Motor Trend - Bastiaan Quast

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 )
```

```
## The following objects are masked from mtcars (position 3):
##
##     am, automatic, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs,
##     wt
## The following objects are masked from mtcars (position 4):
##
##     am, automatic, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs,
##     wt
## The following objects are masked from mtcars (position 5):
##
##     am, automatic, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs,
##     wt
## The following objects are masked from mtcars (position 7):
##
##     am, automatic, carb, cyl, disp, drat, gear, hp, mpg, qsec, vs,
##     wt
```

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
## -9.392 -3.092 -0.297  3.244  9.508
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      24.39       1.36   17.94  < 2e-16 ***
## automaticTRUE    -7.24       1.76   -4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 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)
```

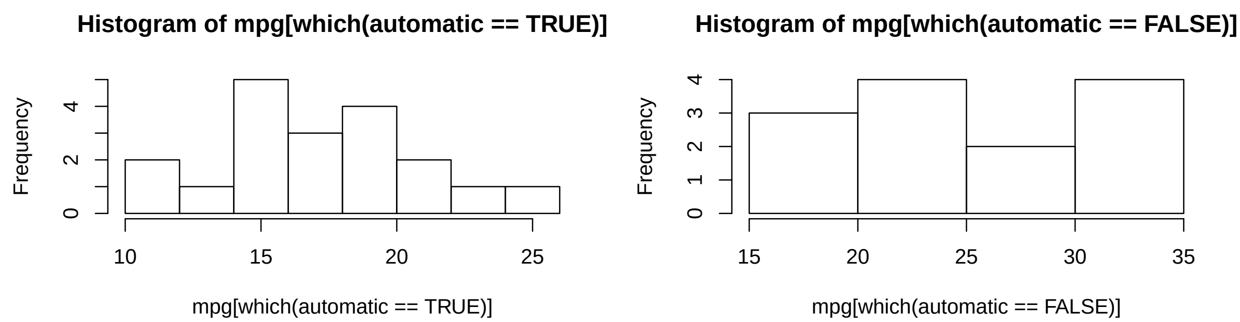
```
##
## Call:
## lm(formula = mpg ~ tsls.fs$fitted.values)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.543 -2.365 -0.125  1.410  6.873
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)              29.08       1.08   26.84  < 2e-16 ***
## tsls.fs$fitted.values   -15.13       1.58   -9.56  1.3e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.05 on 30 degrees of freedom
## Multiple R-squared:  0.753,  Adjusted R-squared:  0.745
## F-statistic: 91.4 on 1 and 30 DF,  p-value: 1.29e-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 )
```