14) Arellano-Bond Estimator

Vitor Kamada

July 2018

Dynamic Model

$$y_{it} = \gamma_1 y_{i,t-1} + \ldots + \gamma_p y_{i,t-p} + x_{it}'\beta + \alpha_i + \epsilon_{it}$$

$$Corr(y_{i,t-1}, \epsilon_{i,t-1}) \neq 0$$

$$Corr(y_{i,t-1}, \bar{\epsilon}_i) \neq 0$$

$$Corr(y_{i,t-1} - \bar{y}_i, \epsilon_{it} - \bar{\epsilon}_i) \neq 0$$

IV Estimation in the FD Model

$$\Delta y_{it} = \gamma_1 \Delta y_{i,t-1} + ... + \gamma_p \Delta y_{i,t-p} + \Delta x_{it}' \beta + \Delta \epsilon_{it}$$

$$Corr(y_{i,t-1} - y_{i,t-2}, \epsilon_{it} - \epsilon_{i,t-1}) \neq 0$$

$$Corr(\Delta \epsilon_{it}, \Delta y_{i,t-k}) = 0 \text{ for } k \geq 2$$

Anderson and Hsiao (1981):

$y_{i,t-2}$ as IV for $\Delta y_{i,t-1}$

Arellano and Bond (1991): Additional Lags as IV

$$\Delta y_{it} = \alpha + \gamma_1 \Delta y_{i,t-1} + \gamma_2 \Delta y_{i,t-2} + \Delta \epsilon_{it}$$

$$t = 4, 5, 6, 7$$

$t = 4$: two IVs (y_{i1} and y_{i2}) uncorrelated with $\Delta\epsilon_{i4}$

$t = 5$: three IVs (y_{i1}, y_{i2}, y_{i3}) uncorrelated with $\Delta\epsilon_{i5}$

Total IVs: 2+3+4+5=14

Panel Study of Income Dynamics 1976-82 (PSID)

Variable	Obs	Mean	Std. Dev.	Min	Max
exp	4,165	19.85378	10.96637	1	51
wks	4,165	46.81152	5.129098	5	52
occ	4,165	.5111645	.4999354	0	1
ind	4,165	.3954382	.4890033	0	1
south	4,165	.2902761	.4539442	0	1
smsa	4,165	.6537815	.475821	0	1
ms	4,165	.8144058	.3888256	0	1
fem	4,165	.112605	.3161473	0	1
union	4,165	.3639856	.4812023	0	1
ed	4,165	12.84538	2.787995	4	17
blk	4,165	.0722689	.2589637	0	1
lwage	4,165	6.676346	.4615122	4.60517	8.537
id	4,165	298	171.7821	1	595
t	4,165	4	2.00024	1	7

Arellano-Bond Estimation

xtabond lwage, lags(2) vce(robust)

```
Arellano-Bond dynamic panel-data estimation  Number of obs      =      2380
Group variable: id                           Number of groups   =       595
Time variable: t
                                             Obs per group:   min =       4
                                                              avg =       4
                                                              max =       4

Number of instruments =     15               Wald chi2(2)       =   1253.03
                                             Prob > chi2        =    0.0000
One-step results
                                  (Std. Err. adjusted for clustering on id)
```

lwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
L1.	.5707517	.0333941	17.09	0.000	.5053005	.6362029
L2.	.2675649	.0242641	11.03	0.000	.2200082	.3151216
_cons	1.203588	.164496	7.32	0.000	.8811814	1.525994

Two-Step GMM Estimation

xtabond lwage, lags(2) twostep vce(robust)

```
Number of instruments =     15              Wald chi2(2)       =    1974.40
                                            Prob > chi2        =     0.0000
Two-step results
                                 (Std. Err. adjusted for clustering on id)
```

lwage	Coef.	WC-Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
L1.	.6095931	.0330542	18.44	0.000	.544808	.6743782
L2.	.2708335	.0279226	9.70	0.000	.2161061	.3255608
_cons	.9182262	.1339978	6.85	0.000	.6555952	1.180857

xtabond lwage, lags(2) vce(robust) maxldep(2)

```
Number of instruments =      9                 Wald chi2(2)      =    1557.28
                                               Prob > chi2       =     0.0000
One-step results
                                   (Std. Err. adjusted for clustering on id)
```

lwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
L1.	.6570311	.0304257	21.59	0.000	.5973978	.7166643
L2.	.2113461	.0258261	8.18	0.000	.160728	.2619643
_cons	.9972351	.1475037	6.76	0.000	.7081331	1.286337

