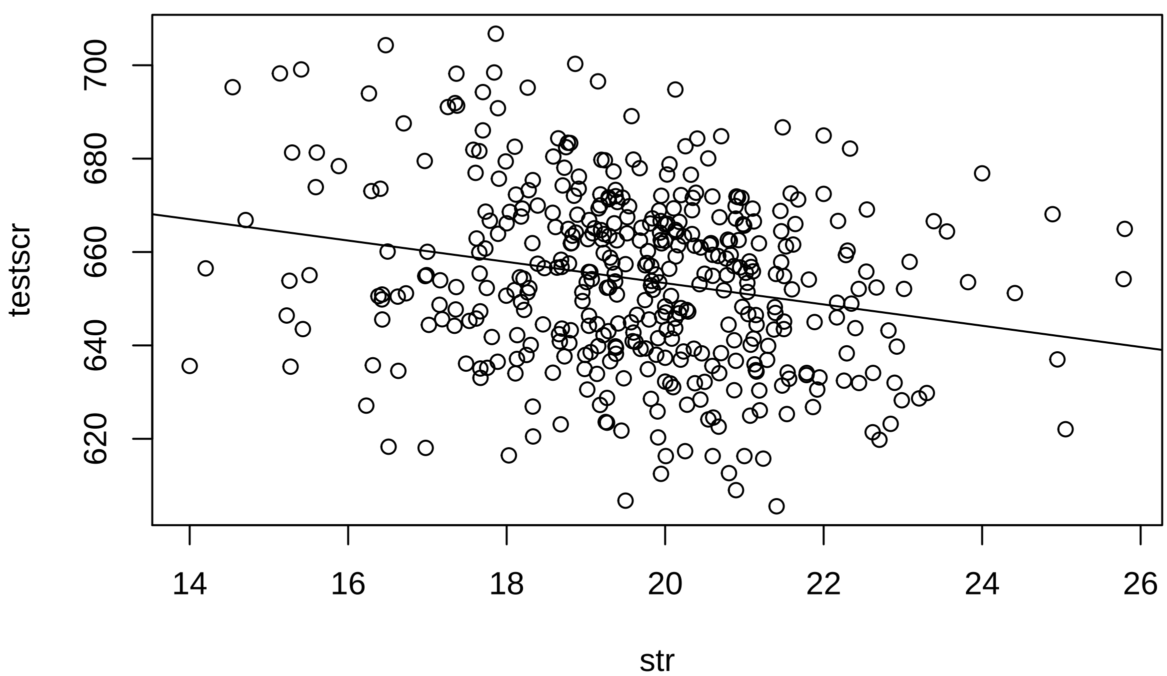
Chapter 8 - Nonlinear Regression Functions with One Regressor

