Problem 2 - Simultaneous Equation Modelling (ILS, 2SLS, and 3SLS)

Taufiqur Rohman

2022-06-20

The problems are written in below.

Suppose the following model:

$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$

$Y_t = \alpha_0 + \alpha_1 R_t + u_{2t}$

Where M_t (money supply) is exogenous, R_t is the interest rate, and Y_t is GDP.

For now, let's upload the data set and clean it first. After that, let's take a look at the problem one by one.

```
path <- '~/Documents/RU/Econometrics II/06-HW-Jun/Exercise_2.xlsx'

colnames_df_exec2 <- c("year", "gdp_y", "m", "i", "r")
df_exec2 <- read_excel(path, skip = 1, col_names = colnames_df_exec2)

head(df_exec2)
```

```
## # A tibble: 6 x 5
##    year gdp_y     m     i     r
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1970 3578   626.  436.  6.56
## 2  1971 3698.  710.  486.  4.51
## 3  1972 3998. 3998.  543   4.47
## 4  1973 4123.  855.  606.  7.72
## 5  1974 4099   902.  562.  7.93
## 6  1975 4084. 1016.  462.  6.12
```

Looking from the data set, there's no problem in it. All of the data types are correct (dbl). We can move on now to the analysis.

Problem 1

1) How would you justify the model?

From the first equation, it shows that money supply and GDP level of a country will affect the interest rate. On the other hand, in the second equation, interest rate will also affect the GDP level of the country. An increase in real GDP (i.e., economic growth) will cause an increase in average interest rates in an economy.

2) Are the equations identified?

- The first equation doesn't have any excluded exogenous variable. But, it includes only one right-hand sided included endogenous variable Y_t, *thus* $k = 0 < g = 1$ *(under-identified)*
- The second equation has one excluded exogenous variable M_t and only one right hand side included endogenous variable R_t, *thus* $k = 1 = g$ *(just-identified)*

3) Using the data given in the table, estimate the parameters of the identified equation(s). Explain which of the method(s) you used in the estimation.

I will estimate the parameter using 2SLS. 2SLS is one of the most popular methods in solving SEM equation since it can be applied to all identified equations (ILS can only be used to a specific set of equations, particularly where the equation is just-identified) and costs less to compute than 3SLS.

In R, there are 2 famous packages that we can use to estimate the 2SLS. The first and the most preferred by many people is "ivreg" package. Let's try first using this package.

```
pr1_ivreg <- ivreg(gdp_y ~ r | m, data = df_exec2)
# Before vertical line: Defining the second stage regression
# After vertical line: Defining the instrumental variable (IV)

summary(pr1_ivreg)
```

```
##
## Call:
## ivreg(formula = gdp_y ~ r | m, data = df_exec2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4985.1 -2201.6  -130.9  1519.1  7868.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14296.8     3492.3   4.094 0.000327 ***
## r            -1244.5      505.1  -2.464 0.020149 *
##
## Diagnostic tests:
##                  df1 df2 statistic  p-value
## Weak instruments   1  28     6.601   0.0158 *
## Wu-Hausman         1  27    81.487 1.22e-09 ***
## Sargan             0  NA        NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2943 on 28 degrees of freedom
## Multiple R-Squared: -2.618,  Adjusted R-squared: -2.747
## Wald test: 6.071 on 1 and 28 DF,  p-value: 0.02015
```

