Econometrics strikes back: GMM and two-way fixed effects

StataCorp LLC

June 28, 2022

Context

- Growing interest in estimation and inference of average treatment effects on the treated
  - Inference: What standard errors should I use
  - Is the tool I am using the correct one to obtain the parameter of interest

Context

- Growing interest in estimation and inference of average treatment effects on the treated
  - Inference: What standard errors should I use
  - Is the tool I am using the correct one to obtain the parameter of interest
- Treatment effect heterogeneity
- Malign two-way fixed effects

The Plan

- Two-way fixed effects allows for desired heterogeneity
    - Known tool with desirable properties

The Plan

- Two-way fixed effects allows for desired heterogeneity
    - Known tool with desirable properties
- How the proposed estimators and standard errors can be obtained using GMM
- Illustrate how we can use `gmm` to fit two sets of estimators
    - Show some programming tips/tricks for `gmm`
    - Show some other programming tools in Stata
- Illustrate how the modeling, not the tool, is the problem

Basic Concepts: Econometric Theory

Framework: Common intervention period

- Notation based on Wooldridge (2021)
- Time: $1 \dots T$
- Intervention: $d \in \{0, 1\}$
- Intervention: at time q.
  - Pre-intervention $t = 1, \dots, q - 1$
  - Intervention $t = q, \dots, T$
- Potential outcome $y_t(d)$
  - $y_t(1)$ under the intervention
  - $y_t(0)$ without the intervention
- Treatment effect at time the $te_t = y_t(1) - y_t(0)$
- Average treatment effect on the treated at time t is $\tau_t \equiv E[y_t(1) - y_t(0)|d = 1]$

Framework: Common intervention period

- The outcome is $y_t = y_t(0) + d\,[y_t(1) - y_t(0)]$

Framework: Common intervention period

- The outcome is $y_t = y_t(0) + d\left[y_t(1) - y_t(0)\right]$

$$
\begin{aligned}
E\left(y_t|d\right) &= E\left[y_t(0)|d\right] + dE\left(te_t|d\right) \\
&= E\left[y_t(0)|d\right] + d\left[(1-d)E\left(te_t|d=0\right) + dE\left(te_t|d=1\right)\right] \\
&= E\left[y_t(0)|d\right] + d\tau_t
\end{aligned}
$$

Framework: Common intervention period $E\left[y_t(0)|d\right]$

- The potential outcome of not receiving treatment is

$$
\begin{aligned}
y_t(0) &= y_1(0) + (y_t(0) - y_1(0)) \\
&= y_1(0) + g_t(0)
\end{aligned}
$$

- Common trends assumption: $E\left[g_t(0)|d\right] = E\left[g_t(0)\right] \equiv \theta_t$

Framework: Common intervention period $E\left[y_t(0)|d\right]$

- The potential outcome of not receiving treatment is

$$\begin{aligned} y_t(0) &= y_1(0) + (y_t(0) - y_1(0)) \\ &= y_1(0) + g_t(0) \end{aligned}$$

- Common trends assumption: $E\left[g_t(0)|d\right] = E\left[g_t(0)\right] \equiv \theta_t$
- Because d is binary $E\left[y_1(0)|d\right] = \lambda + \zeta d$

Framework: Common intervention period $E\left[y_t(0)|d\right]$

- The potential outcome of not receiving treatment is

$$\begin{aligned} y_t(0) &= y_1(0) + (y_t(0) - y_1(0)) \\ &= y_1(0) + g_t(0) \end{aligned}$$

- Common trends assumption: $E\left[g_t(0)|d\right] = E\left[g_t(0)\right] \equiv \theta_t$
- Because d is binary $E\left[y_1(0)|d\right] = \lambda + \zeta d$
- Therefore $E\left[y_t(0)|d\right] = \lambda + \zeta d + \theta_t$

$$\begin{aligned} E\left(y_t|d\right) &= E\left[y_t(0)|d\right] + d\tau_t \\ &= \lambda + \theta_t + \zeta d + d\tau_t \end{aligned}$$

Framework: Common intervention period

- $E\left(y_t|d\right) = \lambda + \theta_t + \zeta d + d\tau_t$

Framework: Common intervention period

- $E\left(y_t|d\right) = \lambda + \theta_t + \zeta d + d\tau_t$
- We have an estimating equation within the potential outcomes framework
- We rely on common trends assumption for identification
- The estimating equation allows for time-varying treatment effects
- We can use our regression methods to estimate the parameters

