

Wooldridge Computer Exercise Ch16.C1

```
library(wooldridge)
data("smoke")
head(smoke)
```

```
##      educ cigpric white age income  cigs restaurn    lincome agesq lcigpric
## 1 16.0   60.506     1  46  20000     0         0  9.903487  2116 4.102743
## 2 16.0   57.883     1  40  30000     0         0 10.308952  1600 4.058424
## 3 12.0   57.664     1  58  30000     3         0 10.308952  3364 4.054633
## 4 13.5   57.883     1  30  20000     0         0  9.903487   900 4.058424
## 5 10.0   58.320     1  17  20000     0         0  9.903487   289 4.065945
## 6  6.0   59.340     1  86   6500     0         0  8.779557  7396 4.083283
```

i)

β_1 is the change in $\log(\text{income})$ for an increase of one cigarette a day.

ii)

We would expect a decrease in consumption for an increase in cigarette prices, therefore $\gamma_5 < 0$, and we would also expect cigarette consumption to decrease if there is a ban on smoking in restaurants, $\gamma_6 < 0$.

iii)

If $\gamma_5 \neq 0$ and $\gamma_6 \neq 0$.

iv)

```
olsinc <- lm(log(income)~cigs+educ+age+I(age^2), data=smoke)
summary(olsinc)
```

```
##
## Call:
## lm(formula = log(income) ~ cigs + educ + age + I(age^2), data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.6237 -0.2978 0.1314 0.4167 1.3542
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.795e+00  1.704e-01  45.741 < 2e-16 ***
## cigs         1.731e-03  1.714e-03   1.010  0.313
## educ         6.036e-02  7.898e-03   7.642 6.10e-14 ***
## age          5.769e-02  7.644e-03   7.548 1.21e-13 ***
## I(age^2)     -6.306e-04  8.338e-05  -7.563 1.08e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6529 on 802 degrees of freedom
## Multiple R-squared:  0.165, Adjusted R-squared:  0.1608
## F-statistic: 39.61 on 4 and 802 DF, p-value: < 2.2e-16
```

β_1 is slightly positive, which means that income slightly increases with an increase in smoking, but it is also not statistically significant.

v)

```
reducedform <- lm(cigs ~ educ+age+I(age^2)+log(cigpric)+restaurn, data=smoke)
summary(reducedform)
```

```
##
## Call:
## lm(formula = cigs ~ educ + age + I(age^2) + log(cigpric) + restaurn,
##     data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.078  -9.312  -6.246   8.038  70.329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.580127  23.695583   0.067  0.94685
## educ        -0.450147   0.161640  -2.785  0.00548 **
## age          0.822541   0.154322   5.330 1.28e-07 ***
## I(age^2)     -0.009590   0.001679  -5.711 1.58e-08 ***
## log(cigpric) -0.351320   5.765550  -0.061  0.95143
## restaurn     -2.736389   1.109693  -2.466  0.01388 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.41 on 801 degrees of freedom
```

```
## Multiple R-squared:  0.051, Adjusted R-squared:  0.04508
## F-statistic:  8.61 on 5 and 801 DF,  p-value: 5.859e-08
```

restaurn is significant, but $\log(\text{cigpric})$ is not. We can still estimate the regression, as we have one IV left.

vi)

```
smoke$fitcigs <- fitted(reducedform)
secondstage <- lm(log(income)~fitcigs+educ+age+I(age^2), data=smoke)
summary(secondstage)

##
## Call:
## lm(formula = log(income) ~ fitcigs + educ + age + I(age^2), data = smoke)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6170 -0.2832  0.1285  0.4057  1.3964
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.7808932  0.1701577  45.728  < 2e-16 ***
## fitcigs      -0.0421257  0.0194080  -2.171  0.03026 *
## educ         0.0396746  0.0120519   3.292  0.00104 **
## age          0.0938182  0.0176573   5.313 1.40e-07 ***
## I(age^2)     -0.0010508  0.0002031  -5.175 2.89e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6514 on 802 degrees of freedom
## Multiple R-squared:  0.1688, Adjusted R-squared:  0.1647
## F-statistic: 40.72 on 4 and 802 DF,  p-value: < 2.2e-16
```

We now get the expected sign for β_1 .

