E392: Problem Set 7

Estimating the Coefficient of Relative Risk Aversion in the US II

Spring 2018

Due: April 18th 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.

Our previous analysis revealed a negative coefficient of Relative Risk Aversion (RRA) with US data using OLS. According to economic theory this means that economic agents are risk loving, which seems a puzzle. We continue using the data set from Campbell (2003) for the US, which has been posted in Canvas (USAQ.txt).

Question 1: OLS Results.

From the discussion in class, is OLS a good estimator for the RRA parameter? Explain. SOL: OLS is biased, because consumption growth is endogenous.

Question 2: TSLS Results I

Using the real interest rate (rf) as the dependent variable, consumption growth as the variables of interest (dc) and four instruments, which are the lagged values of the nominal interest rate (z1), inflation (z2), consumption growth (z3) and the log dividend-price ratio (z4), run the first-stage regression (endogenous variable on instruments). Why is this regression useful for IV analysis? Are the instruments valid? Get the fitted values of the first-stage and run the second stage. What is the sign of the TSLS estimator of the RRA? Compare with OLS from the previous Homework. Does this TSLS solve the puzzle?

```
rm(list = ls())
setwd("C:/Users/jescanci/Dropbox/teaching/2017-2018/e390-bigdata/Rcode")
usa=read.delim("USAQ.txt")
usa=usa[3:208,]
attach(usa)
dim(usa)
```

```
## [1] 206  12
```

```
y<-usa$rf
x<-usa$dc
z1<-as.numeric(usa$z1)
```

```r
z2<-as.numeric(usa$z2)
z3<-as.numeric(usa$z3)
z4<-as.numeric(usa$z4)
```

```r
z<-cbind(z1,z2,z3,z4)
fit2<-lm(x~z)
summary(fit2)
```

```
##
## Call:
## lm(formula = x ~ z)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0168871 -0.0023263  0.0001446  0.0032673  0.0139087
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.539e-03  1.294e-03   3.508 0.000556 ***
## zz1          6.309e-06  6.370e-06   0.990 0.323128
## zz2          5.548e-06  7.375e-06   0.752 0.452758
## zz3         -2.033e-05  7.508e-06  -2.708 0.007347 **
## zz4          1.061e-05  6.295e-06   1.686 0.093326 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005187 on 201 degrees of freedom
## Multiple R-squared:  0.07772,    Adjusted R-squared:  0.05936
## F-statistic: 4.234 on 4 and 201 DF,  p-value: 0.002593
```

```r
xhat<-fit2$fitted.values
fit3<-lm(y~xhat)
summary(fit3)
```

```
##
## Call:
## lm(formula = y ~ xhat)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0156946 -0.0045257  0.0001993  0.0033147  0.0217606
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.019392   0.001686   11.50  < 2e-16 ***
```

```
## xhat        -1.489487   0.325921   -4.57 8.44e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006958 on 204 degrees of freedom
## Multiple R-squared:  0.09287,    Adjusted R-squared:  0.08843
## F-statistic: 20.89 on 1 and 204 DF,  p-value: 8.438e-06
```

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

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0142823 -0.0056124  0.0001483  0.0038472  0.0246162
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0129036  0.0006886  18.739   <2e-16 ***
## x           -0.1801831  0.0945599  -1.905   0.0581 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007241 on 204 degrees of freedom
## Multiple R-squared:  0.01749,    Adjusted R-squared:  0.01267
## F-statistic: 3.631 on 1 and 204 DF,  p-value: 0.05812
```

SOL: The first-stage regression is useful to see if the instruments predict the endogenous variable (that is, if the instruments are valid, correlated with x). The very low value of the F test reveals the predictive power of the instruments is very low, so they don't seem to be valid. The sign of the TSLS is still negative, and indeed, much more negative than OLS. TSLS does not solve the puzzle.

Question 3: TSLS Results II

The previous two-step approach can be done in one shot using the command `tslsl` from the `hdm` package. This avoids using the wrong standard error provided by the second step using OLS. Run tsls and check you get the same coefficient but different standard error. Is the correct standard error from tsls larger or smaller than the wrong standard error from the second-stage OLS?

```r
fit4<-tsls(y~x|z,data=usa)
summary(fit4)
```

```
## [1] "Estimates and Significance Testing from from tsls"
##              Estimate Std. Error t value  p value
## x           -1.489487   0.472418  -3.153  0.00162 **
## (Intercept)  0.019392   0.002444   7.933 2.13e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SOL: Using tsls the coefficient is the same but the standard error is larger with tsls, as expected since it accounts for the uncertainty of the first-stage.