Framework: Common intervention period

- $E\left(y_t|d\right) = \lambda + \theta_t + \zeta d + d\tau_t$
- We have an estimating equation within the potential outcomes framework
- We rely on common trends assumption for identification
- The estimating equation allows for time-varying treatment effects
- We can use our regression methods to estimate the parameters
- Following a similar argument we can make the effect change with covariates
- We can use `margins` or `gmm` to obtain the objects of interest

Framework: Staggered intervention

- Intervention occurs at different times $r \in \{q, q+1, \ldots T\}$
- Potential outcome $y_t(r)$ with never treated at $y_t(\infty)$
- $te_t(r) = y_t(r) - y_t(\infty)$ and $\tau_{rt} = E[te_t(r)|d_r = 1]$
- Common trends
  $E[y_t(\infty) - y_1(\infty)|d_q, \ldots, d_T] = E[y_t(\infty) - y_1(\infty)] \equiv \theta_t$

Framework: Staggered intervention

- $y_t = y_t(\infty) + d_q[y_t(q) - y_t(\infty)] + \ldots + d_T[y_t(T) - y_t(\infty)]$

Framework: Staggered intervention

- $y_t = y_t(\infty) + d_q[y_t(q) - y_t(\infty)] + \ldots + d_T[y_t(T) - y_t(\infty)]$

$$\begin{aligned} E\left(y_t|\mathbf{d}\right) &= E\left[y_t\left(\infty\right)|\mathbf{d}\right] + d_q E\left[te_t(q)|\mathbf{d}\right] + \ldots + d_T E\left[te_t(T)|\mathbf{d}\right] \\ &= E\left[y_t\left(\infty\right)|\mathbf{d}\right] + d_q E\left[te_t(q)|d_q = 1\right] + \ldots + \\ &\qquad d_T E\left[te_t(T)|d_T = 1\right] \end{aligned}$$

- Using common trends and $y_t\left(\infty\right) = y_1\left(\infty\right) + g_t\left(\infty\right)$ we have that:

$$\begin{aligned} E\left[y_t\left(\infty\right)\right] &= E\left[y_1\left(\infty\right)|\mathbf{d}\right] + E\left[g_t\left(\infty\right)|\mathbf{d}\right] \\ &= \eta + \lambda_q d_q + \ldots + \lambda_T d_T + \theta_t \end{aligned}$$

Framework: Staggered intervention

- $y_t = y_t(\infty) + d_q[y_t(q) - y_t(\infty)] + \ldots + d_T[y_t(T) - y_t(\infty)]$

$$\begin{aligned} E\left(y_t|\mathbf{d}\right) &= E\left[y_t\left(\infty\right)|\mathbf{d}\right] + d_q E\left[te_t(q)|\mathbf{d}\right] + \ldots + d_T E\left[te_t(T)|\mathbf{d}\right] \\ &= E\left[y_t\left(\infty\right)|\mathbf{d}\right] + d_q E\left[te_t(q)|d_q = 1\right] + \ldots + \\ &\qquad d_T E\left[te_t(T)|d_T = 1\right] \end{aligned}$$

- Using common trends and $y_t\left(\infty\right) = y_1\left(\infty\right) + g_t\left(\infty\right)$ we have that:

$$\begin{aligned} E\left[y_t\left(\infty\right)\right] &= E\left[y_1\left(\infty\right)|\mathbf{d}\right] + E\left[g_t\left(\infty\right)|\mathbf{d}\right] \\ &= \eta + \lambda_q d_q + \ldots + \lambda_T d_T + \theta_t \end{aligned}$$

- Our estimating equation can then be written as

$$E\left(y_t|\mathbf{d}\right) = \eta + \theta_t + \lambda_q d_q + \ldots + \lambda_T d_t + \tau_{qt} d_q + \ldots + \tau_{Tt} d_T$$

Framework: Staggered intervention

- Although treatment timing differs we reach analogous conclusions

Framework: Staggered intervention

- Although treatment timing differs we reach analogous conclusions
- Our potential outcome understanding holds
- Our concept of common trends as an indentifying assumption holds
- We can use regression tools to obtain the parameters of interest

