```
# Instrumental Variables in R
# Copyright 2013 by Ani Katchova

# install.packages("AER")
library(AER)
# install.packages("systemfit")
library(systemfit)

mydata <- read.csv("C:/Econometrics/Data/iv_health.csv")
attach(mydata)

# Defining variables (Y1 dependent variable, Y2 endogenous variable)
# (X1 exogenous variables, X2 instruments, X2 instruments, overidentified case)
Y1 <- cbind(logmedexpense)
Y2 <- cbind(healthinsu)
X1 <- cbind(illnesses, age, logincome)
X2 <- cbind(ssiratio)
X2alt <- cbind(ssiratio, firmlocation)

# Descriptive statistics
summary(Y1)
summary(Y2)
summary(X1)
summary(X2)

# OLS regression
olsreg <- lm(Y1 ~ Y2 + X1)
summary(olsreg)

# 2SLS estimation
ivreg <- ivreg(Y1 ~ Y2 + X1 | X1 + X2)
summary(ivreg)

# 2SLS estimation (details)
olsreg1 <- lm (Y2 ~ X1 + X2)
summary(olsreg1)
Y2hat <- fitted(olsreg1)

olsreg2 <- lm(Y1 ~ Y2hat + X1)
summary(olsreg2)

# 2SLS estimation, over-identified case
ivreg_o <- ivreg(Y1 ~ Y2 + X1 | X1 + X2alt)
summary(ivreg_o)

# Hausman test for endogeneity of regressors
cf_diff <- coef(ivreg) - coef(olsreg)
vc_diff <- vcov(ivreg) - vcov(olsreg)
x2_diff <- as.vector(t(cf_diff) %*% solve(vc_diff) %*% cf_diff)
pchisq(x2_diff, df = 2, lower.tail = FALSE)

# Systems of equations

# Defining equations for systems of equations (2SLS and 3SLS)
# (X12 exogenous variable for eq2, X22 instrument for eq2)
```

```
X12 <- cbind(illnesses)
X22 <- cbind(firmlocation)
eq1 <- Y1 ~ Y2 + X1 + X2
eq2 <- Y2 ~ Y1 + X12 + X22
inst <- ~ X1 + X2 + X22
system <- list(eq1 = eq1, eq2 = eq2)

# 2SLS estimation
reg2sls <- systemfit(system, "2SLS", inst = inst, data = mydata)
summary(reg2sls)

# 3SLS estimation
reg3sls <- systemfit(system, "3SLS", inst = inst, data = mydata)
summary(reg3sls)
```

```
> # Instrumental Variables in R
> # Copyright 2013 by Ani Katchova
>
> # install.packages("AER")
> library(AER)
Loading required package: car
Loading required package: MASS
Loading required package: nnet
Loading required package: Formula
Loading required package: lmtest
Loading required package: zoo

Attaching package: ‘zoo’

The following object(s) are masked from ‘package:base’:

    as.Date, as.Date.numeric

Loading required package: sandwich
Loading required package: strucchange
Loading required package: survival
Loading required package: splines
Warning message:
package ‘AER’ was built under R version 2.15.3
> # install.packages("systemfit")
> library(systemfit)
Loading required package: Matrix
Loading required package: lattice
Warning message:
package ‘systemfit’ was built under R version 2.15.3
>
> mydata <- read.csv("C:/Econometrics/Data/iv_health.csv")
> attach(mydata)
>
> # Defining variables (Y1 dependent variable, Y2 endogenous variable)
> # (X1 exogenous variables, X2 instruments, X2 instruments, overidentified case)
> Y1 <- cbind(logmedexpense)
> Y2 <- cbind(healthinsu)
> X1 <- cbind(illnesses, age, logincome)
> X2 <- cbind(ssiratio)
> X2alt <- cbind(ssiratio, firmlocation)
>
> # Descriptive statistics
> summary(Y1)
 logmedexpense
 Min.   : 0.000