The second one is using systemfit package. Even though not so many people are using this, it is useful for Simultaneous Equation Modelling in general. We can move from OLS, to 2SLS, even 3SLS by changing the hyper parameter "method".

```
pr1_D <- r ~ m + gdp_y
pr1_S <- gdp_y ~ r
pr1_sys <- list(pr1_D, pr1_S)
pr1_instr <- ~ m
```

```r
pr1.sys <- systemfit(formula = pr1_sys, inst = pr1_instr, method = "2SLS", data = df_exec2)
summary(pr1.sys)
```

```
##
## systemfit results
## method: 2SLS
##
##         N DF       SSR detRCov   OLS-R2 McElroy-R2
## system 60 55 242559747 9259394 -2.61814   0.814755
##
##      N DF        SSR         MSE       RMSE        R2    Adj R2
## eq1 30 27 2.2827e+02 8.45445e+00    2.90765 -0.282531 -0.377534
## eq2 30 28 2.4256e+08 8.66284e+06 2943.27029 -2.618150 -2.747370
##
## The covariance matrix of the residuals
##            eq1        eq2
## eq1    8.45445    7998.76
## eq2 7998.75774 8662839.97
##
## The correlations of the residuals
##          eq1      eq2
## eq1 1.000000 0.934652
## eq2 0.934652 1.000000
##
##
## 2SLS estimates for 'eq1' (equation 1)
## Model Formula: r ~ m + gdp_y
## Instruments: ~m
##
##                 Estimate   Std. Error t value Pr(>|t|)
## (Intercept)  1.59426e+01  2.85663e+07       0        1
## m            1.64032e-03  1.05186e+04       0        1
## gdp_y       -2.27865e-03  9.45931e+03       0        1
##
## Residual standard error: 2.907653 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 228.270063 MSE: 8.454447 Root MSE: 2.907653
## Multiple R-Squared: -0.282531 Adjusted R-Squared: -0.377534
##
##
## 2SLS estimates for 'eq2' (equation 2)
## Model Formula: gdp_y ~ r
## Instruments: ~m
##
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 14296.824   3492.345  4.09376 0.00032664 ***
## r           -1244.519    505.102 -2.46390 0.02014851 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2943.270285 on 28 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 28
## SSR: 242559519.127439 MSE: 8662839.968837 Root MSE: 2943.270285
```

```
## Multiple R-Squared: -2.61815 Adjusted R-Squared: -2.74737
```

$\hat{Y}_t = 14296.824 - 1244.519 R_t$

Problem 2

Suppose we change the model as follows:

$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + \beta_3 Y_{t-1} + u_{1t}$

$Y_t = \alpha_0 + \alpha_1 R_t + u_{2t}$

1) Find out if the system is identified.

- The first equation doesn't have any excluded exogenous variable. But, it includes only one right-hand sided included endogenous variable Y_t, *thus* $k = 0 < g = 1$ *(under-identified)*
- The second equation has two excluded exogenous variable; M_t and Y_{t-1}; and only one right hand side included endogenous variable R_t, *thus* $k = 2 > g = 1$ *(over-identified)*

2) Using the date given in the table, estimate the parameters of the identified equation(s). Explain which of the method(s) you used in the estimation.

First, we should create the lagged variable for GDP.

```
df_exec2["gdp_y_lag1"] <- Lag(df_exec2$gdp_y, +1)
head(df_exec2)
```

```
## # A tibble: 6 x 6
##    year gdp_y     m     i     r gdp_y_lag1
##   <dbl> <dbl> <dbl> <dbl> <dbl>      <dbl>
## 1  1970 3578   626.  436.  6.56         NA
## 2  1971 3698.  710.  486.  4.51       3578
## 3  1972 3998. 3998.  543   4.47       3698.
## 4  1973 4123.  855.  606.  7.72       3998.
## 5  1974 4099   902.  562.  7.93       4123.
## 6  1975 4084. 1016.  462.  6.12       4099
```

Again, I will use 2SLS for this equation. From now on, I will use the systemfit package to estimate the parameter.

```
pr2_D <- r ~ m + gdp_y + gdp_y_lag1
pr2_S <- gdp_y ~ r
pr2_sys <- list(pr2_D, pr2_S)
pr2_instr <- ~ m + gdp_y_lag1

pr2.sys <- systemfit(formula = pr2_sys, inst = pr2_instr, method = "2SLS", data = df_exec2)
summary(pr2.sys)
```

```
##
## systemfit results
## method: 2SLS
##
##         N DF       SSR detRCov   OLS-R2 McElroy-R2
```

```
## system 58 52 164952861 9664353 -1.66236   0.729941
##
##      N DF         SSR         MSE       RMSE        R2    Adj R2
## eq1 29 25 1.45795e+02 5.83180e+00    2.41491  0.180506  0.082167
## eq2 29 27 1.64953e+08 6.10936e+06 2471.71192 -1.662362 -1.760968
##
## The covariance matrix of the residuals
##           eq1        eq2
## eq1    5.8318    5095.51
## eq2 5095.5116 6109359.83
##
## The correlations of the residuals
##          eq1      eq2
## eq1 1.000000 0.853667
## eq2 0.853667 1.000000
##
##
## 2SLS estimates for 'eq1' (equation 1)
## Model Formula: r ~ m + gdp_y + gdp_y_lag1
## Instruments: ~m + gdp_y_lag1
##
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.558444896        NaN     NaN      NaN
## m           -0.000463047        NaN     NaN      NaN
## gdp_y       -0.009949803        NaN     NaN      NaN
## gdp_y_lag1   0.010000000        NaN     NaN      NaN
##
## Residual standard error: 2.414913 on 25 degrees of freedom
## Number of observations: 29 Degrees of Freedom: 25
## SSR: 145.795108 MSE: 5.831804 Root MSE: 2.414913
## Multiple R-Squared: 0.180506 Adjusted R-Squared: 0.082167
##
##
## 2SLS estimates for 'eq2' (equation 2)
## Model Formula: gdp_y ~ r
## Instruments: ~m + gdp_y_lag1
##
##               Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 12878.976   2755.637  4.67368 7.3209e-05 ***
## r           -1024.400    397.179 -2.57919   0.015671 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2471.711922 on 27 degrees of freedom
## Number of observations: 29 Degrees of Freedom: 27
## SSR: 164952715.339747 MSE: 6109359.827398 Root MSE: 2471.711922
## Multiple R-Squared: -1.662362 Adjusted R-Squared: -1.760968
```