Staggered intervention: Callaway and Sant'Anna

- Treatment effects are estimated for each treatment cohort at different points in time
- Reduce the problem to multiple two period problems
- Fits into potential outcome framework
- Similar identifying assumptions
- They propose three estimators: IPW, RA, and IPWRA

IPW

- Remember: $E\left[te_t(r)|d_r = 1\right] \equiv \tau_{rt}$
- The IPW estimator in Callaway and Sant'Anna is given by:

IPW

- Remember: $E\left[te_t(r)|d_r = 1\right] \equiv \tau_{rt}$
- The IPW estimator in Callaway and Sant'Anna is given by:

$$\tau_{rt} = E\left[\left(\frac{d_r}{E\left[d_r\right]} - \frac{\frac{p_r(X)d_\infty}{1-p_r(X)}}{E\left[\frac{p_r(X)d_\infty}{1-p_r(X)}\right]}\right)(Y_t - Y_{r-1})\right]$$

- $p_r\left(X\right)$ is an estimate of the probability of belonging to cohort r

RA

$$\tau_{rt} = E\left[\frac{d_r}{E[d_r]}\left(Y_t - Y_{r-1} - m_{rt}(X)\right)\right]$$
$$m_{rt}(X) = E\left[Y_t - Y_{r-1} | X, d_\infty = 1\right]$$

- $m_{rt}(X)$ is a regression using the never treated observations

IPWRA

$$\tau_{rt} = E\left[\left(\frac{d_r}{E\left[d_r\right]} - \frac{\frac{p_r(X)d_\infty}{1-p_r(X)}}{E\left[\frac{p_r(X)d_\infty}{1-p_r(X)}\right]}\right)\left(Y_t - Y_{r-1} - m_{rt}\left(X\right)\right)\right]$$

- Callaway and Sant'Anna in their implementation have that $p_r\left(.\right)$ and $m_{rt}\left(X\right)$ use the same covariates

What can we say

- Two-way fixed effects is an adequate tool, if we incorporate the heterogeneity we want to model
- Wooldridge (2021) and Callaway and Sant'Anna (2020) provide estimators that can be framed within GMM and fit using `gmm`
- Wooldridge (2021) and Callaway and Sant'Anna (2020) use methods different than GMM.

What can we say

- Two-way fixed effects is an adequate tool, if we incorporate the heterogeneity we want to model
- Wooldridge (2021) and Callaway and Sant'Anna (2020) provide estimators that can be framed within GMM and fit using `gmm`
- Wooldridge (2021) and Callaway and Sant'Anna (2020) use methods different than GMM.
- `gmm` gives equivalent point estimates but allows a wider array of standard errors
- `gmm` illustrates the costs of allowing for more heterogeneity (hidden in the Callaway and Sant'Anna framework)

Basic Concepts: `gmm` and `margins`

gmm

- gmm solves moment conditions of the form: $E\left[Z'e\left(X,\theta\right)\right]=0$
- $e\left(X,\theta\right)$ are residuals for regression and scores of probit or logit likelihoods.
- You specify moments using parenthesis before options and Z using the `instruments()` option
- You could specify gmm as a command or create a program (.ado)

margins

- `margins` uses and `expression` to obtain effects after estimation command
- Usually the `expression` is a command's default prediction
- Any function of the fitted model parameters is valid (`nlcom`, `lincom`)
- Effects could be population averaged effects or effects at a point

Linear regression and contrasts/marginal effects

```
. sysuse auto, clear
(1978 automobile data)
. regress mpg price i.foreign##c.length, vce(robust) noheader
```

mpg	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
price	-.0002262	.0001654	-1.37	0.176	-.0005561	.0001037
foreign						
Foreign	15.60087	14.27441	1.09	0.278	-12.87581	44.07754
length	-.1846372	.0257499	-7.17	0.000	-.2360068	-.1332677
foreign# c.length						
Foreign	-.0930218	.0804212	-1.16	0.251	-.2534577	.067414
_cons	57.41443	4.776039	12.02	0.000	47.8865	66.94237

```
. estimates store regress
```

Linear regression and contrasts/marginal effects

```
. margins, dydx(foreign) post vce(unconditional)
Average marginal effects                                    Number of obs = 74
Expression: Linear prediction, predict()
dy/dx wrt:  1.foreign
```

	dy/dx	Unconditional std. err.	t	P>|t|	[95% conf.	interval]
foreign						
Foreign	-1.880953	1.629301	-1.15	0.252	-5.131319	1.369413

```
Note: dy/dx for factor levels is the discrete change from the base level.
. estimates store dydx
```