Ercio Munoz

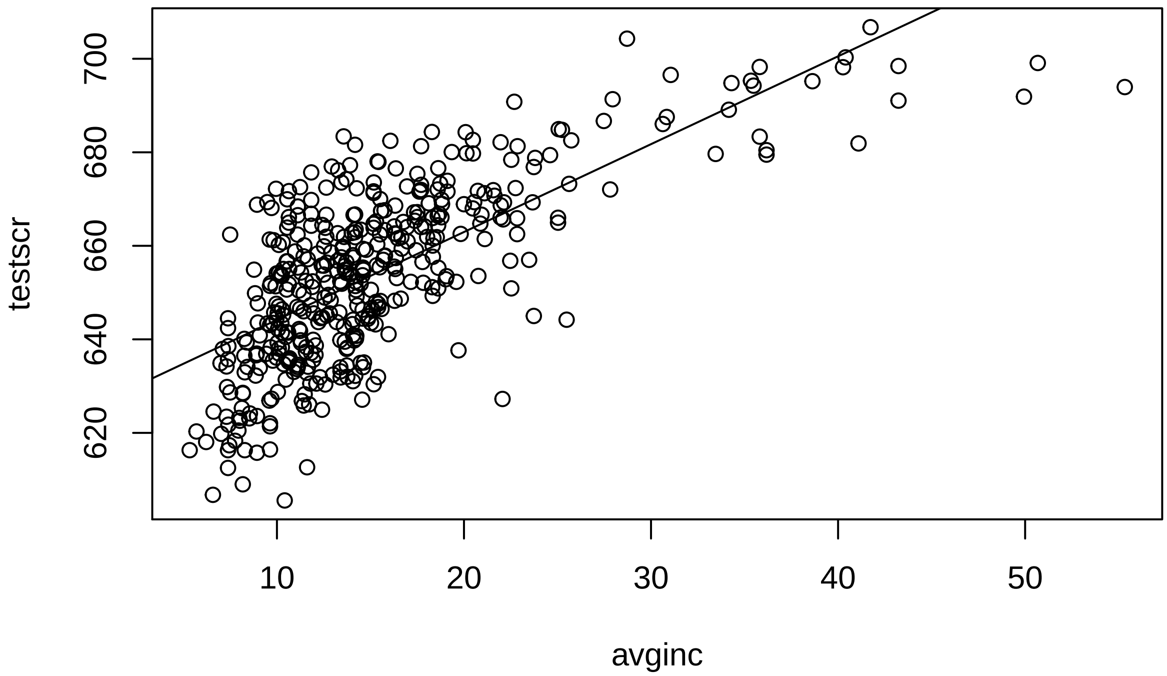
October 24, 2018

In this chapter, we tackle the problem of nonlinear relationships between a dependent variable y and a regressor x. We order the data set according to average income in order to have well behaved lines. First, we fit test scores against str and average income.

```
library(sandwich)
library(lmtest)
library(foreign)
a=
"http://fmwww.bc.edu/ec-p/data/stockwatson/caschool.dta"
data_set = read.dta(a)
odata = data_set[order(data_set$avginc),]
attach(odata)
reg = lm(testscr~str)
plot(str,testscr)
abline(reg)
```




```
reg2 = lm(testscr~avginc)
plot(avginc,testscr)
abline(reg2)
```

