problemset_3b.R

Aaryan Agarwal

2023-12-02

```
# PROBLEM SET 3B
# This assignment is done in a group of three
# Prajwal Kaushal, Sumukha Sharma, Aaryan Agarwal

###############################################################################
# Question 1
###############################################################################

# Loading data
data <- read.table("USAQ.txt", header = TRUE, sep = "\t")

# inspecting it by lookinng at the first few rows
head(data)
```

```
##     DATE           r        dp         rf        inf          dc          rr
## 1 1947.1 -0.01379268 -3.049576 0.00088960 0.01998129  0.01121201 -0.03377397
## 2 1947.2 -0.00734189 -2.982244 0.00086962 0.00456233 -0.00368396 -0.01190422
## 3 1947.3  0.01849884 -2.954207 0.00087961 0.04299992 -0.01253346 -0.02450109
## 4 1947.4  0.03528509 -2.900204 0.00121926 0.01868837  0.00417873  0.01659672
## 5 1948.1 -0.00347102 -2.850472 0.00205788 0.00000000  0.00871671 -0.00347102
## 6 1948.2  0.10871987 -2.923430 0.00233727 0.02809028 -0.00268308  0.08062959
##           rrf          z1         z2         z3          z4
## 1 -0.01909169           .          .          .           .
## 2 -0.00369271           .          .          .           .
## 3 -0.04212031 -3.04957619 0.00088960 0.01998129  0.01121201
## 4 -0.01746911 -2.98224366 0.00086962 0.00456233 -0.00368396
## 5  0.00205788 -2.95420653 0.00087961 0.04299992 -0.01253346
## 6 -0.02575301 -2.90020394 0.00121926 0.01868837  0.00417873
```

```
# checking the structure
str(data)
```

```
## 'data.frame':    208 obs. of  12 variables:
##  $ DATE: num  1947 1947 1947 1947 1948 ...
##  $ r   : num  -0.01379 -0.00734 0.0185 0.03529 -0.00347 ...
##  $ dp  : num  -3.05 -2.98 -2.95 -2.9 -2.85 ...
##  $ rf  : num  0.00089 0.00087 0.00088 0.00122 0.00206 ...
##  $ inf : num  0.01998 0.00456 0.043 0.01869 0 ...
##  $ dc  : num  0.01121 -0.00368 -0.01253 0.00418 0.00872 ...
##  $ rr  : num  -0.03377 -0.0119 -0.0245 0.0166 -0.00347 ...
##  $ rrf : num  -0.01909 -0.00369 -0.04212 -0.01747 0.00206 ...
```

```
## $ z1  : chr  "." "." "-3.04957619" "-2.98224366" ...
## $ z2  : chr  "." "." "0.00088960" "0.00086962" ...
## $ z3  : chr  "." "." "0.01998129" "0.00456233" ...
## $ z4  : chr  "." "." "0.01121201" "-0.00368396" ...
```

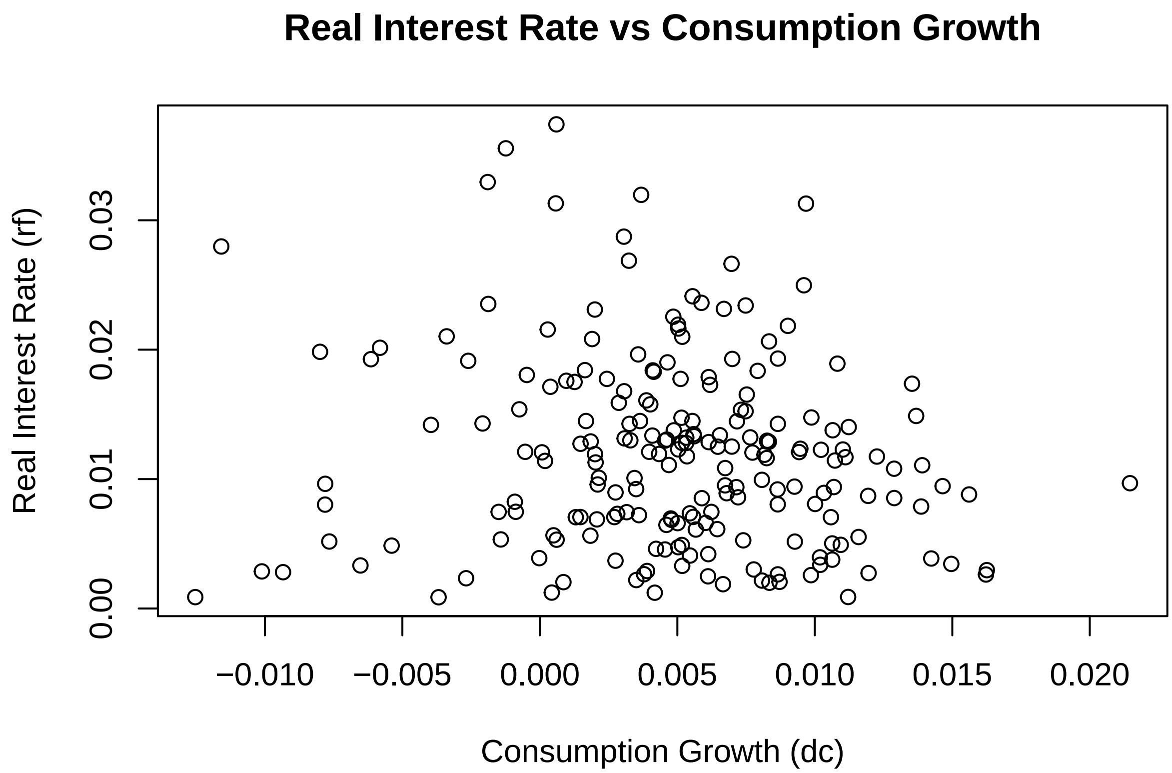
```r
# summary
summary(data)
```

```
