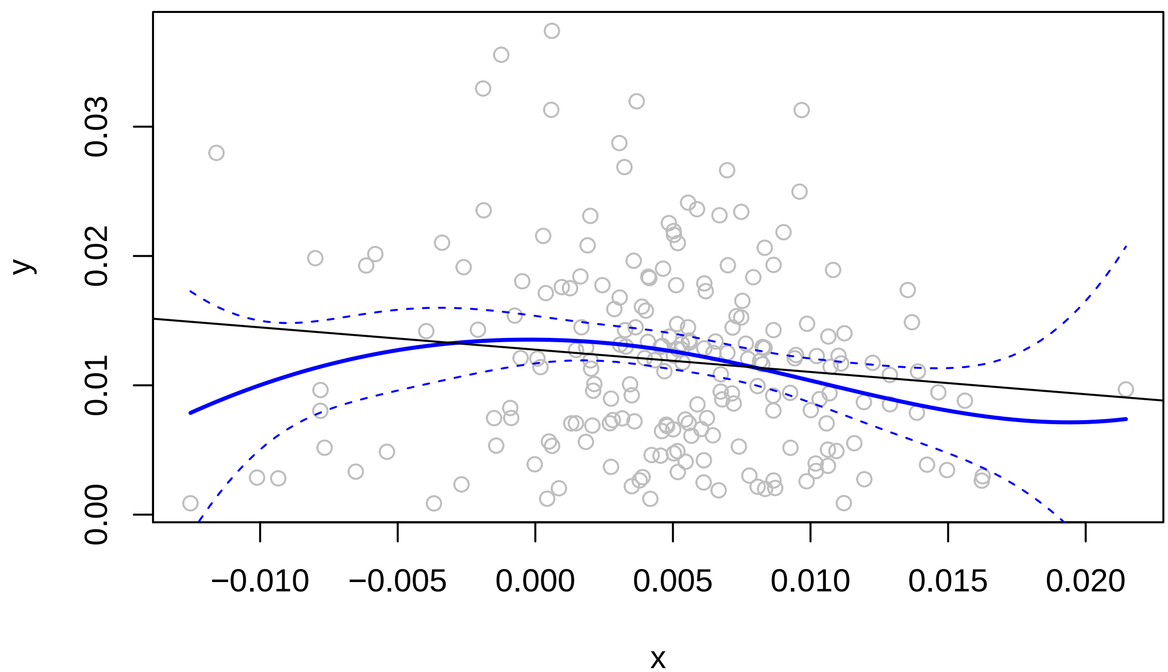
The interpretation is that an increase in consumption growth of one unit leads to a drecrease in interest rates of .17 points. A negative RRA indicates the utility is convex and people are risk lovers.

Question 3: Nonlinear Regression Analysis

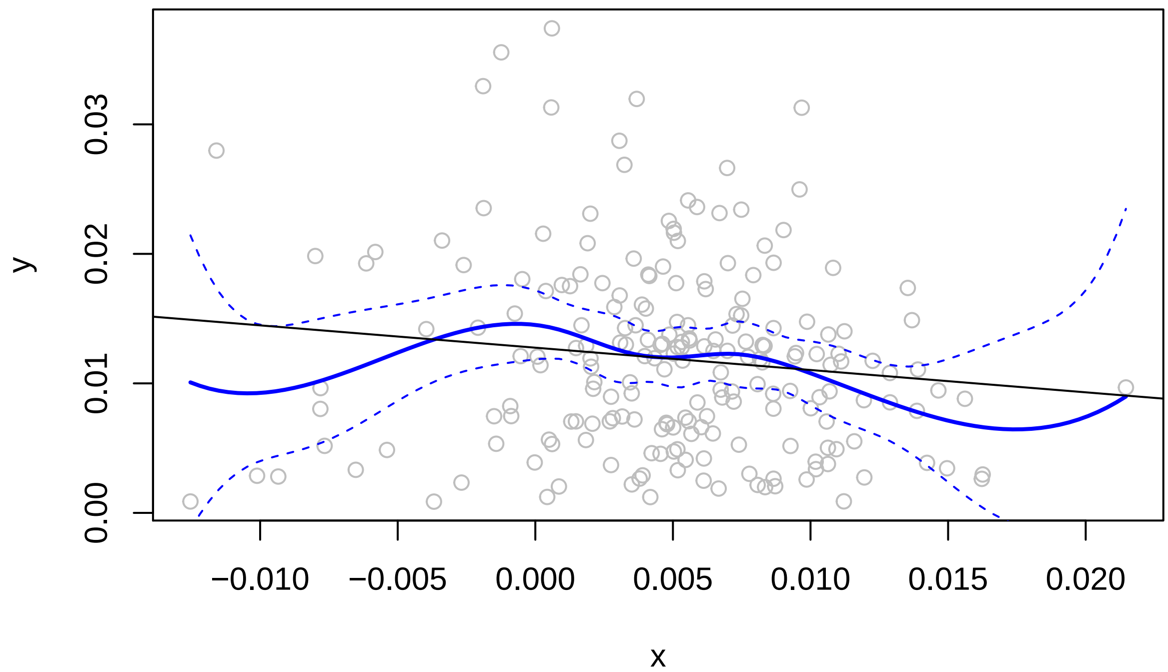
The wrong "sign" of the RRA might be due to nonlinearity bias. To that end we apply cubic splines to this data set. Estimate a cubic spline with one internal knot (at the median), and plot the cubic spline together with the regression fit. Is there evidence of nonlinearity in consumption growth? What if we use five knots instead? Does the variance of the cubic fit increase or decrease relative to the one knot case? How is the slope of the relation between interest rates and consumption growth for the most part, negative or positive?

```r
library(splines)
fit2=lm(y~bs(x,df=4)) #bs creates the cubic spline basis
xlims=range(x)
```

```
x.grid=seq(from=xlims[1],to=xlims[2],length.out = 100)
pred2=predict(fit2,newdata=list(x=x.grid),se=T)
plot(x,y,col="gray")
lines(x.grid,pred2$fit,lwd=2,col="blue")
lines(x.grid,pred2$fit+2*pred2$se,col="blue",lty="dashed")
lines(x.grid,pred2$fit-2*pred2$se,col="blue",lty="dashed")
abline(fit1)
```




```
fit2=lm(y~bs(x,df=8)) #bs creates the cubic spline basis
xlims=range(x)
x.grid=seq(from=xlims[1],to=xlims[2],length.out = 100)
pred2=predict(fit2,newdata=list(x=x.grid),se=T)
plot(x,y,col="gray")
lines(x.grid,pred2$fit,lwd=2,col="blue")
lines(x.grid,pred2$fit+2*pred2$se,col="blue",lty="dashed")
lines(x.grid,pred2$fit-2*pred2$se,col="blue",lty="dashed")
abline(fit1)
```