$\hat{Y}_t = 12878.976 - 1024.400 R_t + u_{2t}$

Problem 3

We further change the model as follows:

$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$

$Y_t = \alpha_0 + \alpha_1 R_t + \alpha_2 I_t + u_{2t}$

We treat I_t (domestic investment) and M_t as exogenous.

1) Determine the identification of the system.

- The first equation has one excluded exogenous variable, I_t. Also, it includes one right-hand sided included endogenous variable Y_t, *thus* k *= 1 =* g *(just-identified)*
- The second equation has one excluded exogenous variable, M_t and only one right hand side included endogenous variable R_t, *thus* k*= 1 =* g *(just-identified)*

All the whole system is identified.

(2) Using the date given in the table, estimate the parameters of the identified equation(s). Explain which of the method(s) you used in the estimation.

Again, I will use 2SLS for this equation. I will use the systemfit package to estimate the parameter.

```
pr3_D <- r ~ m + gdp_y
pr3_S <- gdp_y ~ r + i
pr3_sys <- list(pr3_D, pr3_S)
pr3_instr <- ~ m + i

pr3.sys <- systemfit(formula = pr3_sys, inst = pr3_instr, method = "2SLS", data = df_exec2)
summary(pr3.sys)
```

```
##
## systemfit results
## method: 2SLS
##
##         N DF      SSR detRCov   OLS-R2 McElroy-R2
## system 60 54 19734474  897301 0.705631    0.92267
##
##      N DF         SSR         MSE      RMSE       R2   Adj R2
## eq1 30 27 1.43857e+02 5.32805e+00   2.30826 0.191740 0.131868
## eq2 30 27 1.97343e+07 7.30901e+05 854.92756 0.705632 0.683827
##
## The covariance matrix of the residuals
##            eq1       eq2
## eq1    5.32805   1731.18
## eq2 1731.17836 730901.13
##
## The correlations of the residuals
##         eq1     eq2
## eq1 1.00000 0.87726
## eq2 0.87726 1.00000
##
##
## 2SLS estimates for 'eq1' (equation 1)
## Model Formula: r ~ m + gdp_y
## Instruments: ~m + i
##
##                Estimate   Std. Error  t value  Pr(>|t|)
## (Intercept)  8.373463704  2.362211691  3.54476 0.0014556 **
```

```
## m           -0.001146765  0.000867188 -1.32240 0.1971296
## gdp_y        0.000227751  0.000711940  0.31990 0.7515049
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.308257 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 143.857421 MSE: 5.328053 Root MSE: 2.308257
## Multiple R-Squared: 0.19174 Adjusted R-Squared: 0.131868
##
##
## 2SLS estimates for 'eq2' (equation 2)
## Model Formula: gdp_y ~ r + i
## Instruments: ~m + i
##
##                Estimate  Std. Error  t value   Pr(>|t|)
## (Intercept) 4870.762604 2085.129969  2.33595   0.027168 *
## r           -334.406852  223.825198 -1.49405   0.146757
## i              3.782842    0.783412  4.82868 4.8268e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 854.927557 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 19734330.469561 MSE: 730901.128502 Root MSE: 854.927557
## Multiple R-Squared: 0.705632 Adjusted R-Squared: 0.683827
```

$\hat{R}_t = 8.373 - 0.001M_t + 0.0002Y_t$

$\hat{Y}_t = 4870.763 - 334.607R_t + 3.783I_t$

Problem 4

Again, we change the model as follows:

$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$

$Y_t = \alpha_0 + \alpha_1 R_t + \alpha_2 I_t + u_{2t}$

$I_t = \gamma_0 + \gamma_1 R_t + u_{3t}$

1) Find out which of the equations are identified.

