Project

Sk Minhazul Islam

Tuesday, September 02, 2014

Executive Summary

We look at estimate an OLS model, regressing mileage on an automatic/manual dummy. Find a significantly negative effect of automatic on mileage. We redo the estimation as TSLS, using weight as an instument for automatic. These results are even more significant.

Analysis

We start by loading the **mtcars** data frame and recode **am** to a logical **automatic** variable.

```
data( mtcars )
mtcars$automatic <- mtcars$am == 0
attach( mtcars )
```

We do some exploratory data analysis on the distribution of the automatic variable (Appendix).

We estimate the Ordinary Least Squares (OLS) model.

```
ols <- lm( mpg ~ automatic, data=mtcars )
summary( ols )
```

```
##
## Call:
## lm(formula = mpg ~ automatic, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     24.392      1.360  17.941  < 2e-16 ***
## automaticTRUE   -7.245      1.764  -4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

There is a significant negative effect of automatic on milage.

We redo the estimation using Two-Stage Least Squares (TSLS), using weight (**wt**) as an instrument for automatic.

```r
tsls.fs <- lm(automatic ~ wt)
tsls <- lm(mpg ~ tsls.fs$fitted.values)
summary(tsls)
```

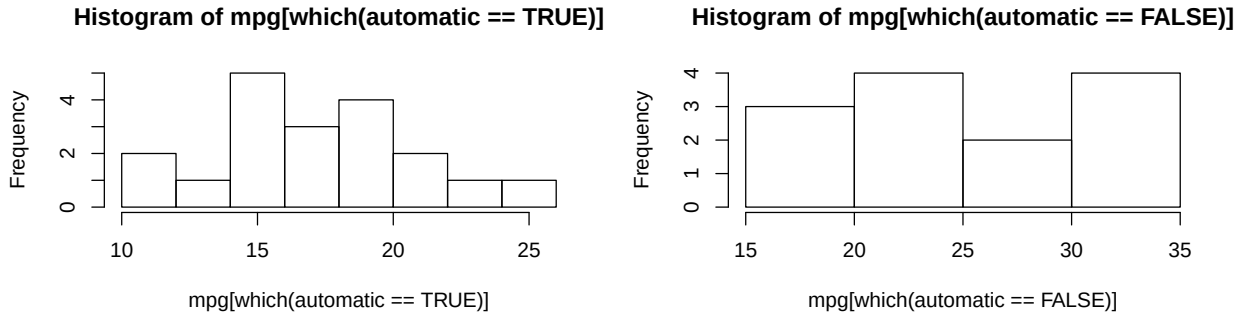
```
##
## Call:
## lm(formula = mpg ~ tsls.fs$fitted.values)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               29.076      1.083  26.841  < 2e-16 ***
## tsls.fs$fitted.values    -15.133      1.583  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

This coefficient is even more negative, and even more significant. We finally plot the residuals for both estimations (Appendix).

The residuals of the OLS estimation appear to be normally distibuted, the TSLS somewhat less, though **n** is too small to make any real claims.

Appendix: Figures

```r
par( mfrow=c(1,2) )
hist( mpg[which(automatic == TRUE)] )
hist( mpg[which(automatic == FALSE)] )
```



We plot the residuals of the OLS and TSLS estimations.

```
par( mfrow=c(1,3) )
hist( ols$residuals )
hist(tsls.fs$residuals)
hist( tsls$residuals )
```