Linear regression and contrasts/marginal effects

```
. estimates restore regress
(results regress are active now)
. margins r.foreign, post  vce(unconditional) contrast(nowald)
Contrasts of predictive margins                          Number of obs = 74
Expression: Linear prediction, predict()
```

	Contrast	Unconditional std. err.	[95% conf.	interval]
foreign (Foreign vs Domestic)	-1.880953	1.629301	-5.131319	1.369413

```
. estimates store contrast
```

Linear regression and contrasts/marginal effects

```
. estimates restore regress
(results regress are active now)
. margins, vce(unconditional) at(foreign=0) at(foreign=1)  ///
>          contrast(at(r) nowald) post
Contrasts of predictive margins                          Number of obs = 74
Expression: Linear prediction, predict()
1._at: foreign = 0
2._at: foreign = 1
```

	Contrast	Unconditional std. err.	[95% conf.	interval]
_at (2 vs 1)	-1.880953	1.629301	-5.131319	1.369413

```
. estimates store atcontrast
```

Linear regression and contrasts/marginal effects

```
. etable, estimates(dydx contrast atcontrast) column(estimates)
```

	dydx	contrast	atcontrast
Car origin			
Foreign	-1.881		
	(1.629)		
Car origin			
(Foreign vs Domestic)		-1.881	
		(1.629)	
_at			
(2 vs 1)			-1.881
			(1.629)
Number of observations	74		

Linear regression and contrasts/marginal effects

```
 . collect remap colname[1.foreign] = colname[r1vs0.foreign]
(13 items remapped in collection ETable)
. collect remap colname[r2vs1._at] = colname[r1vs0.foreign]
(8 items remapped in collection ETable)
. collect layout
Collection: ETable
      Rows: coleq#colname[]#result[_r_b _r_se] result[N]
   Columns: etable_estimates#stars[value]
   Table 1: 4 x 3
───────────────────────────────────────────────────
                          dydx  contrast atcontrast
───────────────────────────────────────────────────
Car origin
  (Foreign vs Domestic)  -1.881   -1.881     -1.881
                        (1.629)  (1.629)    (1.629)
Number of observations      74
───────────────────────────────────────────────────
```

Effects using gmm

```
. local xb  {b1}*price + {b2}*length + {b3}*1.foreign + {b4}*c.length#1.foreign
. local xb0 {b1}*price + {b2}*length + {b0}
. local xb1 `xb0´ + {b3} + {b4}*length
```

Effects using gmm

```
. gmm (mpg:   mpg  - (`xb´ + {b0}))                                 ///
>     (at0:  `xb0´ - {at0})                                         ///
>     (dydx: `xb1´ - {at0} - {dydx}),                               ///
>     instruments(mpg: price i.foreign##c.length)                   ///
>     winitial(unadjusted, independent) onestep iterlogonly
Iteration 0:   GMM criterion Q(b) =  475.49917
Iteration 1:   GMM criterion Q(b) =  2.076e-20
Iteration 2:   GMM criterion Q(b) =  3.729e-28
```

Effects using gmm

```
. gmm
GMM estimation
Number of parameters =   7
Number of moments    =   7
Initial weight matrix: Unadjusted                Number of obs   =         74
```

	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
/b1	-.0002262	.0001597	-1.42	0.157	-.0005392	.0000867
/b2	-.1846372	.0248647	-7.43	0.000	-.2333711	-.1359034
/b3	15.60087	13.78374	1.13	0.258	-11.41477	42.6165
/b4	-.0930218	.0776567	-1.20	0.231	-.2452262	.0591826
/b0	57.41443	4.611861	12.45	0.000	48.37535	66.45351
/at0	21.32035	.7128755	29.91	0.000	19.92314	22.71756
/dydx	-1.880953	1.573294	-1.20	0.232	-4.964552	1.202647

```
Instruments for equation mpg: price 0b.foreign 1.foreign length
    0b.foreign#co.length 1.foreign#c.length _cons
Instruments for equation at0: _cons
Instruments for equation dydx: _cons
```