##       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                 rrf
## Min.   :-0.017866   Min.   :-0.012533   Min.   :-0.30989   Min.   :-0.042120
## 1st Qu.: 0.004144   1st Qu.: 0.002062   1st Qu.:-0.02174   1st Qu.:-0.001211
## Median : 0.007885   Median : 0.005167   Median : 0.02975   Median : 0.002551
## Mean   : 0.009778   Mean   : 0.004944   Mean   : 0.01995   Mean   : 0.002126
## 3rd Qu.: 0.014130   3rd Qu.: 0.008283   3rd Qu.: 0.06769   3rd Qu.: 0.006769
## Max.   : 0.043000   Max.   : 0.021462   Max.   : 0.20117   Max.   : 0.029566
##       z1                 z2                 z3                 z4
## Length:208         Length:208         Length:208         Length:208
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```r
# subset of the data focusing on real interest rate(rf) and consumption growth(dc)
analysis_data <- data[, c("rf", "dc")]

# Summary for the subset
summary(analysis_data)
```

```
##        rf                  dc
## Min.   :0.0008696   Min.   :-0.012533
## 1st Qu.:0.0065634   1st Qu.: 0.002062
## Median :0.0117161   Median : 0.005167
## Mean   :0.0119036   Mean   : 0.004944
## 3rd Qu.:0.0158045   3rd Qu.: 0.008283
## Max.   :0.0374114   Max.   : 0.021462
```

```r
