PS7 R Solutions

Matthew Alampay Davis

December 9, 2021

Question 1

This is non-empirical. See the official solutions.

Question 2

```
wage <- read.dta13("wage2.dta")
head(wage)
```

```
##   wage hours  IQ KWW educ exper tenure age married black south urban sibs
## 1  769    40  93  35   12    11      2  31       1     0     0     1    1
## 2  808    50 119  41   18    11     16  37       1     0     0     1    1
## 3  825    40 108  46   14    11      9  33       1     0     0     1    1
## 4  650    40  96  32   12    13      7  32       1     0     0     1    4
## 5  562    40  74  27   11    14      5  34       1     0     0     1   10
## 6 1400    40 116  43   16    14      2  35       1     1     0     1    1
##   brthord meduc feduc    lwage
## 1       2     8     8 6.645091
## 2      NA    14    14 6.694562
## 3       2    14    14 6.715384
## 4       3    12    12 6.476973
## 5       6     6    11 6.331502
## 6       2     8    NA 7.244227
```

Part a

```
mod.2sls <- iv_robust(lwage ~ educ + exper + tenure + black |
    sibs + exper + tenure + black, data = wage, se_type = "HC0")
summary(mod.2sls)
```

```
##
## Call:
## iv_robust(formula = lwage ~ educ + exper + tenure + black | sibs +
##     exper + tenure + black, data = wage, se_type = "HC0")
##
```

```
## Standard error type:  HC0
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|)  CI Lower CI Upper  DF
## (Intercept)  5.21598   0.517418  10.081 9.422e-23  4.200535  6.23142 930
## educ         0.09363   0.031833   2.941 3.349e-03  0.031160  0.15611 930
## exper        0.02092   0.008297   2.521 1.185e-02  0.004638  0.03721 930
## tenure       0.01155   0.002757   4.189 3.064e-05  0.006138  0.01696 930
## black       -0.18333   0.050082  -3.661 2.658e-04 -0.281615 -0.08504 930
##
## Multiple R-squared:  0.1685 ,    Adjusted R-squared:  0.165
## F-statistic: 25.37 on 4 and 930 DF,  p-value: < 2.2e-16
```

Just a note here that heteroskedasticity-robust standard errors that Stata uses are HC1 except when doing a 2SLS regression, in which case it defaults to HC0. No clue why, but that's why se_type = 'stata' doesn't actually give you the Stata default in this particular case.

Part b

This question asks us to do the first-stage and second-stage regressions yourself manually:

```
# First stage
mod.manual.1 <- lm(educ ~ sibs + exper + tenure + black, data = wage)
wage$educ.hat <- mod.manual.1$fitted.values
# Second stage
mod.manual.2 <- lm(lwage ~ educ.hat + exper + tenure + black,
    data = wage)
summary(mod.manual.2)
```

```
##
## Call:
## lm(formula = lwage ~ educ.hat + exper + tenure + black, data = wage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97409 -0.25720  0.00997  0.26147  1.25198
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.215976   0.568815   9.170  < 2e-16 ***
## educ.hat     0.093632   0.035293   2.653 0.008114 **
## exper        0.020922   0.008779   2.383 0.017368 *
## tenure       0.011548   0.002868   4.027  6.1e-05 ***
## black       -0.183329   0.052476  -3.494 0.000499 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4028 on 930 degrees of freedom
## Multiple R-squared:  0.08912,    Adjusted R-squared:  0.0852
## F-statistic: 22.75 on 4 and 930 DF,  p-value: < 2.2e-16
```

The coefficient on educ.hat is the same but the standard errors are slightly too large

Part c

```
# First stage
mod.wrong.1 <- lm(educ ~ sibs, data = wage)
wage$educ.tilde <- mod.wrong.1$fitted.values
# Second stage
mod.wrong.2 <- lm(lwage ~ educ.tilde + exper + tenure + black,
    data = wage)
summary(mod.wrong.2)
```

```
##
## Call:
## lm(formula = lwage ~ educ.tilde + exper + tenure + black, data = wage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97409 -0.25720  0.00997  0.26147  1.25198
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.771022   0.360376  16.014  < 2e-16 ***
## educ.tilde   0.069975   0.026376   2.653  0.00811 **
## exper       -0.000394   0.003121  -0.126  0.89957
## tenure       0.013975   0.002691   5.193 2.54e-07 ***
## black       -0.241631   0.041528  -5.819 8.16e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4028 on 930 degrees of freedom
## Multiple R-squared:  0.08912,    Adjusted R-squared:  0.0852
## F-statistic: 22.75 on 4 and 930 DF,  p-value: < 2.2e-16
```

When instead we (incorrectly) use educ.tilde in the second stage regression, its coefficient is 0.0700, and the corresponding standard error is 0.0264. Both are too low. The reduction in the estimated return to education from about 9.4% to 7.0% is not trivial. This illustrates that it is best to avoid doing 2SLS manually.

Question 3

This is non-empirical. See official solutions.