

## Test Exercise 4

Jader Martins

10/31/2020

(a)

```
reg.lm <- lm("logw ~ educ + exper + exper^2 + smsa + south", data)
summary(reg.lm)
```

```
##
## Call:
## lm(formula = "logw ~ educ + exper + exper^2 + smsa + south",
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.70732 -0.23590  0.02543  0.25134  1.37509
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.788491   0.063148  75.829  <2e-16 ***
## educ         0.081326   0.003525  23.071  <2e-16 ***
## exper        0.040312   0.002234  18.047  <2e-16 ***
## smsa         0.154139   0.015947   9.666  <2e-16 ***
## south       -0.178928   0.014748 -12.132  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3841 on 3005 degrees of freedom
## Multiple R-squared:  0.2518, Adjusted R-squared:  0.2508
## F-statistic: 252.9 on 4 and 3005 DF,  p-value: < 2.2e-16
```

Every year in education adds to the logwage 0.0813259 . So for every year we multiply the age by  $\exp(0.081)$  which is 1.08.

(b)

```
reg.lm <- lm("educ ~ exper", data)
summary(reg.lm)
```

```
##
## Call:
## lm(formula = "educ ~ exper", data = data)
##
## Residuals:
```

```
##      Min      1Q Median      3Q      Max
## -7.827 -1.625 -0.157  1.375  5.219
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.001009   0.087262  194.83  <2e-16 ***
## exper      -0.422029   0.008926  -47.28  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.028 on 3008 degrees of freedom
## Multiple R-squared:  0.4264, Adjusted R-squared:  0.4262
## F-statistic: 2236 on 1 and 3008 DF,  p-value: < 2.2e-16
```

The experience and education years is dependent on peoples age, and age generally influences people wages, so this could cause an endogenous relation. As experience is double attributed it could approximate the variance of age.

(c)

As the residual can be correlated to experience, the variable age could mitigate this effect in the residual.

(d)

```
reg.lm <- lm("educ ~ age + age^2 + nearc + daded + momed", data)
educ.hat <- fitted.values(reg.lm)
reg.lm <- lm("logw ~ educ.hat + exper + exper^2 + smsa + south", data)
summary(reg.lm)
```

```
##
## Call:
## lm(formula = "logw ~ educ.hat + exper + exper^2 + smsa + south",
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.76098 -0.25633  0.02148  0.27536  1.49069
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.247834   0.092492  56.738  < 2e-16 ***
## educ.hat     0.064811   0.006403  10.122  < 2e-16 ***
## exper       0.012918   0.001912   6.757 1.69e-11 ***
## smsa        0.167289   0.017109   9.778  < 2e-16 ***
## south      -0.196440   0.015913 -12.345  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4098 on 3005 degrees of freedom
## Multiple R-squared:  0.1483, Adjusted R-squared:  0.1472
## F-statistic: 130.8 on 4 and 3005 DF,  p-value: < 2.2e-16
```

Now education has a lower effect on logwage, so it could be a better estimation.

(e)

```
reg.lm <- lm("educ ~ smsa + south + age + age^2 + nearc + daded + momed", data)
educ.hat <- fitted.values(reg.lm)
reg.lm <- lm("exper ~ smsa + south + age + nearc + daded + momed", data)
exper.hat <- fitted.values(reg.lm)
reg.lm <- lm("exper^2 ~ smsa + south + age^2 + nearc + daded + momed", data)
exper2.hat <- fitted.values(reg.lm)
reg.lm <- lm("logw ~ educ.hat + exper.hat + exper2.hat + smsa + south", data)
summary(reg.lm)
```

```
##
## Call:
## lm(formula = "logw ~ educ.hat + exper.hat + exper2.hat + smsa + south",
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.68926 -0.23988  0.01859  0.26789  1.48377
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.185380   0.699296   4.555 5.44e-06 ***
## educ.hat      0.110319   0.008918  12.371 < 2e-16 ***
## exper.hat     0.346499   0.154099   2.249  0.0246 *
## exper2.hat   -0.015480   0.007794  -1.986  0.0471 *
## smsa          0.105852   0.023638   4.478 7.81e-06 ***
## south        -0.127096   0.023992  -5.297 1.26e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3928 on 3004 degrees of freedom
## Multiple R-squared:  0.2178, Adjusted R-squared:  0.2165
## F-statistic: 167.3 on 5 and 3004 DF,  p-value: < 2.2e-16
```

Adjusting for education and experience the r-squared is much lower giving a better estimation of variability.

(f)

```
reg.lm <- ivreg("logw ~ educ + exper + exper^2 + smsa + south |
               age + smsa + south + age^2 + nearc + daded + momed", data=data)
summary(reg.lm, diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = "logw ~ educ + exper + exper^2 + smsa + south | \n
##       data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.74204 -0.23933  0.02796  0.24792  1.32271
##
## Coefficients:
```

```

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.560379   0.097311  46.864 < 2e-16 ***
## educ        0.098785   0.006662  14.827 < 2e-16 ***
## exper       0.040460   0.002249  17.987 < 2e-16 ***
## smsa        0.138215   0.016859   8.198 3.57e-16 ***
## south      -0.162593   0.015758 -10.318 < 2e-16 ***
##
## Diagnostic tests:
##              df1  df2 statistic p-value
## Weak instruments (educ)      4 3003   178.120 < 2e-16 ***
## Weak instruments (exper)     4 3003  1562.559 < 2e-16 ***
## Wu-Hausman                   1 3004    9.747 0.00181 **
## Sargan                       2  NA     4.434 0.10895
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3867 on 3005 degrees of freedom
## Multiple R-Squared: 0.2416, Adjusted R-squared: 0.2406
## Wald test: 214.7 on 4 and 3005 DF, p-value: < 2.2e-16

Instruments seems to be valid (are uncorrelated with the error term).

```