Using R functions

```
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```

## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
## Loading required package: sandwich
## Loading required package: survival
iv <- ivreg(log(income)~cigs+educ+age+I(age^2)|educ+age+I(age^2)+log(cigpric)+restaurn, data = smoke)
summary(iv,vcov = sandwich, diagnostics = TRUE)

##
## Call:
## ivreg(formula = log(income) ~ cigs + educ + age + I(age^2) |
##      educ + age + I(age^2) + log(cigpric) + restaurn, data = smoke)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -4.13055 -0.44952 -0.05524  0.52926  3.09278
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.780893   0.257970  30.162  < 2e-16 ***
## cigs         -0.042126   0.024197  -1.741   0.0821 .
## educ          0.039675   0.015457   2.567   0.0104 *
## age           0.093818   0.022497   4.170 3.38e-05 ***
## I(age^2)     -0.001051   0.000257  -4.088 4.78e-05 ***
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments    2 801      3.860 0.0215 *
## Wu-Hausman          1 801      5.520 0.0190 *
## Sargan              1  NA      1.356 0.2443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.88 on 802 degrees of freedom
## Multiple R-Squared:  -0.5169, Adjusted R-squared:  -0.5245
## Wald test: 21.88 on 4 and 802 DF, p-value: < 2.2e-16

#se slightly off, but acceptable.
#-----

```

#not really working

```
library(systemfit)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Please cite the 'systemfit' package as:
```

```
## Arne Henningsen and Jeff D. Hamann (2007). systemfit: A Package for Estimating System  
40. http://www.jstatsoft.org/v23/i04/.
```

```
##
```

```
## If you have questions, suggestions, or comments regarding the 'systemfit' package, pl  
Forge site:
```

```
## https://r-forge.r-project.org/projects/systemfit/
```

```
system <- systemfit(formula=list(log(income)~cigs+educ+age+I(age^2),cigs~log(income)  
inst=~log(cigpric)+restaurn,  
method="2SLS",  
data=smoke)
```

```
summary(system)
```

```
##
```

```
## systemfit results
```

```
## method: 2SLS
```

```
##
```

```
##           N   DF   SSR detRCov   OLS-R2 McElroy-R2
```

```
## system 1614 1602 296987 28396.9 -0.951766    -1.3822
```

```
##
```

```
##           N   DF   SSR   MSE   RMSE           R2      Adj R2
```

```
## eq1  807  802 100662 125.513 11.2033 -244.878845 -246.105173
```

```
## eq2  807  800 196325 245.406 15.6654  -0.293708  -0.303411
```

```
##
```

```
## The covariance matrix of the residuals
```

```
##           eq1      eq2
```

```
## eq1 125.5135  49.0393
```

```
## eq2  49.0393 245.4062
```

```
##
```

```
## The correlations of the residuals
```

```
##           eq1      eq2
```

```
## eq1 1.000000 0.279419
```

```
## eq2 0.279419 1.000000
```

```
##
```

```
##
```

```
## 2SLS estimates for 'eq1' (equation 1)
```

```
## Model Formula: log(income) ~ cigs + educ + age + I(age^2)
```

```
## Instruments: ~log(cigpric) + restaurn
```

```
##
```

```

##               Estimate   Std. Error t value Pr(>|t|)
## (Intercept) -1.82390e+01  3.57041e+06  -1e-05  1.00000
## cigs        -3.02246e-01  7.79634e+04   0e+00  1.00000
## educ        -2.48101e+00  3.49785e+05  -1e-05  0.99999
## age          3.20194e+00  1.91080e+05   2e-05  0.99999
## I(age^2)     -3.54495e-02  2.19669e+03  -2e-05  0.99999
##
## Residual standard error: 11.20328 on 802 degrees of freedom
## Number of observations: 807 Degrees of Freedom: 802
## SSR: 100661.811051 MSE: 125.51348 Root MSE: 11.20328
## Multiple R-Squared: -244.878845 Adjusted R-Squared: -246.105173
##
##
## 2SLS estimates for 'eq2' (equation 2)
## Model Formula: cigs ~ log(income) + educ + age + I(age^2) + log(cigpric) + restaurn
## Instruments: ~log(cigpric) + restaurn
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.9597446         NA      NA      NA
## log(income)    1.1356359         NA      NA      NA
## educ          -0.7684853         NA      NA      NA
## age            0.2825605         NA      NA      NA
## I(age^2)       -0.0089433         NA      NA      NA
## log(cigpric)   2.2793212         NA      NA      NA
## restaurn      -3.5719844         NA      NA      NA
##
## Residual standard error: 15.665445 on 800 degrees of freedom
## Number of observations: 807 Degrees of Freedom: 800
## SSR: 196324.925379 MSE: 245.406157 Root MSE: 15.665445
## Multiple R-Squared: -0.293708 Adjusted R-Squared: -0.303411

```