Jugal Marfatia

Instrumental Variable (PS 3)

Question 1

In [1]:
```
import pandas as pd

df = pd.read_excel('/Users/jugalmarfatia/Downloads/excelfiles/affairs.xls')

df = df.rename(columns=lambda x: x.strip())
df.head()
```

Out[1]:

	id	male	age	yrsmarr	kids	relig	educ	occup	ratemarr	naffairs	affair	vryhap	hapavg	avgmarr	unhap	vryrel	smerel
0	4	1	37.0	10.0	0	3	18	7	4	0	0	0	1	0	0	0	0
1	5	0	27.0	4.0	0	4	14	6	4	0	0	0	1	0	0	0	1
2	6	1	27.0	1.5	0	3	18	4	4	3	1	0	1	0	0	0	0
3	11	0	32.0	15.0	1	1	12	1	4	0	0	0	1	0	0	0	0
4	12	0	27.0	4.0	1	3	17	1	5	3	1	1	0	0	0	0	0

A. Regression

$Kids_i = \alpha_0 + \alpha_1 \; Age_i + \alpha_2 \; Yrsmarr_i + e_i$

In [2]:
```
import statsmodels.formula.api as sm

ols = sm.ols(formula="kids ~ age + yrsmarr", data=df).fit()

print(ols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   kids   R-squared:                       0.330
Model:                            OLS   Adj. R-squared:                  0.327
Method:                 Least Squares   F-statistic:                     147.0
Date:                Fri, 21 Sep 2018   Prob (F-statistic):           1.21e-52
Time:                        13:29:51   Log-Likelihood:                -254.32
No. Observations:                 601   AIC:                             514.6
Df Residuals:                     598   BIC:                             527.8
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.3990      0.063      6.364      0.000       0.276       0.522
age           -0.0029      0.003     -1.116      0.265      -0.008       0.002
yrsmarr        0.0502      0.004     11.626      0.000       0.042       0.059
==============================================================================
Omnibus:                       18.966   Durbin-Watson:                   2.076
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               15.057
Skew:                          -0.300   Prob(JB):                     0.000538
Kurtosis:                       2.508   Cond. No.                         145.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

B. Regression

First run two regressions:

$Kids_i = \alpha_0 + \alpha_1 \ Age_i + e_i$

$Age_i = \beta_0 + \beta_1 \ Yrsmarr_i + u_i$

Predict $\hat{e}_i$ and $\hat{u}_i$

Run regression:

$\hat{e}_i = \theta_0 + \theta_1 \ \hat{u}_i + v_i$

In [3]:
```
ols_y = sm.ols(formula="kids ~ age ", data=df).fit()
ols_x1 = sm.ols(formula="age ~ yrsmarr", data=df).fit()

df['kids_error'] = df['kids'] - ols_y.predict()
df['age_error'] = df['age'] - ols_x1.predict()

ols = sm.ols(formula="kids_error ~ age_error", data=df).fit()

print(ols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             kids_error   R-squared:                       0.111
Model:                            OLS   Adj. R-squared:                  0.110
Method:                 Least Squares   F-statistic:                     75.14
Date:                Fri, 21 Sep 2018   Prob (F-statistic):           4.10e-17
Time:                        13:29:52   Log-Likelihood:                -280.04
No. Observations:                 601   AIC:                             564.1
Df Residuals:                     599   BIC:                             572.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1.023e-16      0.016    6.5e-15      1.000      -0.031       0.031
age_error     -0.0234      0.003     -8.668      0.000      -0.029      -0.018
==============================================================================
Omnibus:                      138.122   Durbin-Watson:                   2.069
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               77.054
Skew:                          -0.740   Prob(JB):                     1.85e-17
Kurtosis:                       2.057   Cond. No.                         5.84
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

C. Regression

Run regressions:

