HW 3

2023-04-18

Question 12.22 From Hansen Econometrics

```
# Load packages
pacman::p_load(tidyverse, magrittr, here, haven, magrittr, AER, fixest, cragg, ivmodel)

# Load data
ajr_df = read_dta(here("/Users/johannaallen/Documents/Erik/EC 581",
                       "AJR2001/AJR2001.dta"))
```

Part a)

```
# Estimate 12.86 loggdp = b0 + b1risk
reg1 = ajr_df %>% lm(loggdp ~ risk,.) |>  summary()

# Estimate 12.87 risk = b0 + b1 logmortality
reg2 = ajr_df %>% lm(risk ~ logmort0,.)
# Get the fitted values
risk_fitted = reg2$fitted.values
reg2 = summary(reg2)

# Estimate 12.88 logGdp = b0 + risk with risk now endogenous
reg3 = lm(ajr_df$loggdp ~ risk_fitted) |>  summary()
```

Part b)

```
# Restimate the above with both homoskedastic and heteroskedastic standard errors. The fixest package a
reg1_het = ajr_df %>% feols(loggdp ~ risk,., vcov = "HC1")
reg2_het = ajr_df %>% feols(risk ~ logmort0,., vcov = "HC1")
reg3_het = ajr_df %>% feols(loggdp ~ 1 | risk ~ logmort0,., vcov = "HC1")

# Compare SE's for 12.86
reg1$coefficients[2,2]; reg1_het$se[2] # Homoskedastic se's are reported
```

```
## [1] 0.063
```

```
##  risk
## 0.051
```

```r
# 12.87
reg2$coefficients[2,2]; reg2_het$se[2] # Homoskedastic se's are reported
```

```
## [1] 0.13

## logmort0
##     0.15
```

```r
# 12.88
reg3$coefficients[2,2]; reg3_het$se[2] # Homoskedastic se's are reported
```

```
## [1] 0.13

## fit_risk
##     0.17
```

Part c)

```r
# Estimate 2SLS with indirect least squares
# First estimate effect of instrument on outcome
reg4 = ajr_df %>% lm(loggdp ~ logmort0,.) |> summary()

# Then divide that effect by effect of instrument on endogenous variable found in reg2
reg4$coefficients[2,1] / reg2$coefficients[2,1] # The same as part b
```

```
## [1] 0.93
```

Part d)

```r
# Estimate 2SLS with two stage approach
reg_2sls = ivreg(loggdp ~ risk | logmort0, data = ajr_df)
coef(reg_2sls)[2] # Same as before
```

```
## risk
## 0.93
```

Part e)

```r
# Get error terms from 2SLS regression and add them to data set
uhat2 = residuals(reg_2sls)

ajr_df = mutate(ajr_df, uhat2 = uhat2)

# Run control variable approach
reg_control = ajr_df %>% lm(loggdp ~ risk + logmort0 + uhat2,.)
coef(reg_control)[2]  #Same as before
```

```
## risk
## 0.93
```

Part f)

```r
# Add latitude and africa as regressors
(reg_africa = ajr_df %>% lm(loggdp ~ risk + latitude + africa,.))
```

```
##
## Call:
## lm(formula = loggdp ~ risk + latitude + africa, data = .)
##
## Coefficients:
## (Intercept)         risk     latitude       africa
##       5.652        0.377        1.382       -0.723
```

```r
 # Africa is significant at any alpha level, lattitude is only significant at the 10% level
```

Part g)

```r
(reg_africa_2sls = ivreg(loggdp ~ risk + africa + latitude | logmort0 + africa + latitude, data = ajr_di
```

```
##
## Call:
## ivreg(formula = loggdp ~ risk + africa + latitude | logmort0 +     africa + latitude, data = ajr_df)
##
## Coefficients:
## (Intercept)         risk       africa     latitude
##      2.9951       0.8000      -0.3479      -0.0553
```

```r
# Africa and latitude are not longer significant
```

Part h)

```r
# Add exponential of mortality to data
ajr_df = mutate(ajr_df, mortality0 = exp(logmort0))

# Run baseline regresssion with mortality
(reg_mortality = ivreg(loggdp ~ risk | mortality0, data = ajr_df))
```

```
##
## Call:
## ivreg(formula = loggdp ~ risk | mortality0, data = ajr_df)
##
## Coefficients:
## (Intercept)         risk
##       0.489        1.161
```

The authors prefered log mortality because it has an easier interpretation than this level of mortality.

Part i)

```r
# Add square root to data set
ajr_df = mutate(ajr_df, mortsquare = logmort0^2)

(reg_square = ivreg(loggdp ~ risk | mortality0 + mortsquare, data = ajr_df))
```

```
##
## Call:
## ivreg(formula = loggdp ~ risk | mortality0 + mortsquare, data = ajr_df)
##
## Coefficients:
## (Intercept)         risk
##       1.934        0.939
```

```r
 # The result is back to the same result as earlier for risk
```

Part j)

```r
stock_yogo_test(X =~ 1, D =~ risk, Z =~ mortality0 + mortsquare, B = 0.1,size_bias = "size", data = ajr_
```

```
## Results of Stock and Yogo test for weak instruments:
##
##      Null Hypothesis:             Instruments are weak
##      Alternative Hypothesis:      Instruments are not weak
##
##      Data:                        ajr_df
##      Controls:                    ~1
##      Treatments:                  ~risk
##      Instruments:                 ~mortality0 + mortsquare
##
##      Alpha:                             0.05
##      Acceptable level of bias:    10% Wald test distortion.
##      Critical Value:                      20
##
##      Cragg-Donald Statistic:              11
##      Df:                                  61
```

The instruments are weak

Part k)

```r
summary(reg_square, vcov = sandwich, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = loggdp ~ risk | mortality0 + mortsquare, data = ajr_df)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4275 -0.5506  0.0837  0.7034  1.7147
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.934      1.128    1.71    0.092 .
## risk           0.939      0.169    5.55  6.3e-07 ***
##
## Diagnostic tests:
##                  df1 df2 statistic p-value
## Weak instruments   2  61      8.58 0.00052 ***
## Wu-Hausman         1  61     18.61   6e-05 ***
## Sargan             1  NA      0.63 0.42624
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.96 on 62 degrees of freedom
## Multiple R-Squared: 0.173,   Adjusted R-squared: 0.159
## Wald test: 30.8 on 1 and 62 DF,  p-value: 6.34e-07
```

The Wu-Hausaman test is significant so we conclude the instruments are exogenous.

Part j)

```
# Create IVmodel object
ivmod = ivmodel(Y = as.numeric(ajr_df$loggdp), D = as.numeric(ajr_df$risk), Z = ajr_df[, c("mortality0"

liml = LIML(ivmod)

liml$point.est
```

```
##      Estimate
## [1,]     0.96
```