- The first equation has one excluded exogenous variable, I_t. Also, it includes one right-hand sided included endogenous variable Y_t, *thus* $k = 1 = g$ *(just-identified)*
- The second equation has one excluded exogenous variable, M_t and two one right hand side included endogenous variables; R_t and I_t; *thus* $k = 1 < g = 2$ *(under-identified)*
- The third equation has two excluded exogenous variable; M_t and Y_t; and only one right hand side included endogenous variable R_t, *thus* $k = 1 = g$ *(over-identified)*

2) Using the date given in the table, estimate the parameters of the identified equation(s) and justify your method(s).

We can use 2SLS here. But, I will try to solve this model using 3SLS. 3SLS is a combination between 2SLS and SUR (Seemingly Unrelated Regression).

The part where we have endogenous variables both in left-hand and right-hand side of the equation is the 2SLS. After that, we try to account whether the error term of each equation are also correlated. That's where the SUR comes in.

I found in the 3 previous equations, the error term of each equation's are correlated between each others. As the data is tiny and it won't affect the cost of calculation that much, there's no harm in trying this method.

```r
pr4_D <- r ~ m + gdp_y
pr4_S <- gdp_y ~ r + i
pr4_A <- i ~ r
pr4_sys <- list(pr4_D, pr4_S, pr4_A)
pr4_instr <- ~ m + i

pr4.sys <- systemfit(formula = pr4_sys, inst = pr4_instr, method = "3SLS", data = df_exec2)
summary(pr4.sys)
```

```
##
## systemfit results
## method: 3SLS
##
##         N DF      SSR    detRCov   OLS-R2 McElroy-R2
## system 90 82 68969400 1833103551 0.012692    0.99425
##
##      N DF         SSR         MSE       RMSE        R2    Adj R2
## eq1 30 27 1.66443e+02 6.16456e+00    2.48285  0.064842 -0.004429
## eq2 30 27 6.10190e+07 2.25996e+06 1503.31696  0.089808  0.022386
## eq3 30 28 7.95026e+06 2.83938e+05  532.85829 -1.823077 -1.923901
##
## The covariance matrix of the residuals used for estimation
##            eq1       eq2       eq3
## eq1    5.32805   1731.18   1133.31
## eq2 1731.17836 730901.13 375829.37
## eq3 1133.31175 375829.37 283937.95
##
## The covariance matrix of the residuals
##            eq1        eq2       eq3
## eq1    6.16456    3674.78   1302.28
## eq2 3674.77676 2259961.87 758555.27
## eq3 1302.28039  758555.27 283937.95
##
## The correlations of the residuals
##          eq1      eq2      eq3
## eq1 1.000000 0.984531 0.984332
## eq2 0.984531 1.000000 0.946945
## eq3 0.984332 0.946945 1.000000
##
##
## 3SLS estimates for 'eq1' (equation 1)
## Model Formula: r ~ m + gdp_y
## Instruments: ~m + i
##
##                 Estimate   Std. Error  t value   Pr(>|t|)
## (Intercept) 12.607919321  1.142368467 11.03665 1.6422e-11 ***
## m            0.000251327  0.000534773  0.46997 0.64215084
```

```
## gdp_y       -0.001105048  0.000288682 -3.82791 0.00069653 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.482854 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 166.443205 MSE: 6.164563 Root MSE: 2.482854
## Multiple R-Squared: 0.064842 Adjusted R-Squared: -0.004429
##
##
## 3SLS estimates for 'eq2' (equation 2)
## Model Formula: gdp_y ~ r + i
## Instruments: ~m + i
##
##                Estimate  Std. Error  t value   Pr(>|t|)
## (Intercept) 7956.076994 1441.595088  5.51894 7.5810e-06 ***
## r           -621.695406  174.412427 -3.56451  0.0013833 **
## i              2.459210    0.442741  5.55451 6.8946e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1503.316956 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 61018970.511525 MSE: 2259961.870797 Root MSE: 1503.316956
## Multiple R-Squared: 0.089808 Adjusted R-Squared: 0.022386
##
##
## 3SLS estimates for 'eq3' (equation 3)
## Model Formula: i ~ r
## Instruments: ~m + i
##
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 2330.9457   627.3674  3.71544 0.00089626 ***
## r           -217.0456    90.7196 -2.39249 0.02368670 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 532.858286 on 28 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 28
## SSR: 7950262.675548 MSE: 283937.952698 Root MSE: 532.858286
## Multiple R-Squared: -1.823077 Adjusted R-Squared: -1.923901
```

Equation:

$\hat{R}_t = 12.608 + 0.0002M_t - 0.001Y_t$

$\hat{I}_t = 2330.946 - 217.046R_t +$