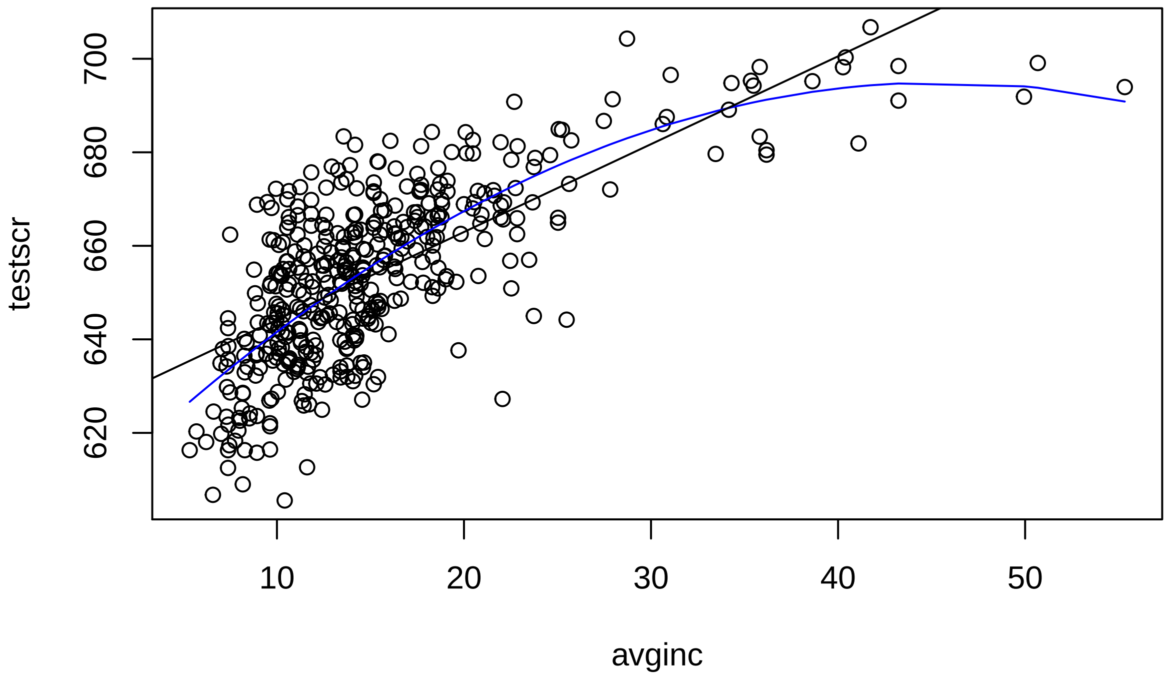


Now, we try a quadratic function on average income. We run a regression and plot the fitted values of this quadratic function.

```
avginc2 = avginc*avginc
reg3 = lm(testscr~avginc+avginc2)
coeftest(reg3, vcov = vcovHC(reg3, "HC1"))
```

```
##
## t test of coefficients:
##
##                 Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 607.3017350   2.9017539 209.2878 < 2.2e-16 ***
## avginc        3.8509947   0.2680941  14.3643 < 2.2e-16 ***
## avginc2      -0.0423085   0.0047803  -8.8505 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(avginc,testscr)
lines(avginc,reg3$fitted.values,col='blue')
abline(reg2)
```



Estimating a cubic function on average income and testing the null hypothesis of a linear relationship:

```
avginc3 = avginc*avginc2
reg4 = lm(testscr~avginc+avginc2+avginc3)
coeftest(reg4, vcov = vcovHC(reg4, "HC1"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)  6.0008e+02  5.1021e+00 117.6150 < 2.2e-16 ***
## avginc       5.0187e+00  7.0735e-01   7.0950 5.606e-12 ***
## avginc2     -9.5805e-02  2.8954e-02  -3.3089  0.001018 **
## avginc3      6.8548e-04  3.4706e-04   1.9751  0.048919 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

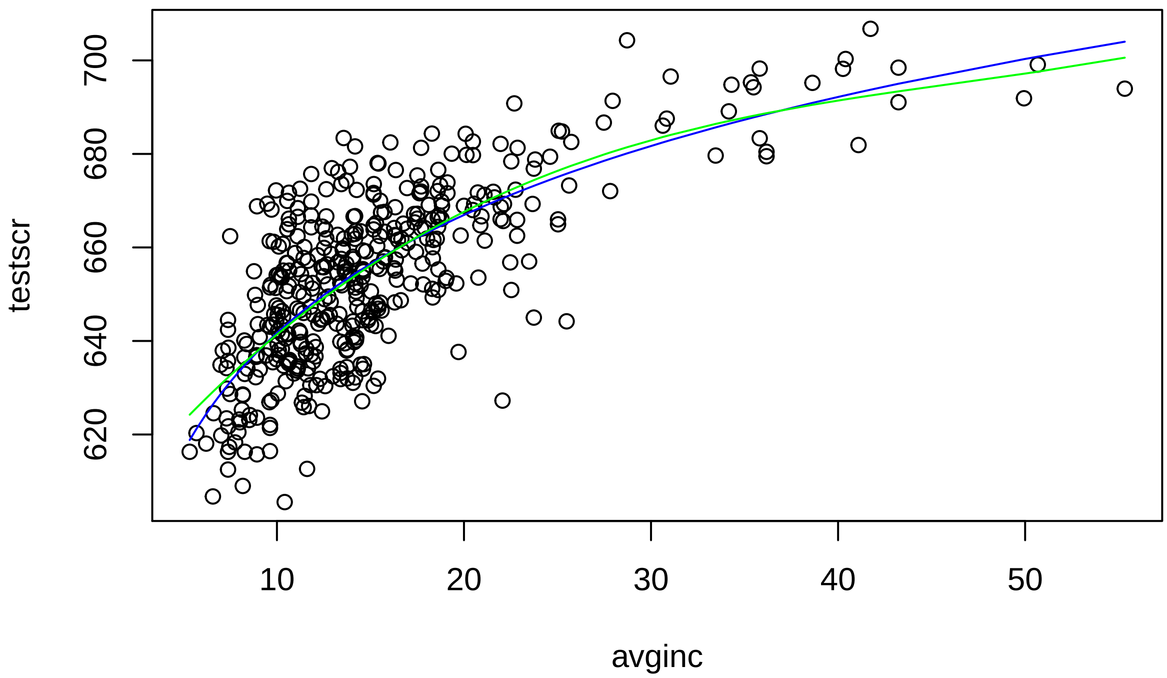
```
library(car)
myH0 <- c("avginc2=0","avginc3=0")
linearHypothesis(reg4, myH0, vcov = vcovHC(reg4, "HC1"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## avginc2 = 0
## avginc3 = 0
##
## Model 1: restricted model
## Model 2: testscr ~ avginc + avginc2 + avginc3
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1    418
```

```
## 2    416  2 37.691 9.043e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