There is some evidence of nonlinearity, although the evidence is not very strong (the linear fit falls within the confidence bands). With more knots the variance increases, as we estimate more parameters. We can see that by the fit being more variable. For the most part the slopes are negative (except at the boundaries where there is less precision).

Question 4: Further Results on Nonlinearity

To have a better control of the bias-variance tradeoff we fit a smoothing spline with a cross-validated choice of lambda (use `smooth.spline(x,y,cv=TRUE)`). Plot the resulting fit together with a cubic spline with eight degress of freedom. According to the plot is eight degrees of freedom for cubic splines overfitting? Is there evidence of nonlinearity with smoothing splines? Does the nonlinear model imply that people are risk lovers?

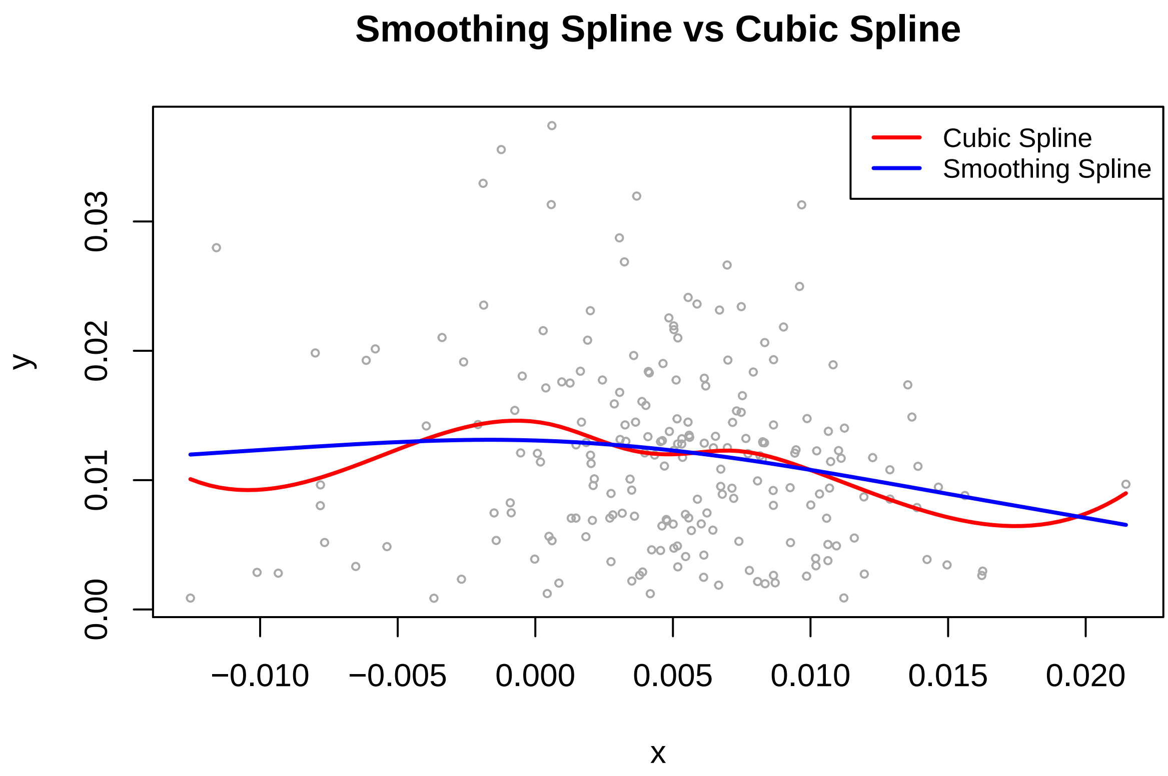
```
plot(x,y,xlim=xlims,cex=.5,col="darkgrey")
title("Smoothing Spline vs Cubic Spline")
fit4=smooth.spline(x,y,cv=TRUE)
```

```
## Warning in smooth.spline(x, y, cv = TRUE): cross-validation with non-unique
## 'x' values seems doubtful
```