 1st Qu.: 5.740
 Median : 6.678
 Mean   : 6.481
 3rd Qu.: 7.430
 Max.   :10.180
> summary(Y2)
   healthinsu
 Min.   :0.0000
 1st Qu.:0.0000
```

```
 Median :0.0000
 Mean   :0.3822
 3rd Qu.:1.0000
 Max.   :1.0000
> summary(X1)
   illnesses          age          logincome
 Min.   :0.000   Min.   :65.00   Min.   :-6.908
 1st Qu.:1.000   1st Qu.:70.00   1st Qu.: 2.233
 Median :2.000   Median :74.00   Median : 2.743
 Mean   :1.861   Mean   :75.05   Mean   : 2.743
 3rd Qu.:3.000   3rd Qu.:80.00   3rd Qu.: 3.315
 Max.   :9.000   Max.   :91.00   Max.   : 5.744
> summary(X2)
    ssiratio
 Min.   :0.0000
 1st Qu.:0.2381
 Median :0.5045
 Mean   :0.5365
 3rd Qu.:0.9091
 Max.   :9.2506
>
> # OLS regression
> olsreg <- lm(Y1 ~ Y2 + X1)
> summary(olsreg)

Call:
lm(formula = Y1 ~ Y2 + X1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.2793 -0.6768  0.1472  0.8517  3.7803

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.780127   0.150891  38.307  < 2e-16 ***
Y2           0.074960   0.026012   2.882  0.00396 **
X1illnesses  0.440653   0.009572  46.035  < 2e-16 ***
X1age       -0.002595   0.001879  -1.381  0.16735
X1logincome  0.017236   0.013787   1.250  0.21124
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.237 on 10084 degrees of freedom
Multiple R-squared: 0.1749,      Adjusted R-squared: 0.1746
F-statistic: 534.4 on 4 and 10084 DF,  p-value: < 2.2e-16

>
> # 2SLS estimation
> ivreg <- ivreg(Y1 ~ Y2 + X1 | X1 + X2)
> summary(ivreg)

Call:
ivreg(formula = Y1 ~ Y2 + X1 | X1 + X2)

Residuals:
```

```
    Min      1Q  Median      3Q     Max
-6.7141 -0.7468  0.1288  0.8907  4.0895

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.589839   0.234676  28.081  < 2e-16 ***
Y2          -0.852201   0.198386  -4.296 1.76e-05 ***
X1illnesses  0.448512   0.010293  43.575  < 2e-16 ***
X1age       -0.011797   0.002789  -4.230 2.36e-05 ***
X1logincome  0.097693   0.022464   4.349 1.38e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.313 on 10084 degrees of freedom
Multiple R-Squared: 0.07094,     Adjusted R-squared: 0.07058
Wald test: 477.3 on 4 and 10084 DF,  p-value: < 2.2e-16

>
> # 2SLS estimation (details)
> olsreg1 <- lm (Y2 ~ X1 + X2)
> summary(olsreg1)

Call:
lm(formula = Y2 ~ X1 + X2)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6817 -0.3882 -0.2413  0.5167  2.5921

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.9591576  0.0568776  16.864  < 2e-16 ***
X1illnesses  0.0113510  0.0036336   3.124  0.00179 **
X1age       -0.0085302  0.0007125 -11.973  < 2e-16 ***
X1logincome  0.0544246  0.0056429   9.645  < 2e-16 ***
X2          -0.1997539  0.0141579 -14.109  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4691 on 10084 degrees of freedom
Multiple R-squared: 0.06839,     Adjusted R-squared: 0.06803
F-statistic: 185.1 on 4 and 10084 DF,  p-value: < 2.2e-16

> Y2hat <- fitted(olsreg1)
>
> olsreg2 <- lm(Y1 ~ Y2hat + X1)
> summary(olsreg2)

Call:
lm(formula = Y1 ~ Y2hat + X1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.2923 -0.6683  0.1525  0.8507  3.6881