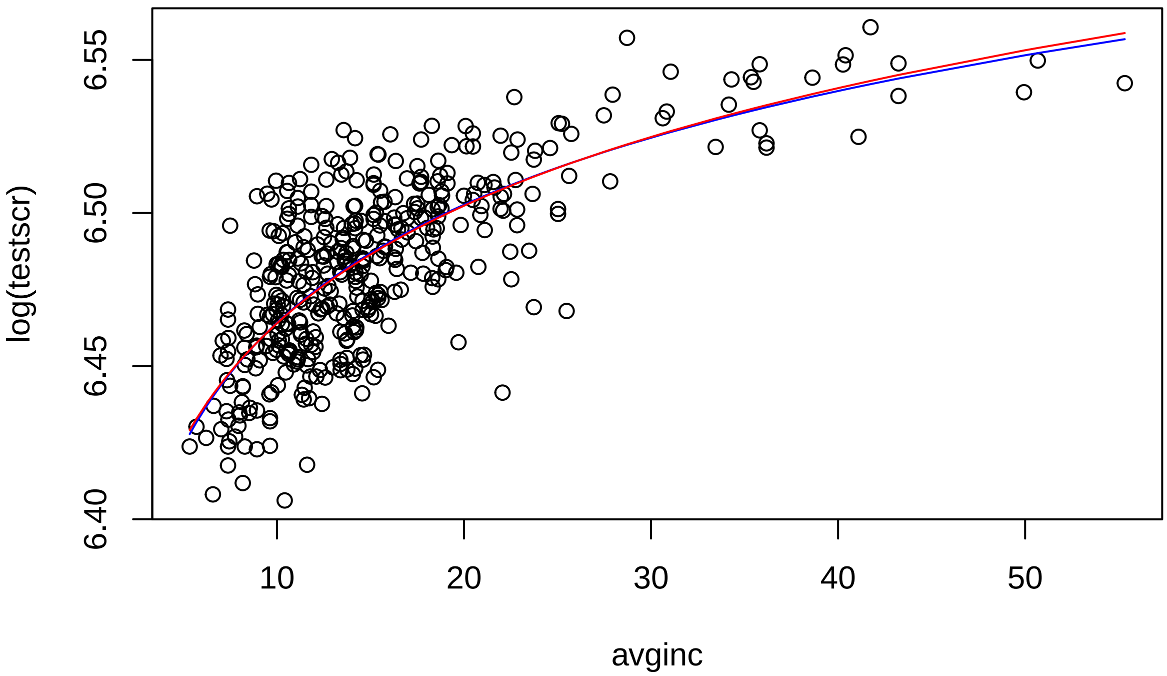
Now estimate a linear-log model and plot the fitted values in blue:

```r
reg5 = lm(testscr~log(avginc))
plot(avginc,testscr)
lines(avginc,reg5$fitted.values,col='blue')
lines(avginc,reg4$fitted.values,col='green')
```



Now a log-log model and a plot of the fitted values in red:

```r
reg6 = lm(log(testscr)~log(avginc))
plot(avginc,log(testscr))
lines(avginc,log(reg5$fitted.values),col='blue')
lines(avginc,reg6$fitted.values,col='red')
```