`gmm` with a program evaluator

- You can also write an evaluator for `gmm`
- Flexibility vs. complexity
- What you would do if you were writing a routine

Evaluator

```
. *! version 1.0.0  25jun2022
. program _twfe_gmm_fr
  1.     version 17
  2.     syntax varlist if, at(name)                                   ///
>                      [                                               ///
>                      *                                               ///
>                      ]
  3.
.       tokenize `varlist´
  4. end
```

Evaluator: specifying equations and linear combinations

```
. *! version 1.0.0  25jun2022
. program _twfe_gmm_fr
  1.      version 17
  2.      syntax varlist if, at(name)                                              ///
>                       [                                                          ///
>                       y(string)                                                  ///
>                       *                                                          ///
>                       ]
  3.
.         tokenize `varlist´
  4.
.         tempvar breg bpom bdydx
  5.      tempname beta
  6.
.         local reg  `1´
  7.      local pom0 `2´
  8.      local dydx `3´
  9.
.         quietly matrix score double `breg´  = `at´ `if´, eq(#1)
 10.      quietly matrix score double `bpom´  = `at´ `if´, eq(#2)
 11.      quietly matrix score double `bdydx´ = `at´ `if´, eq(#3)
 12.
.         quietly replace `reg´  = `y´ - `breg´    `if´
 13.      quietly replace `pom0´ = `breg´ - `bpom´ `if´
 14.      quietly replace `dydx´ = `breg´ - `bpom´ - `bdydx´ `if´
 15. end
```

Tricking gmm to do at()

```
. quietly regress mpg price i.foreign##c.length, vce(robust) noheader
. matrix beta = e(b)
. _fv_term_info 0b.foreign 1.foreign, individuals fvrestripe matrix(beta)
. ret list
scalars:
              r(tsops) =  0
            r(k_terms) =  2
macros:
        r(individuals) : "price __000002 __000003 length"
            r(varlist) : "0b.foreign 1.foreign"
              r(type2) : "variable"
              r(type1) : "variable"
matrices:
              r(mean2) :  1 x 1
              r(mean1) :  1 x 1
```

Tricking gmm to do at()

```
. mat list beta
beta[1,7]
                                                          co.length#
          price     __000002     __000003       length  co.__000002
y1   -.00022623            0    15.600867   -.18463724            0
        c.length#
     c.__000003        _cons
y1   -.09302183    57.414433
```

Tricking gmm to do at()

```
. program _twfe_gmm_fr, sortpreserve
  1.          version 17
  2.          syntax varlist if, at(name)                                       ///
>                         [                                                   ///
>                         y(string)                                           ///
>                         atlist(string)                                      ///
>                         *                                                   ///
>                         ]
  3.          * OMMITING OUTPUT
.             quietly replace `reg´  = `y´ - `breg´    `if´
  4.          // Forming at() objects
.             local k: colsof `at´
  5.          matrix `beta´ = `at´[1, 1..`k´-2]
  6.          _fv_term_info `atlist´, individuals fvrestripe matrix(`beta´)
  7.          local f0: word 2 of `r(individuals)´
  8.          local f1: word 3 of `r(individuals)´
  9.          replace `f0´ = 1
 10.          replace `f1´ = 0
 11.          matrix score double `xb0´ = `beta´ `if´
 12.          quietly replace `pom0´ = `xb0´ - `bpom´ `if´
 13.          quietly replace `f0´ = 0
 14.          quietly replace `f1´ = 1
 15.          matrix score double `xb1´ = `beta´ `if´
 16.          quietly replace `dydx´ = `xb1´ - `bpom´ - `bdydx´ `if´
 17. end
```

What have we learned

- Two-way fixed effects is not broken
- Heterogeneous treatment effects fall into our potential outcome framework
- We can think of the problem as a set of estimating equations
- Getting effects can be done via `margins` or `gmm`

Stata Examples

Wooldridge (2021): Using `margins`

- Staggered treatment and heterogeneity in the covariates
- Key variables:
  - Define a cohort variable
  - Define an observation level indicator of treatment w_{it} (`w`)

Data

```
. describe
Contains data from staggered_6.dta
 Observations:         3,270
    Variables:             7                  24 Jun 2022 08:37
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
id              int     %9.0g                 cross-sectional identifier
year            int     %9.0g                 2011 to 2016
y               float   %9.0g                 outcome, levels
w               byte    %9.0g                 =1 if treated
x1              byte    %9.0g                 time constant control
x2              byte    %9.0g                 time constant control
logy            float   %9.0g                 outcome variable, natural log
--------------------------------------------------------------------------------
Sorted by: id
     Note: Dataset has changed since last saved.
```

