4.2) Control Function Approach

Vitor Kamada

August 2019

Endogeneity

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + u_1$$

$$E(z'u_1) = 0$$

$$y_2 = z\pi_2 + v_2$$

$$E(z'v_2) = 0$$

Endogeneity: $Cov(u_1, v_2) \neq 0$

$$u_1 = \rho_1 v_2 + e_1$$

Control Function Approach

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + u_1$$

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + \rho_1 v_2 + e_1$$

$$y_2 = z\pi_2 + v_2$$

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + \rho_1 \hat{v}_2 + error$$

$$error_i = e_{i1} + \rho_1 z_i(\hat{\pi}_2 - \pi_2)$$

Card (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling

```
import pandas as pd
from linearmodels import IVGMM
```

```
file="https://github.com/VitorKamada/ECO7110/raw/master/Data/card.dta
df = pd.read_stata(file)
```

```
df['const'] = 1
Demo = ['const', 'black', 'exper', 'expersq',
        'south','smsa','smsa66']
Region = ['reg662', 'reg663', 'reg664', 'reg665',
          'reg666', 'reg667', 'reg668','reg669']
df['blackeduc'] = df['educ']*df['black']
df['blacknearc4'] = df['black']*df['nearc4']
```

```
# Reduce decimal points to 2
pd.options.display.float_format = '{:,.2f}'.format
```

```
print(df[['lwage','educ','nearc4'] +
         Demo].describe(percentiles=[]))
```

	lwage	educ	nearc4
count	3,010.00	3,010.00	3,010.00
mean	6.26	13.26	0.68
std	0.44	2.68	0.47
min	4.61	1.00	0.00
50%	6.29	13.00	1.00
max	7.78	18.00	1.00

expersq	south	smsa	smsa66
3,010.00	3,010.00	3,010.00	3,010.00
95.58	0.40	0.71	0.65
84.62	0.49	0.45	0.48
0.00	0.00	0.00	0.00
64.00	0.00	1.00	1.00
529.00	1.00	1.00	1.00

Card (1995): OLS

$$Log(wage) = \alpha_1 educ + \alpha_2 black \cdot educ + z_1\delta_1 + u_1$$

```
OLS = IVGMM(df.lwage, df[['educ']+['blackeduc']+Demo+Region],
            None, None).fit()
```

	Parameter	Std. Err.	T-stat	P-value
educ	0.0708	0.0039	18.070	0.0000
blackeduc	0.0179	0.0060	2.9607	0.0031
const	4.6846	0.0772	60.676	0.0000
black	-0.4191	0.0769	-5.4532	0.0000
exper	0.0822	0.0068	12.168	0.0000
expersq	-0.0021	0.0003	-6.6535	0.0000
south	-0.1442	0.0281	-5.1293	0.0000
smsa	0.1341	0.0192	6.9873	0.0000
smsa66	0.0250	0.0185	1.3480	0.1777
reg662	0.0989	0.0351	2.8151	0.0049

```
IV = IVGMM(df.lwage, df[Demo+Region], df[['educ','blackeduc']],
           df[['nearc4','blacknearc4']]).fit()
```

	Parameter	Std. Err.	T-stat	P-value
const	3.7346	0.9410	3.9688	0.0001
black	-0.2828	0.4998	-0.5658	0.5716
exper	0.1059	0.0249	4.2576	0.0000
expersq	-0.0022	0.0005	-4.5839	0.0000
south	-0.1425	0.0298	-4.7795	0.0000
smsa	0.1112	0.0310	3.5890	0.0003
smsa66	0.0180	0.0205	0.8775	0.3802
educ	0.1274	0.0560	2.2741	0.0230
blackeduc	0.0109	0.0398	0.2739	0.7842

```
First1 = IVGMM(df.educ, df[['nearc4']+['blacknearc4']+Demo+Region],
               None, None).fit()
df['Res1'] = First1.resids
```

	Parameter	Std. Err.	T-stat	P-value
nearc4	0.3192	0.0956	3.3391	0.0008
blacknearc4	0.0030	0.1695	0.0176	0.9860
const	16.639	0.2193	75.877	0.0000
black	-0.9375	0.1462	-6.4119	0.0000
exper	-0.4126	0.0321	-12.866	0.0000
expersq	0.0009	0.0017	0.5094	0.6105
south	-0.0517	0.1419	-0.3645	0.7155
smsa	0.4021	0.1107	3.6311	0.0003
smsa66	0.0254	0.1105	0.2303	0.8179
reg662	-0.0786	0.1854	-0.4242	0.6715

```
First2 = IVGMM(df.blackeduc, df[['nearc4']+['blacknearc4']+Demo+Region],
               None, None).fit()
df['Res2'] = First2.resids
```

	Parameter	Std. Err.	T-stat	P-value
nearc4	-0.0909	0.0327	-2.7798	0.0054
blacknearc4	0.8747	0.1806	4.8445	0.0000
const	0.2570	0.1398	1.8375	0.0661
black	11.550	0.1523	75.816	0.0000
exper	0.0533	0.0303	1.7622	0.0780
expersq	-0.0079	0.0017	-4.7835	0.0000
south	-0.2528	0.0959	-2.6367	0.0084
smsa	0.1953	0.0634	3.0780	0.0021
smsa66	0.0469	0.0683	0.6874	0.4918
reg662	-0.1564	0.0655	-2.3872	0.0170

Endogeneity Test

```
Exog = df[['educ']+['blackeduc']+['Res1']+['Res2']+Demo+Region]
CF = IVGMM(df.lwage, Exog, None, None).fit()
```

	Parameter	Std. Err.	T-stat	P-value
educ	0.1274	0.0536	2.3766	0.0175
blackeduc	0.0109	0.0380	0.2871	0.7740
Res1	-0.0568	0.0538	-1.0562	0.2909
Res2	0.0070	0.0387	0.1812	0.8562