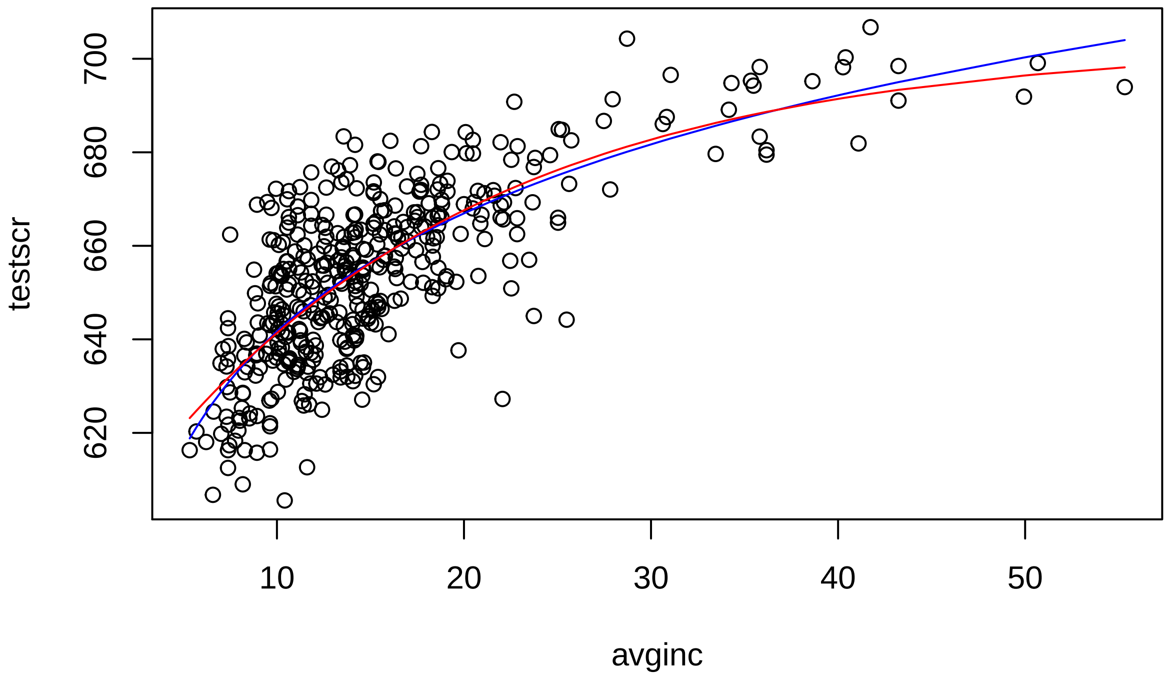


Nonlinear least squares:

```
reg7 = nls(testscr~b0*(1-exp(b1*(avginc-b2))), start = c(b0=720, b1=-.05, b2=-34))
summary(reg7)
```

```
##
## Formula: testscr ~ b0 * (1 - exp(b1 * (avginc - b2)))
##
## Parameters:
##     Estimate Std. Error t value Pr(>|t|)
## b0 703.22279    6.69762 104.996  < 2e-16 ***
## b1  -0.05523    0.00910  -6.069 2.89e-09 ***
## b2 -34.00415    5.67691  -5.990 4.54e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.67 on 417 degrees of freedom
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 4.075e-06
```

```
plot(avginc,testscr)
lines(avginc,fitted(reg5),col='blue')
lines(avginc,fitted(reg7),col='red')
```