$Kids_i = \omega_0 + \omega_1 \hat{u}_i + v_i$

In [4]:
```
ols = sm.ols(formula="kids ~ age_error", data=df).fit()

print(ols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   kids   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.8372
Date:                Fri, 21 Sep 2018   Prob (F-statistic):              0.361
Time:                        13:29:52   Log-Likelihood:                -374.04
No. Observations:                 601   AIC:                             752.1
Df Residuals:                     599   BIC:                             760.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.7155      0.018     38.838      0.000       0.679       0.752
age_error     -0.0029      0.003     -0.915      0.361      -0.009       0.003
==============================================================================
Omnibus:                      261.852   Durbin-Watson:                   2.049
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              120.379
Skew:                          -0.952   Prob(JB):                     7.25e-27
Kurtosis:                       1.914   Cond. No.                         5.84
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Yes they are the same because:

$u_i = Age_i - \beta_0 + \beta_1 \ Yrsmarr_i$

Therefore:

$Kids_i = \omega_0 + \omega_1 \ \hat{u}_i + v_i = \omega_0 + \omega_1(Age_i - \beta_0 + \beta_1 \ Yrsmarr_i) + v_i$

Implies: We get the same coeffecient estimates.

Question 2

IV estimate using GMM

In [5]:
```
from linearmodels.iv import IVGMM

mod = IVGMM.from_formula(formula="kids ~ 1 + [age ~  yrsmarr]", data=df).fit()

print(mod.summary)
```

```
                          IV-GMM Estimation Summary
==============================================================================
Dep. Variable:                   kids   R-squared:                      0.0789
Estimator:                     IV-GMM   Adj. R-squared:                 0.0774
No. Observations:                 601   F-statistic:                    225.97
Date:                Fri, Sep 21 2018   P-value (F-stat)                0.0000
Time:                        13:29:52   Distribution:                  chi2(1)
Cov. Estimator:                robust

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
Intercept     -0.4481     0.0834    -5.3700     0.0000     -0.6117     -0.2846
age            0.0358     0.0024     15.032     0.0000      0.0311      0.0405
==============================================================================

Endogenous: age
Instruments: yrsmarr
GMM Covariance
Debiased: False
Robust (Heteroskedastic)
```

IV estimate using 2SLS

In [6]:
```
ols = sm.ols(formula="age ~ yrsmarr", data=df).fit()
df['age_hat'] = ols.predict()

ols = sm.ols(formula="kids ~ age_hat", data=df).fit()

print(ols.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   kids   R-squared:                       0.328
Model:                            OLS   Adj. R-squared:                  0.327
Method:                 Least Squares   F-statistic:                     292.6
Date:                Fri, 21 Sep 2018   Prob (F-statistic):           1.04e-53
Time:                        13:29:52   Log-Likelihood:                -254.94
No. Observations:                 601   AIC:                             513.9
Df Residuals:                     599   BIC:                             522.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.4481      0.070     -6.431      0.000      -0.585      -0.311
age_hat        0.0358      0.002     17.105      0.000       0.032       0.040
==============================================================================
Omnibus:                       19.672   Durbin-Watson:                   2.083
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               15.594
Skew:                          -0.306   Prob(JB):                     0.000411
Kurtosis:                       2.502   Cond. No.                         154.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

IV using Coeffecient ratio i.e.

$\beta_{iv} = \frac{\beta_{yz}}{\beta_{xz}}$

In [7]:
```
ols = sm.ols(formula="age ~ yrsmarr", data=df).fit()
ols1 = sm.ols(formula="kids ~ yrsmarr", data=df).fit()

print("Beta_IV_beta_ratio: " +  str(ols1.params[1]/ols.params[1]))
```

```
Beta_IV_beta_ratio: 0.035816413472169316
```

IV using sample covariance ratio i.e.

$\beta_{iv} = \frac{scov(y,z)}{scov(x,z)}$

In [8]:
```
scov_zx = df[['age', 'yrsmarr']].cov()
scov_zy = df[['kids', 'yrsmarr']].cov()

print("Beta_IV_scov_ratio: " +  str(scov_zy.loc['kids', 'yrsmarr']/scov_zx.loc['age', 'yrsmarr']))
```

```
Beta_IV_scov_ratio: 0.03581641347216924
```

The results from above four methods show that indeed we get the same IV estimates using the different techinques.