Cohort

```
 . generate double cohort = 0
. bysort id: generate ttimes = year[_n] if w==1
(2,787 missing values generated)
. bysort id: egen cohort0 = min(ttimes)
(2,172 missing values generated)
. replace cohort = cohort0 if cohort0!=.
(1,098 real changes made)
. tab cohort
```

cohort	Freq.	Percent	Cum.
0	2,172	66.42	66.42
2014	804	24.59	91.01
2015	192	5.87	96.88
2016	102	3.12	100.00
Total	3,270	100.00	

Regressors

- Treatment indicator interacted with cohort and year
  - `i.w#2014bn.cohort#2014bn.year`
    `i.w#2014bn.cohort#2015bn.year ...`
    `i.w#2016bn.cohort#2016bn.year`
- Treatment indicator interacted with cohort and year and covariate `x1`
  - `i.w#2014bn.cohort#2015bn.year#c.x1 ...`
- Interaction and levels of covariates, cohort, and time
  - `(c.x1)##(2014bn.year 2015bn.year 2016bn.year i.cohort)`

regress

```
. qui reg logy i.w#2014bn.cohort#2014bn.year i.w#2014bn.cohort#2015bn.year  ///
>         i.w#2014bn.cohort#2016bn.year i.w#2015bn.cohort#2015bn.year      ///
>         i.w#2015bn.cohort#2016bn.year i.w#2016bn.cohort#2016bn.year      ///
>         i.w#2014bn.cohort#2014bn.year#c.x1                               ///
>         i.w#2014bn.cohort#2015bn.year#c.x1                               ///
>         i.w#2014bn.cohort#2016bn.year#c.x1                               ///
>         i.w#2015bn.cohort#2015bn.year#c.x1                               ///
>         i.w#2015bn.cohort#2016bn.year#c.x1                               ///
>         i.w#2016bn.cohort#2016bn.year#c.x1                               ///
>         (c.x1)##(2012bn.year 2013bn.year 2014bn.year                     ///
>         2015bn.year 2016bn.year i.cohort),                               ///
>         vce(cluster id)
```

margins to compute heterogeneous effects

```
. quietly generate over = cohort if  cohort!=0
. quietly margins 2014.year 2015.year 2016.year,      ///
>                 dydx(w) over(over) vce(unconditional)     ///
>                                noestimcheck post
. _coef_table
                                          (Std. err. adjusted for 545 clusters in id)
```

	Coefficient	Unconditional std. err.	t	P>|t|	[95% conf.	interval]
0.w	(base outcome)					
1.w						
over#year						
2014 2014	.1800395	.0224137	8.03	0.000	.1360115	.2240676
2014 2015	.1758216	.0229169	7.67	0.000	.1308052	.2208379
2014 2016	.1849706	.0251249	7.36	0.000	.135617	.2343243
2015 2014	0	(omitted)				
2015 2015	.0978163	.0414103	2.36	0.019	.0164725	.17916
2015 2016	.1327046	.0447888	2.96	0.003	.0447245	.2206847
2016 2014	0	(omitted)				
2016 2015	0	(omitted)				
2016 2016	.092621	.0654263	1.42	0.157	-.0358982	.2211401