Type of Regressors

Predetermined or weakly exogenous regressors:

$$E(x_{it}\epsilon_{is}) \neq 0 \text{ for } s < t$$

$$E(x_{it}\epsilon_{is}) = 0 \text{ for } s \geq t$$

x_{it} is instrumented by $x_{i,t-1}, x_{i,t-2}, \ldots$

Contemporaneously endogenous regressors:

$$E(x_{it}\epsilon_{is}) \neq 0 \text{ for } s \leq t$$

$$E(x_{it}\epsilon_{is}) = 0 \text{ for } s > t$$

x_{it} is instrumented by $x_{i,t-2}, \ldots$

Baltagi and Khanti-Akom (1990)

wks: weeks worked

occ: dummy for blue-collar occupation

ind: industry

smsa: metropolitan statistical area

ms: marital status

ed: education

xtabond lwage occ south smsa ind, lags(2) maxldep(3) pre(wks,lag(1,2)) endogenous(ms,lag(0,2)) endogenous(union,lag(0,2)) twostep vce(robust) artests(3)

lwage	Coef.	WC-Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
L1.	.611753	.0373491	16.38	0.000	.5385501	.6849559
L2.	.2409058	.0319939	7.53	0.000	.1781989	.3036127
wks						
--.	-.0159751	.0082523	-1.94	0.053	-.0321493	.000199
L1.	.0039944	.0027425	1.46	0.145	-.0013807	.0093695
ms	.1859324	.144458	1.29	0.198	-.0972	.4690649
union	-.1531329	.1677842	-0.91	0.361	-.4819839	.1757181
occ	-.0357509	.0347705	-1.03	0.304	-.1038999	.032398
south	-.0250368	.2150806	-0.12	0.907	-.446587	.3965134
smsa	-.0848223	.0525243	-1.61	0.106	-.187768	.0181235
ind	.0227008	.0424207	0.54	0.593	-.0604422	.1058437
_cons	1.639999	.4981019	3.29	0.001	.6637377	2.616261

If ϵ_{it} are serially uncorrelated, then:

$$Cov(\epsilon_{it} - \epsilon_{i,t-1}, \epsilon_{i,t-1} - \epsilon_{i,t-2})$$

$$-Cov(\epsilon_{i,t-1}, \epsilon_{i,t-1}) \neq 0 \text{ but}$$

$$\Delta\epsilon_{it} \text{ not correlated } \Delta\epsilon_{i,t-k} \text{ for } k \geq 2$$

estat abond

Order	z	Prob > z
1	-4.5244	0.0000
2	-1.6041	0.1087
3	.35729	0.7209

H0: no autocorrelation

Test of Overidentifying Restrictions

quietly xtabond lwage occ south smsa ind, lags(2) maxldep(3) pre(wks,lag(1,2)) endogenous(ms,lag(0,2)) /// endogenous(union,lag(0,2)) twostep artests(3)

estat sargan

```
Sargan test of overidentifying restrictions
        H0: overidentifying restrictions are valid

        chi2(29)      =  39.87571
        Prob > chi2   =    0.0860
```

Wooldridge (2010): Dynamic Airfare Equation

$$t = 1997, 1998, 1999, \text{ and } 2000$$

$$lfare_{it} = \theta_t + \beta_1 lfare_{i,t-1} + \beta_2 concen_{it} + c_i + u_{it}$$

$$\Delta lfare_{it} = \eta_t + \rho \Delta lfare_{i,t-1} + \gamma \Delta concen_{it} + \Delta u_{it}$$

$$t = 1999 \text{ and } 2000$$

$$t = 1999: \ lfare_{i,1997}$$

$$t = 2000: \ lfare_{i,1998}, lfare_{i,1997}$$

reg d(lfare l.lfare concen) y99 y00, vce(cluster id)

D.lfare	Coef.	Robust Std. Err.	t	P>|t|
lfare				
LD.	-.1264673	.0267104	-4.73	0.000
concen				
D1.	.0762671	.0527226	1.45	0.148
y99	-.0473536	.0050308	-9.41	0.000
y00	0	(omitted)		
_cons	.0624434	.0032977	18.94	0.000

xtabond lfare concen, lags(1) noconsta diffva(y99 y00) vce(robust)

lfare	Coef.	Robust Std. Err.	z	P>|z|
lfare				
L1.	.3326355	.0633024	5.25	0.000
concen	.1519406	.057848	2.63	0.009
y99	.0051715	.0043914	1.18	0.239
y00	.0577598	.0038228	15.11	0.000