Finally, a table comparing different nonlinear specifications and their respective p-values of a test with the null hypothesis of a linear relationship:

```r
model1 = lm(testscr~str+el_pct+meal_pct)
model2 = lm(testscr~str+el_pct+meal_pct+log(avginc))
HiEL = (el_pct>=10)
HiELstr = HiEL*str
model3 = lm(testscr~str+HiEL+HiELstr)
model4 = lm(testscr~str+HiEL+HiELstr+meal_pct+log(avginc))
str2 = str^2
str3 = str^3
model5 = lm(testscr~str+str2+str3+HiEL+meal_pct+log(avginc))
HiELstr2 = HiEL*str^2
HiELstr3 = HiEL*str^3
model6 = lm(testscr~str+str2+str3+HiEL+HiELstr+HiELstr2+HiELstr3+meal_pct+log(avginc))
model7 = lm(testscr~str+str2+str3+el_pct+meal_pct+log(avginc))

library(sandwich)
library(lmtest)

rse_1 = sqrt(diag(vcovHC(model1, type = "HC1")))
F_1 = waldtest(model1, vcov = vcovHC(model1, type = "HC1"))
rse_2 = sqrt(diag(vcovHC(model2, type = "HC1")))
F_2 = waldtest(model2, vcov = vcovHC(model2, type = "HC1"))
rse_3 = sqrt(diag(vcovHC(model3, type = "HC1")))
F_3 = waldtest(model3, vcov = vcovHC(model3, type = "HC1"))
rse_4 = sqrt(diag(vcovHC(model4, type = "HC1")))
F_4 = waldtest(model4, vcov = vcovHC(model4, type = "HC1"))
rse_5 = sqrt(diag(vcovHC(model5, type = "HC1")))
F_5 = waldtest(model5, vcov = vcovHC(model5, type = "HC1"))
rse_6 = sqrt(diag(vcovHC(model6, type = "HC1")))
F_6 = waldtest(model6, vcov = vcovHC(model6, type = "HC1"))
rse_7 = sqrt(diag(vcovHC(model7, type = "HC1")))
```

```
F_7 = waldtest(model7, vcov = vcovHC(model7, type = "HC1"))

pv_a3 = linearHypothesis(model3,c("str","HiELstr"),vcov=vcovHC(model3, "HC1"))
pv_a4 = linearHypothesis(model4,c("str","HiELstr"),vcov=vcovHC(model4, "HC1"))
pv_a5 = linearHypothesis(model5,c("str","str2","str3"),vcov=vcovHC(model5, "HC1"))
pv_a6 = linearHypothesis(model6,c("str","str2","str3","HiELstr","HiELstr2","HiELstr3"),vcov=vcovHC(mode
pv_a7 = linearHypothesis(model7,c("str","str2","str3"),vcov=vcovHC(model7, "HC1"))

pv_b5 = linearHypothesis(model5,c("str2","str3"),vcov=vcovHC(model5, "HC1"))
pv_b6 = linearHypothesis(model6,c("str2","str3"),vcov=vcovHC(model6, "HC1"))
pv_b7 = linearHypothesis(model7,c("str2","str3"),vcov=vcovHC(model7, "HC1"))

pv_c6 = linearHypothesis(model6,c("HiELstr","HiELstr2","HiELstr3"),vcov=vcovHC(model6, "HC1"))

library(stargazer)
stargazer(list(model1,model2,model3,model4,model5,model6,model7),
          type="text",keep.stat=c("rsq","n"),
          se=list(rse_1,rse_2,rse_3,rse_4,rse_5,rse_6,rse_7),
          add.lines = list(c("a)","","",format(pv_a3$`Pr(>F)`[2],digits=2),format(pv_a4$`Pr(>F)`[2],dig
                        c("b)","","","","",format(pv_b5$`Pr(>F)`[2],digits=2),format(pv_b6$`Pr(>F)`[2],
          c("c)","","","","","",format(pv_c6$`Pr(>F)`[2],digits=2),"")))
```

```
##
## ==========================================================================================
##                                        Dependent variable:
##              ----------------------------------------------------------------------------
##                                               testscr
##                  (1)        (2)        (3)        (4)        (5)       (6)        (7)
## ------------------------------------------------------------------------------------------
## str          -0.998***  -0.734***   -0.968     -0.531   64.339*** 83.701*** 65.285***
##               (0.270)    (0.257)   (0.589)    (0.342)   (24.861)  (28.497)  (25.259)
##
## el_pct       -0.122***  -0.176***                                             -0.166***
##               (0.033)    (0.034)                                               (0.034)
##
## str2                                                    -3.424*** -4.381***  -3.466***
##                                                          (1.250)   (1.441)    (1.271)
##
## str3                                                    0.059***   0.075***  0.060***
##                                                          (0.021)   (0.024)    (0.021)
##
## meal_pct     -0.547***  -0.398***             -0.411*** -0.420*** -0.418***  -0.402***
##               (0.024)    (0.033)               (0.029)   (0.029)   (0.029)    (0.033)
##
## log(avginc)             11.569***             12.124*** 11.748*** 11.800***  11.509***
##                          (1.819)               (1.798)   (1.771)   (1.778)    (1.806)
##
## HiEL                                 5.639      5.498   -5.474*** 816.075**
##                                    (19.515)   (9.795)    (1.034)  (327.674)
##
## HiELstr                             -1.277     -0.578             -123.282**
##                                     (0.967)    (0.496)              (50.213)
##
## HiELstr2                                                            6.121**
```

```
##                                                                        (2.542)
##
## HiELstr3                                                               -0.101**
##                                                                        (0.043)
##
## Constant     700.150*** 658.552*** 682.246*** 653.666***  252.051     122.354    244.809
##               (5.568)    (8.642)    (11.868)   (9.869)   (163.634)  (185.519)  (165.722)
##
## ------------------------------------------------------------------------------------------
## a)                                   0.0038     0.0029    0.00034    6.4e-05    0.00059
## b)                                                         0.0023     0.0033     0.0028
## c)                                                                    0.046
## Observations   420        420        420        420        420        420        420
## R2            0.775      0.796      0.310      0.797      0.801      0.803      0.801
## ==========================================================================================
## Note:                                                         *p<0.1; **p<0.05; ***p<0.01
```