# Plotting the data to see relationship between dc and rf
plot(analysis_data$dc, analysis_data$rf, main = "Real Interest Rate vs Consumption Growth",
     xlab = "Consumption Growth (dc)", ylab = "Real Interest Rate (rf)")
```

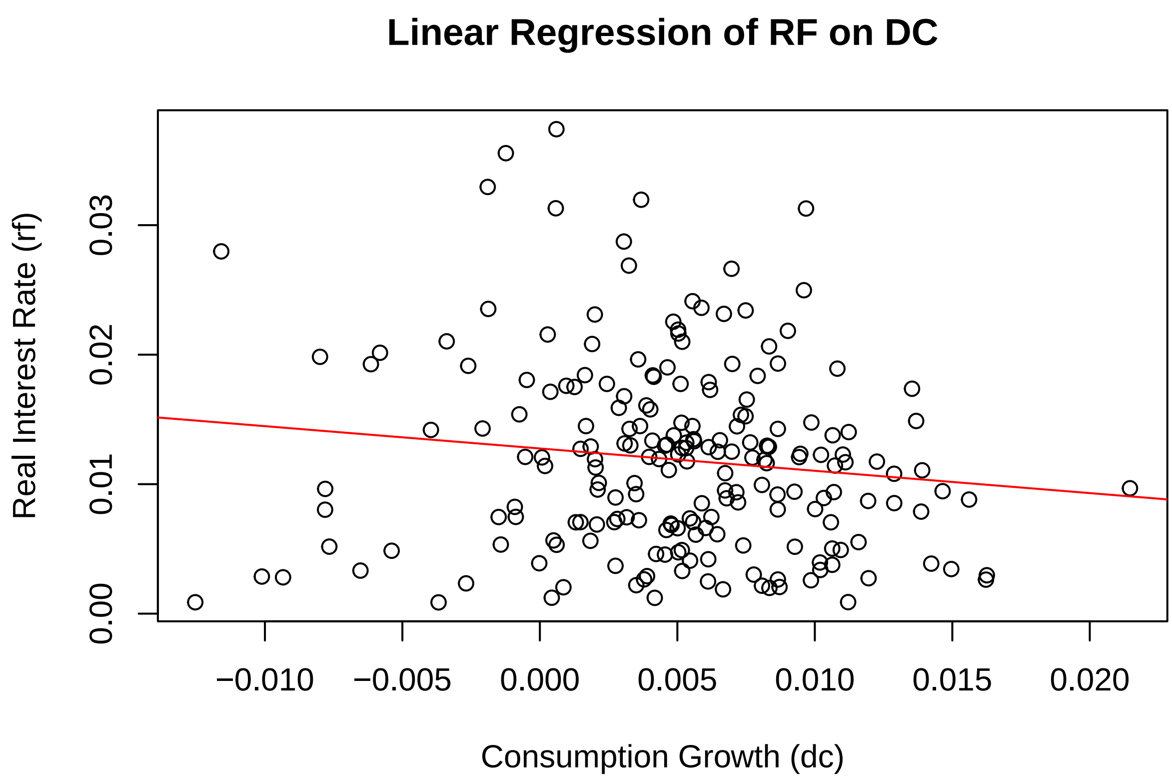



```r
##############################################################################
# Question 2
##############################################################################

# linear regression
model <- lm(rf ~ dc, data = analysis_data)

# summary of the regression model to interpret the coefficients
summary(model)
```

```
##
## Call:
## lm(formula = rf ~ dc, data = analysis_data)
##
## 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 ***
## dc          -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
```

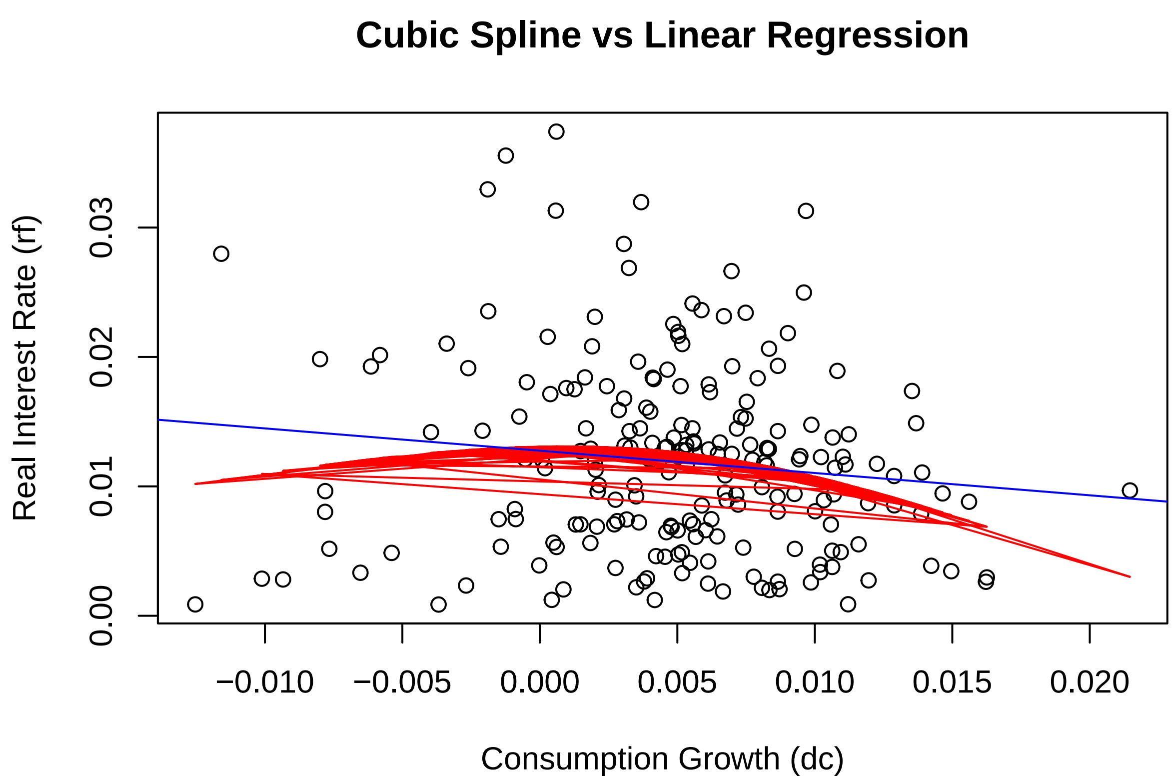
```
## Multiple R-squared:  0.01595,    Adjusted R-squared:  0.01117
## F-statistic: 3.338 on 1 and 206 DF,  p-value: 0.06913
```

```
# Plotting the original data points
plot(analysis_data$dc, analysis_data$rf, main = "Linear Regression of RF on DC",
     xlab = "Consumption Growth (dc)", ylab = "Real Interest Rate (rf)")

# Adding the regression line to the plot
abline(model, col = "red")
```

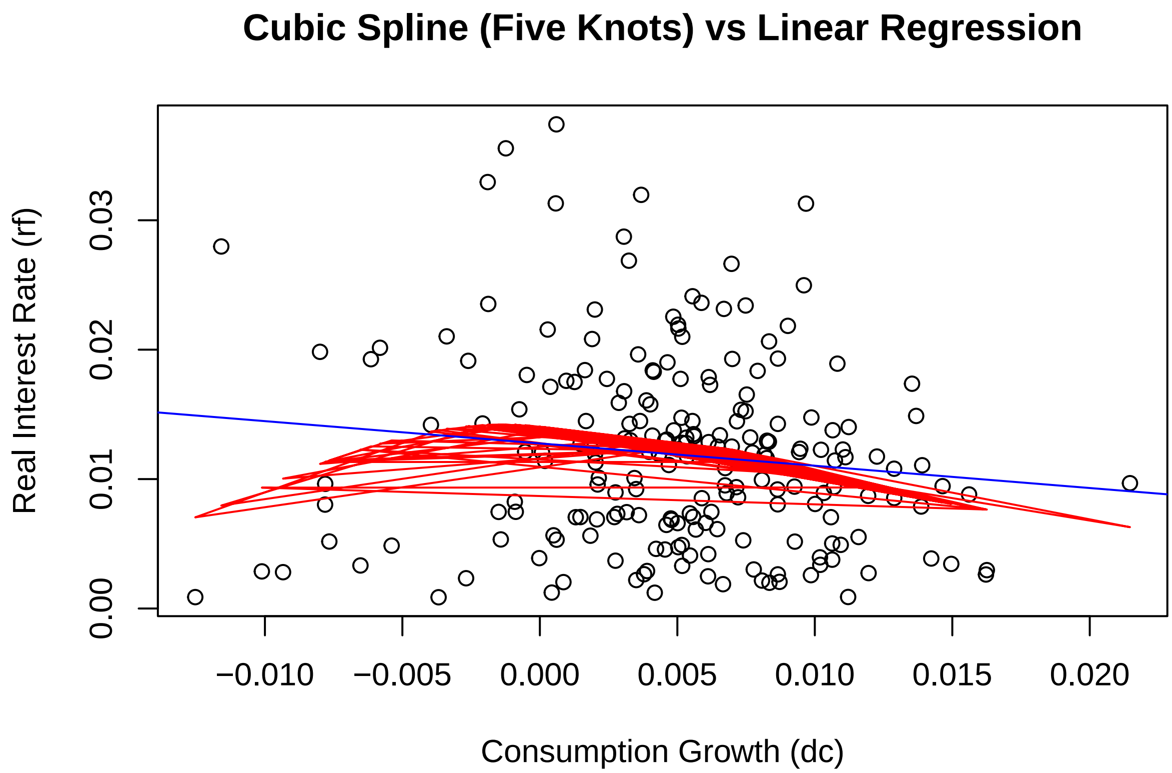