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.589839   0.221021  29.815  < 2e-16 ***
Y2hat       -0.852201   0.186843  -4.561 5.15e-06 ***
X1illnesses  0.448512   0.009694  46.267  < 2e-16 ***
X1age       -0.011797   0.002627  -4.492 7.15e-06 ***
X1logincome  0.097693   0.021157   4.617 3.93e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.237 on 10084 degrees of freedom
Multiple R-squared: 0.1759,      Adjusted R-squared: 0.1756
F-statistic: 538.1 on 4 and 10084 DF,  p-value: < 2.2e-16

>
> # 2SLS estimation, over-identified case
> ivreg_o <- ivreg(Y1 ~ Y2 + X1 | X1 + X2alt)
> summary(ivreg_o)

Call:
ivreg(formula = Y1 ~ Y2 + X1 | X1 + X2alt)

Residuals:
    Min      1Q  Median      3Q     Max
-6.7692 -0.7664  0.1183  0.9073  4.1775

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.692387   0.228705  29.262  < 2e-16 ***
Y2          -0.969624   0.186385  -5.202 2.01e-07 ***
X1illnesses  0.449508   0.010427  43.111  < 2e-16 ***
X1age       -0.012963   0.002728  -4.752 2.04e-06 ***
X1logincome  0.107882   0.021821   4.944 7.78e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.333 on 10084 degrees of freedom
Multiple R-Squared: 0.04295,     Adjusted R-squared: 0.04257
Wald test: 465.7 on 4 and 10084 DF,  p-value: < 2.2e-16

>
> # Hausman test for endogeneity of regressors
> cf_diff <- coef(ivreg) - coef(olsreg)
> vc_diff <- vcov(ivreg) - vcov(olsreg)
> x2_diff <- as.vector(t(cf_diff) %*% solve(vc_diff) %*% cf_diff)
> pchisq(x2_diff, df = 2, lower.tail = FALSE)
[1] 1.493317e-05
>
> # Systems of equations
>
> # Defining equations for systems of equations (2SLS and 3SLS)
> # (X12 exogenous variable for eq2, X22 instrument for eq2)
> X12 <- cbind(illnesses)
> X22 <- cbind(firmlocation)
> eq1 <- Y1 ~ Y2 + X1 + X2
```

```
> eq2 <- Y2 ~ Y1 + X12 + X22
> inst <- ~ X1 + X2 + X22
> system <- list(eq1 = eq1, eq2 = eq2)
>
> # 2SLS estimation
> reg2sls <- systemfit(system, "2SLS", inst = inst, data = mydata)
> summary(reg2sls)

systemfit results
method: 2SLS

           N    DF     SSR detRCov   OLS-R2 McElroy-R2
system 20178 20168 25414.9 0.67827 -0.20465  -0.243292

        N    DF      SSR     MSE     RMSE        R2   Adj R2
eq1 10089 10083 22323.18 2.21394 1.487932 -0.192789 -0.19338
eq2 10089 10085  3091.76 0.30657 0.553687 -0.297834 -0.29822

The covariance matrix of the residuals
         eq1      eq2
eq1 2.213942 0.021393
eq2 0.021393 0.306570

The correlations of the residuals
          eq1       eq2
eq1 1.0000000 0.0259672
eq2 0.0259672 1.0000000


2SLS estimates for 'eq1' (equation 1)
Model Formula: Y1 ~ Y2 + X1 + X2
Instruments: ~X1 + X2 + X22

              Estimate Std. Error  t value   Pr(>|t|)
(Intercept)  7.3766350  0.5575224 13.23110 < 2.22e-16 ***
Y2          -1.6725004  0.5499930 -3.04095 0.00236440 **
X1illnesses  0.4578235  0.0131069 34.92997 < 2.22e-16 ***
X1age       -0.0187948  0.0052074 -3.60925 0.00030857 ***
X1logincome  0.1423374  0.0348757  4.08127 4.5135e-05 ***
X2          -0.1638580  0.1186859 -1.38060 0.16743185
