3) Generalized Method of Moments (GMM)

Vitor Kamada

August 2019

Reference

Wooldridge (2010). **Econometric Analysis of Cross Section and Panel Data.** Ch 8

https://ebookcentral.proquest.com/lib/wayne/detail.action?docID=3339196&

Method of Moments

$$E[y-\mu] = \frac{1}{N}\sum_{i=1}^{N}(y_i-\mu) = 0$$

$$\hat{\mu}_{MM} = \frac{1}{N}\sum_{i=1}^{N} y_i = \overline{y}$$

$$E[u|\mathbf{x}] \rightarrow E[\mathbf{x}u] = 0$$

$$E[\mathbf{x}(y-\mathbf{x}'\beta)] = \frac{1}{N}\sum_{i=1}^{N} x_i(y_i - x_i'\beta) = 0$$

$$\hat{\beta}_{MM} = (\Sigma_i x_i x_i')^{-1}(\Sigma_i x_i y_i) = 0$$

Generalized Method of Moments (GMM)

$$Q_N(\beta) = \{\tfrac{1}{N}(y - X\beta)'Z\}W_N\{\tfrac{1}{N}Z'(y - X\beta)\}$$

$$\frac{\partial Q_n(\beta)}{\partial \beta} = -2[\tfrac{1}{N}X'Z]W_N[\tfrac{1}{N}Z'(y - X\beta)] = 0$$

$$\hat{\beta}_{GMM} = (X'ZW_NZ'X)^{-1}X'ZW_NZ'y$$

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y$$

$$\hat{\beta}_{2SLS} = \{X'Z(Z'Z)^{-1}Z'X\}^{-1}X'Z(Z'Z)^{-1}Z'y$$

Optimal GMM

$$\hat{\beta}_{OGMM} = (X'Z\hat{S}^{-1}Z'X)^{-1}X'Z\hat{S}^{-1}Z'y$$

$$\hat{S} = \frac{1}{N}\sum_{i=1}^{N}\hat{u}^2 z_i z_i' = \frac{Z'DZ}{N}$$

If $E[u_i^2|z_i] = \sigma^2$, then $\hat{S} = \frac{s^2 Z'Z}{N}$

Medical Expenditure Panel Survey (MEPS): Individuals over the age of 65 years

ldrugexp: the log of total out-of-pocket expenditures on prescribed medications

hi_empunion: indicator for whether the individual holds either employer or union-sponsored health insurance

totchr: # of chronic conditions

sociodemographic variables: age, female, blhisp, and linc

ssiratio: ratio of an individual's social security income to the individual's income from all sources

multlc: if the firm is a large operator with multiple locations

pip install linearmodels

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from linearmodels import IV2SLS, IVLIML, IVGMM
```

```python
file="https://github.com/VitorKamada/ECO7110/raw/master/Data/mus06data.dta"
data = pd.read_stata(file)
data.head()
```

	ssiratio	age	educyr	drugexp	private	female	hisp
0	0.149877	74	16	595	1	1	0
1	0.395856	73	8	1783	1	0	0
2	1.000000	80	12	176	0	1	0
3	0.206639	70	17	2437	1	0	0
4	0.537192	91	16	330	0	0	0

Summary Statistics

```
controls = ['totchr', 'female', 'age', 'linc','blhisp']
print(data[['ldrugexp','hi_empunion'] +
          controls].describe(percentiles=[]))
```

	ldrugexp	hi_empunion	linc	blhisp
count	10391.000000	10391.000000	10089.000000	10391.000000
mean	6.479664	0.379655	2.743271	0.170340
std	1.363393	0.485324	0.913144	0.375956
min	0.000000	0.000000	-6.907755	0.000000
50%	6.677083	0.000000	2.743160	0.000000
max	10.180172	1.000000	5.744476	1.000000

```python
instruments = ['ssiratio', 'lowincome', 'multlc', 'firmsz']
print(data[instruments].describe(percentiles=[]))
```

```
           ssiratio     lowincome        multlc        firmsz
count  10391.000000  10391.000000  10391.000000  10391.000000
mean       0.520628      0.185641      0.060341      0.136501
std        0.374589      0.388836      0.238128      2.138914
min       -2.100647      0.000000      0.000000      0.000000
50%        0.490139      0.000000      0.000000      0.000000
max        9.250620      1.000000      1.000000     50.000000
```

```python
data[['hi_empunion'] + instruments].corr()
```

	hi_empunion	ssiratio	lowincome	multlc	firmsz
hi_empunion	1.000000	-0.196284	-0.114425	0.119054	0.037173
ssiratio	-0.196284	1.000000	0.249820	-0.174394	-0.040481
lowincome	-0.114425	0.249820	1.000000	-0.060701	-0.007802
multlc	0.119054	-0.174394	-0.060701	1.000000	0.187394
firmsz	0.037173	-0.040481	-0.007802	0.187394	1.000000

```python
from statsmodels.api import OLS, add_constant
data['const'] = 1
controls = ['const'] + controls
```

```python
ivolsmod = IV2SLS(data.ldrugexp,
      data[['hi_empunion'] + controls], None, None)
res_ols = ivolsmod.fit()
print(res_ols)
```

```
              Parameter  Std. Err.     T-stat    P-value
--------------------------------------------------------
hi_empunion      0.0739     0.0260     2.8441     0.0045
const            5.8611     0.1570     37.320     0.0000
totchr           0.4404     0.0094     47.049     0.0000
female           0.0578     0.0254     2.2797     0.0226
age             -0.0035     0.0019    -1.8228     0.0683
linc             0.0105     0.0137     0.7646     0.4445
blhisp          -0.1513     0.0341    -4.4353     0.0000
```

2SLS