```
# Intercept (0.0127557): This is the expected value of rf when dc is zero.
# The intercept is significantly different from zero, as indicated by the p-value
# (<2e-16, which is practically zero).

# dc Coefficient (-0.1723330): This is the estimated coefficient for consumption growth.
# A negative coefficient shows that there is an inverse relationship between consumption
# growth and the real interest rate.
# negative value implies risk lovinng behavior.

##############################################################################
# Question 3
##############################################################################

library(splines)
```

```r
# cubic spline with one internal knot
median_dc <- median(analysis_data$dc)
model_spline_one_knot <- lm(rf ~ ns(dc, knots = median_dc), data = analysis_data)

# Plotting the cubic spline with the linear regression
plot(analysis_data$dc, analysis_data$rf, main = "Cubic Spline vs Linear Regression",
     xlab = "Consumption Growth (dc)", ylab = "Real Interest Rate (rf)")
points(analysis_data$dc, predict(model_spline_one_knot), col = "red", type = "l")
abline(lm(rf ~ dc, data = analysis_data), col = "blue")
```




```r
# cubic spline with five internal knots
quantiles_dc <- quantile(analysis_data$dc, probs = c(0.2, 0.4, 0.6, 0.8))
model_spline_five_knots <- lm(rf ~ ns(dc, knots = quantiles_dc), data = analysis_data)

# Plotting the data with the cubic spline (five knots)
plot(analysis_data$dc, analysis_data$rf, main = "Cubic Spline (Five Knots) vs Linear Regression",
     xlab = "Consumption Growth (dc)", ylab = "Real Interest Rate (rf)")
points(analysis_data$dc, predict(model_spline_five_knots), col = "red", type = "l")
abline(lm(rf ~ dc, data = analysis_data), col = "blue")
```




```r
# Compare the variance of the one knot and five knot models
summary(model_spline_one_knot)
```

```
##
## Call:
## lm(formula = rf ~ ns(dc, knots = median_dc), data = analysis_data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0118548 -0.0057351 -0.0002457  0.0039253  0.0243245
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.0101869  0.0027358   3.724 0.000254 ***
## ns(dc, knots = median_dc)1  0.0007682  0.0051313   0.150 0.881135
## ns(dc, knots = median_dc)2 -0.0097575  0.0034049  -2.866 0.004595 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007224 on 205 degrees of freedom
## Multiple R-squared:  0.03891,    Adjusted R-squared:  0.02953
## F-statistic: 4.149 on 2 and 205 DF,  p-value: 0.01712
```

```r
summary(model_spline_five_knots)
```

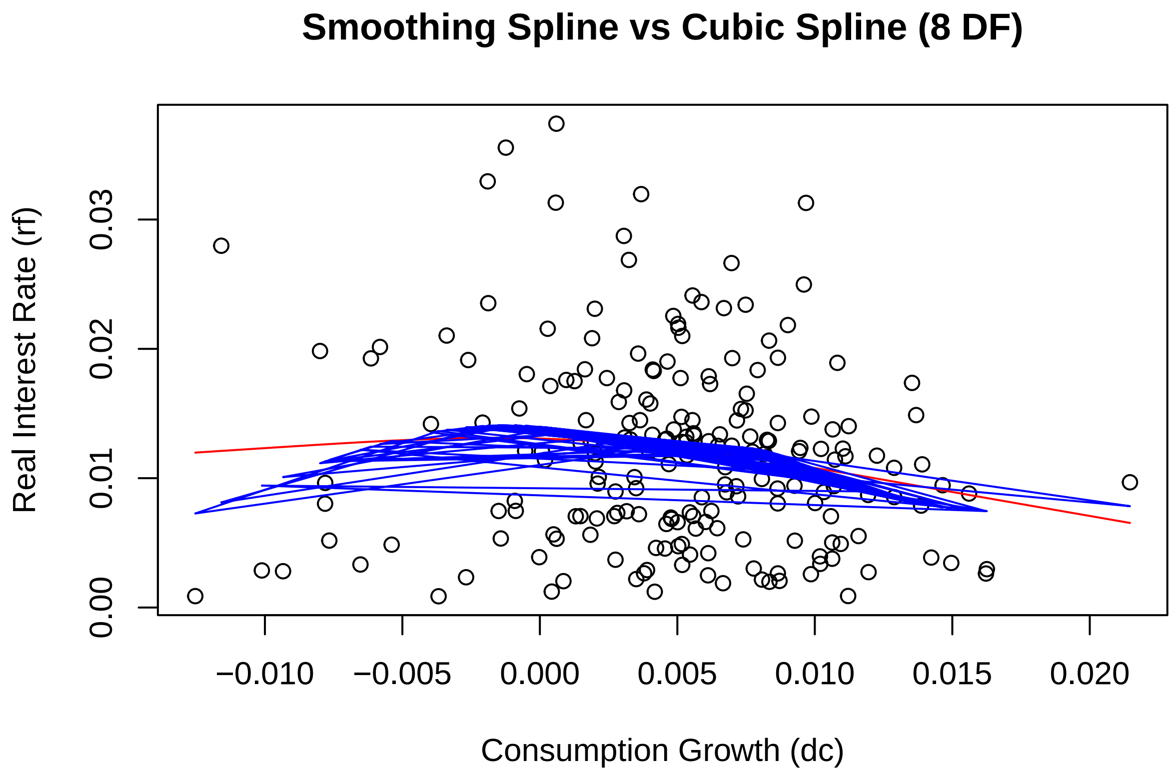
```
##
## Call:
## lm(formula = rf ~ ns(dc, knots = quantiles_dc), data = analysis_data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0129157 -0.0056947 -0.0000513  0.0038501  0.0235440
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    0.007046   0.003807   1.851   0.0657 .