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.487932 on 10083 degrees of freedom
Number of observations: 10089 Degrees of Freedom: 10083
SSR: 22323.176017 MSE: 2.213942 Root MSE: 1.487932
Multiple R-Squared: -0.192789 Adjusted R-Squared: -0.19338


2SLS estimates for 'eq2' (equation 2)
Model Formula: Y2 ~ Y1 + X12 + X22
Instruments: ~X1 + X2 + X22

              Estimate Std. Error  t value   Pr(>|t|)
(Intercept) -0.9720771  0.4658588 -2.08663  0.0369462 *
```

```
Y1           0.2348305  0.0820282  2.86280  0.0042078 **
X12         -0.0995706  0.0361380 -2.75529  0.0058746 **
X22          0.2828365  0.0269203 10.50644 < 2.22e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.553687 on 10085 degrees of freedom
Number of observations: 10089 Degrees of Freedom: 10085
SSR: 3091.755376 MSE: 0.30657 Root MSE: 0.553687
Multiple R-Squared: -0.297834 Adjusted R-Squared: -0.29822

>
> # 3SLS estimation
> reg3sls <- systemfit(system, "3SLS", inst = inst, data = mydata)
> summary(reg3sls)

systemfit results
method: 3SLS

           N    DF     SSR  detRCov    OLS-R2 McElroy-R2
system 20178 20168 24851.7 0.661561 -0.177955  -0.226796

        N    DF      SSR     MSE     RMSE        R2    Adj R2
eq1 10089 10083 21759.98 2.15809 1.469043 -0.162695 -0.163272
eq2 10089 10085  3091.76 0.30657 0.553687 -0.297834 -0.298220

The covariance matrix of the residuals used for estimation
         eq1      eq2
eq1 2.213942 0.021393
eq2 0.021393 0.306570

The covariance matrix of the residuals
           eq1        eq2
eq1 2.15808602 0.00657595
eq2 0.00657595 0.30656970

The correlations of the residuals
           eq1        eq2
eq1 1.00000000 0.00808462
eq2 0.00808462 1.00000000


3SLS estimates for 'eq1' (equation 1)
Model Formula: Y1 ~ Y2 + X1 + X2
Instruments: ~X1 + X2 + X22

               Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)  7.27396672  0.55748322 13.04787 < 2.22e-16 ***
Y2          -1.59903581  0.54998166 -2.90743 0.00365201 **
X1illnesses  0.45636542  0.01310671 34.81922 < 2.22e-16 ***
X1age       -0.01774624  0.00520695 -3.40818 0.00065653 ***
X1logincome  0.13598108  0.03487310  3.89931 9.7092e-05 ***
X2          -0.13395362  0.11867718 -1.12872 0.25904171
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 1.469043 on 10083 degrees of freedom
Number of observations: 10089 Degrees of Freedom: 10083
SSR: 21759.981368 MSE: 2.158086 Root MSE: 1.469043
Multiple R-Squared: -0.162695 Adjusted R-Squared: -0.163272


3SLS estimates for 'eq2' (equation 2)
Model Formula: Y2 ~ Y1 + X12 + X22
Instruments: ~X1 + X2 + X22

              Estimate Std. Error  t value   Pr(>|t|)
(Intercept) -0.9720771  0.4658588 -2.08663  0.0369462 *
Y1           0.2348305  0.0820282  2.86280  0.0042078 **
X12         -0.0995706  0.0361380 -2.75529  0.0058746 **
X22          0.2828365  0.0269203 10.50644 < 2.22e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.553687 on 10085 degrees of freedom
Number of observations: 10089 Degrees of Freedom: 10085
SSR: 3091.755374 MSE: 0.30657 Root MSE: 0.553687
Multiple R-Squared: -0.297834
```