Question 4: TSLS Results III

It seems that the previous instrumental variables are weak, in the sense that they are not able to predict x. Create a large matrix of instruments by binding splines for each of the four previous instruments with df=8 (use e.g `bs(z1,df=8)`). The number of new instruments should be 32. Run tsls with all instruments. To avoid a large variance from using all instruments, use Lasso to select the instruments in the first stage and run TSLS post-lasso. This can be done in one shot with `lasso.IV<-rlassoIV(y~x|z,select.X = FALSE,select.Z = TRUE)`. Report the new TSLS estimator of the Risk aversion parameter. Is it now consistent with risk aversion? How many instruments does Lasso with cross-validated lambda select? Use cv.glmnet from glmnet and plot.

```r
z=cbind(bs(z1,df=8),bs(z2,df=8),bs(z3,df=8),bs(z4,df=8))
dim(z)
```

```
## [1] 206  32
```

```r
fit5<-tsls(y~x|z,data=usa)
summary(fit5)
```

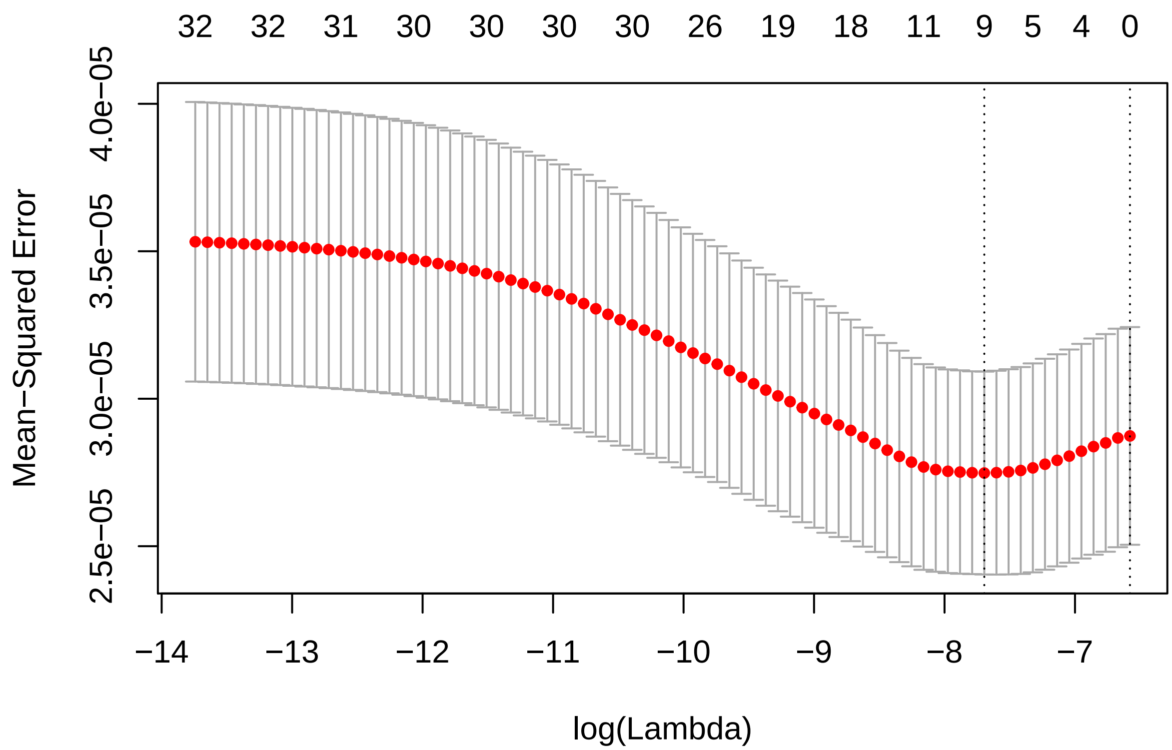
```
## [1] "Estimates and Significance Testing from from tsls"
##              Estimate Std. Error t value p value
## x           -0.648322   0.211370  -3.067 0.00216 **
## (Intercept)  0.015224   0.001176  12.948 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lasso.IV<-rlassoIV(y~x|z,select.X = FALSE,select.Z = TRUE)
summary(lasso.IV)
```

```
## [1] "Estimates and significance testing of the effect of target variables in the IV r
##   coeff.   se. t-value p-value
## x  2.424 0.220   11.02  <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lasso.mod=glmnet(z,x,alpha=1)
cv.out=cv.glmnet(z,x,alpha=1)
plot(cv.out)
```



SOL: The coefficient now is 2.4, positive, which is consistent with risk aversion of agents. Nevertheless, the high sensitive of estimates to the selected instruments is an indication of weak identification (low signal on the parameter of interest). Lasso selects 10 instruments out of the 32.