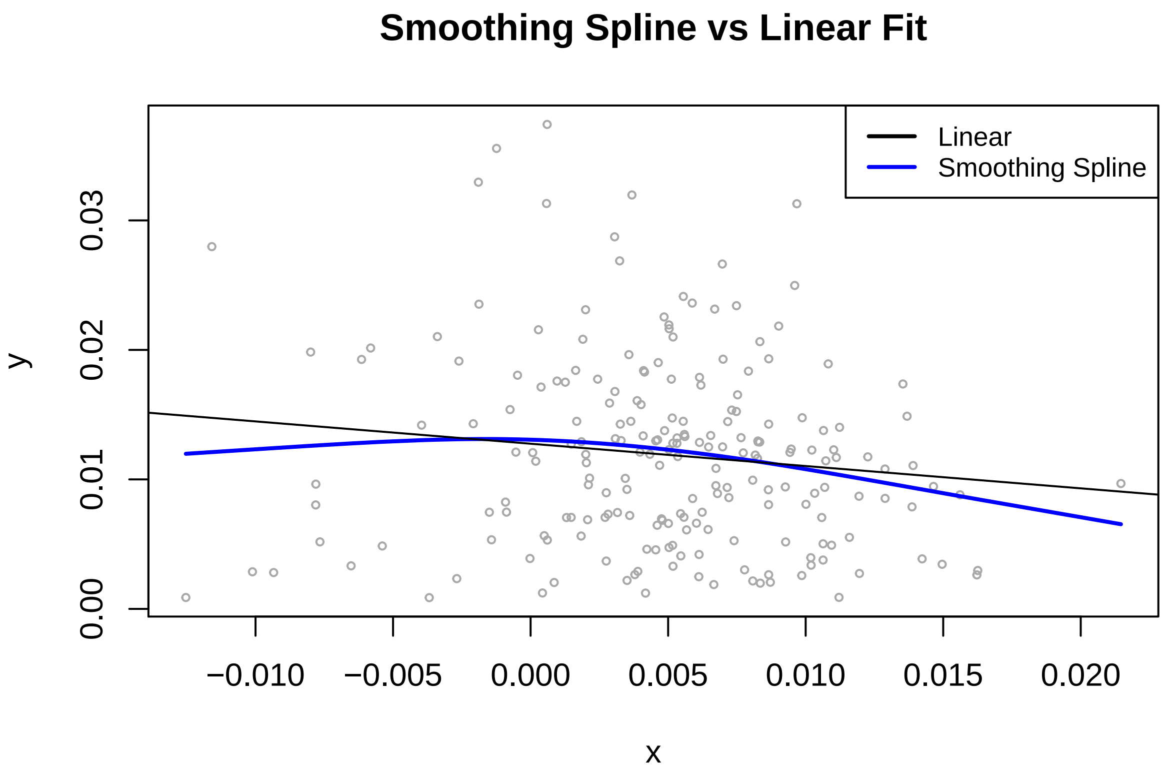
```
fit4$df
```

```
## [1] 2.838691
```

```
lines(x.grid,pred2$fit,lwd=2,col="red")
lines(fit4,col="blue",lwd=2)
legend("topright",legend=c("Cubic Spline","Smoothing Spline"),col=c("red","blue"),lty=1,
```




```
plot(x,y,xlim=xlims,cex=.5,col="darkgrey")
title("Smoothing Spline vs Linear Fit")
lines(fit4,col="blue",lwd=2)
abline(fit1)
legend("topright",legend=c("Linear","Smoothing Spline"),col=c("black","blue"),lty=1,lwd=
```



The smoothing spline is smoother than the cubic spline, which suggests overfitting (larger variance than the optimal one) for the cubic spline. The smoothing spline still has some curvature in it. Since the slope is still negative for the most part we conclude that the nonlinear models also imply risk loving. This is a bit of a puzzle, that will study later in more detail.