```python
ivmod = IV2SLS(data.ldrugexp, data[controls],
  data.hi_empunion, data[['ssiratio','multlc']])
res_2sls = ivmod.fit()
print(res_2sls.summary)
```

	Parameter	Std. Err.	T-stat	P-value
const	6.8752	0.2579	26.660	0.0000
totchr	0.4512	0.0103	43.769	0.0000
female	-0.0278	0.0322	-0.8653	0.3869
age	-0.0141	0.0029	-4.8753	0.0000
linc	0.0943	0.0219	4.3079	0.0000
blhisp	-0.2237	0.0396	-5.6514	0.0000
hi_empunion	-0.9899	0.2046	-4.8386	0.0000

OGMM

```python
ivmod = IVGMM(data.ldrugexp, data[controls],
  data.hi_empunion, data[['ssiratio','multlc']])
res_gmm = ivmod.fit()
print(res_gmm)
```

	Parameter	Std. Err.	T-stat	P-value
const	6.8778	0.2580	26.658	0.0000
totchr	0.4510	0.0103	43.738	0.0000
female	-0.0282	0.0322	-0.8752	0.3815
age	-0.0142	0.0029	-4.8773	0.0000
linc	0.0945	0.0219	4.3142	0.0000
blhisp	-0.2231	0.0396	-5.6344	0.0000
hi_empunion	-0.9933	0.2047	-4.8530	0.0000

```
print(res_gmm.first_stage)
```

const	0.9834
	(16.780)
totchr	0.0133
	(3.6234)
female	-0.0727
	(-7.5644)
age	-0.0081
	(-11.311)
linc	0.0444
	(6.7838)
blhisp	-0.0679
	(-5.5484)
ssiratio	-0.1823
	(-7.8326)
multlc	0.1209
	(5.8212)

```python
ivmod = IVLIML(data.ldrugexp, data[controls],
               data.hi_empunion, data[['ssiratio','multlc']])
res_liml = ivmod.fit(cov_type='robust')
from linearmodels.iv.results import compare
print(compare({'2SLS':res_2sls,'LIML': res_liml, 'GMM': res_gmm}))
```

	2SLS	GMM	LIML
const	6.8752	6.8778	6.8807
	(26.660)	(26.658)	(26.577)
totchr	0.4512	0.4510	0.4513
	(43.769)	(43.738)	(43.730)
female	-0.0278	-0.0282	-0.0283
	(-0.8653)	(-0.8752)	(-0.8776)
age	-0.0141	-0.0142	-0.0142
	(-4.8753)	(-4.8773)	(-4.8781)
linc	0.0943	0.0945	0.0947
	(4.3079)	(4.3142)	(4.3114)
blhisp	-0.2237	-0.2231	-0.2241
	(-5.6514)	(-5.6344)	(-5.6531)
hi_empunion	-0.9899	-0.9933	-0.9957
	(-4.8386)	(-4.8530)	(-4.8361)

Overidentified Test (OID), Hansen's Test, and Sargan's Test

$$Q(\hat{\beta}) = \{\frac{1}{N}(y - X\hat{\beta})'Z\}\hat{S}^{-1}\{\frac{1}{N}Z'(y - X\hat{\beta})\}$$

$$Z'(y - X\hat{\beta}) \simeq 0, \text{ so } Q(\hat{\beta}) \simeq 0$$

$$Q(\hat{\beta}) \overset{a}{\sim} \chi^2_r,$$

r is the $\#$ of overidentifying restrictions

$$H_0 : E\{Z'(y - X\beta)\} = 0$$

Rejection means that at least one of the instruments is not valid

Test of Overidentifying Restrictions

```
res_gmm.j_stat
```

$$H_0 : \text{Overidentifying Restriction is Valid}$$

```
H0: Expected moment conditions are equal to 0
Statistic: 1.0475
P-value: 0.3061
Distributed: chi2(1)
WaldTestStatistic, id: 0x7f1d6b251160
```

Four Available Instruments

```
ivmod = IVGMM(data.ldrugexp, data[controls],
        data.hi_empunion, data[instruments])
res_gmm_all = ivmod.fit()
res_gmm_all.j_stat
```

```
H0: Expected moment conditions are equal to 0
Statistic: 11.5903
P-value: 0.0089
Distributed: chi2(3)
WaldTestStatistic, id: 0x7f1d6af14710
```

```
print(res_gmm_all)
```

	Parameter	Std. Err.	T-stat	P-value
const	6.7126	0.2426	27.670	0.0000
totchr	0.4495	0.0100	44.738	0.0000
female	-0.0105	0.0307	-0.3406	0.7334
age	-0.0125	0.0027	-4.5364	0.0000
linc	0.0797	0.0203	3.9162	0.0001
blhisp	-0.2061	0.0383	-5.3828	0.0000
hi_empunion	-0.8124	0.1846	-4.3999	0.0000

print(res_gmm_all.first_stage)

totchr	0.0133
	(3.6494)
female	-0.0724
	(-7.5497)
age	-0.0080
	(-11.206)
linc	0.0410
	(6.3552)
blhisp	-0.0676
	(-5.5369)
ssiratio	-0.1690
	(-7.3289)
lowincome	-0.0637
	(-5.1947)
multlc	0.1151
	(5.4799)
firmsz	0.0037
	(1.9286)