## ns(dc, knots = quantiles_dc)1  0.004663   0.003705   1.258   0.2097
## ns(dc, knots = quantiles_dc)2  0.005898   0.004332   1.361   0.1749
## ns(dc, knots = quantiles_dc)3 -0.002512   0.003583  -0.701   0.4841
## ns(dc, knots = quantiles_dc)4  0.008999   0.009355   0.962   0.3372
## ns(dc, knots = quantiles_dc)5 -0.005637   0.006074  -0.928   0.3545
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007249 on 202 degrees of freedom
## Multiple R-squared:  0.04639,    Adjusted R-squared:  0.02279
## F-statistic: 1.965 on 5 and 202 DF,  p-value: 0.08526
```

```
#summary(fit_splines)


# The coefficients of the cubic spline terms are mixed.
# The residual standard error is 0.007224, which measures the standard deviation of the residuals.
# Multiple R-square is 0.03891, which means a slight improvement compared to the linear model.

# The coefficients are not statistically significant, as indicated by their p-values.
# This suggests a more complex, possibly non-linear relationship.
# The residual standard error is 0.007249, slightly higher than the one-knot model,
# => a marginal increase in the variance.
# Multiple R-squared is 0.04639, which is slightly higher than the one-knot model,
# => a small improvement in the model's ability to explain the variability in rf.

# the slope is negative

##################################################################################
# Question 4
##################################################################################

# smoothing spline with cross-validation
smooth_spline_model <- smooth.spline(analysis_data$dc, analysis_data$rf, cv = TRUE)
```

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

```
# cubic spline with eight degrees of freedom
cubic_spline_model <- lm(rf ~ ns(dc, df = 8), data = analysis_data)

# Plotting the data with the smoothing spline and cubic spline fits
```

```r
plot(analysis_data$dc, analysis_data$rf, main = "Smoothing Spline vs Cubic Spline (8 DF)",
     xlab = "Consumption Growth (dc)", ylab = "Real Interest Rate (rf)")
lines(smooth_spline_model, col = "red")
points(analysis_data$dc, predict(cubic_spline_model), col = "blue", type = "l")
```




```r
# Interpretation based on the plot
# The smoothing spline (red line) shows a curve that deviates from a straight line,
# => there are nonlinearities in the relationship between dc and (rf).
# The cubic spline with eight degrees of freedom (blue lines) follows the data's minor
# fluctuations more closely than the smoothing spline.
# This could indicate that the cubic spline is overfitting the data
# The overall slope of the smoothing spline seems to trend downwards as consumption growth increases.
# This would suggest risk-loving behavior.
```