gmm to compute heterogeneous effects

```
. gmm
Iteration 0:   EE criterion =  1.498e-24
Iteration 1:   EE criterion =  6.227e-32
Heterogeneous-treatment-effects regression             Number of obs = 3,270
Data type: Repeated cross-sectional
Estimator: Two-way fixed-effects
                                  (Std. err. adjusted for 545 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |       ATET   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
    __cohort#|
        year |
   2014 2014 |   .1800395   .0222936     8.08   0.000     .1363449    .2237342
   2014 2015 |   .1758216    .022794     7.71   0.000     .1311461    .2204971
   2014 2016 |   .1849706   .0249902     7.40   0.000     .1359907    .2339505
   2015 2015 |   .0978163   .0411884     2.37   0.018     .0170885    .1785441
   2015 2016 |   .1327046   .0445487     2.98   0.003     .0453907    .2200185
   2016 2016 |    .092621   .0650757     1.42   0.155    -.034925      .220167
-------------+----------------------------------------------------------------
OME1         |
    __cohort#|
        year |
   2014 2014 |   2.401441   .0765251    31.38   0.000     2.251454    2.551427
-------------+----------------------------------------------------------------
OME0         |
    __cohort#|
        year |
   2014 2014 |   2.221401   .0753911    29.47   0.000     2.073637    2.369165
-------------+----------------------------------------------------------------
OME1         |
    __cohort#|
        year |
   2014 2015 |   2.303018   .0743222    30.99   0.000     2.157349    2.448687
-------------+----------------------------------------------------------------
OME0         |
    __cohort#|
        year |
   2014 2015 |   2.127196   .0753975    28.21   0.000      1.97942    2.274973
-------------+----------------------------------------------------------------
OME1         |
    __cohort#|
        year |
```

Nonlinear models for heterogeneous effects (maybe)

```
. use did_staggered_6_corner, clear
. generate double cohort = 0
. bysort id: generate ttimes = year[_n] if w==1
(4,786 missing values generated)
. bysort id: egen cohort0 = min(ttimes)
(3,018 missing values generated)
. replace cohort = cohort0 if cohort0!=.
(2,982 real changes made)
```

Nonlinear models for heterogeneous effects (maybe)

```
. qui poisson y i.w#2004bn.cohort#2004bn.year i.w#2004bn.cohort#2005bn.year ///
>          i.w#2004bn.cohort#2006bn.year i.w#2005bn.cohort#2005bn.year      ///
>          i.w#2005bn.cohort#2006bn.year i.w#2006bn.cohort#2006bn.year      ///
>          i.w#2004bn.cohort#2004bn.year#c.x                                ///
>          i.w#2004bn.cohort#2005bn.year#c.x                                ///
>          i.w#2004bn.cohort#2006bn.year#c.x                                ///
>          i.w#2005bn.cohort#2005bn.year#c.x                                ///
>          i.w#2005bn.cohort#2006bn.year#c.x                                ///
>          i.w#2006bn.cohort#2006bn.year#c.x                                ///
>          (c.x)##(2002bn.year 2003bn.year 2004bn.year 2005bn.year          ///
>                  2006bn.year i.cohort),                                   ///
>          vce(cluster id)
```

Nonlinear models for heterogeneous effects (maybe)

```
. quietly generate over = cohort if  cohort!=0
. margins 2004bn.year 2005bn.year 2006bn.year,                           ///
>                 dydx(w) over(over) noestimcheck vce(unconditional)
note: 3018 observations omitted because of missing values in over() variable.
Average marginal effects                                Number of obs    = 6,000
                                                        Subpop. no. obs = 2,982

Expression: Predicted number of events, predict()
dy/dx wrt:  1.w
Over:       over
                                (Std. err. adjusted for 1,000 clusters in id)
------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
0.w          |  (base outcome)
-------------+----------------------------------------------------------------
1.w          |
   over#year |
  2004 2004  |   1.017501   1.033521     0.98   0.325    -1.008164    3.043166
  2004 2005  |    6.00713   2.162626     2.78   0.005      1.76846     10.2458
  2004 2006  |   4.569667   1.369919     3.34   0.001     1.884675    7.254658
  2005 2004  |          0  (omitted)
  2005 2005  |   7.170127   3.355386     2.14   0.033     .5936913    13.74656
  2005 2006  |   7.185492   2.781751     2.58   0.010      1.73336    12.63762
  2006 2004  |          0  (omitted)
  2006 2005  |          0  (omitted)
  2006 2006  |   13.73294   10.32555     1.33   0.184    -6.504777    33.97065
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Callaway and Santana `gmm`

- Obtain group and time cohorts
- Compute effects of interest for each group and time cohort
- Form the moment conditions
- Example for IPW. Remeber:

$$\tau_{rt} = E\left[\left(\frac{d_r}{E\left[d_r\right]} - \frac{\frac{p_r(X)d_\infty}{1-p_r(X)}}{E\left[\frac{p_r(X)d_\infty}{1-p_r(X)}\right]}\right)\left(Y_t - Y_{r-1}\right)\right]$$

Group and time cohorts

```
. // Group and time computation
. generate keep  = inlist(year, 2013, 2014) & inlist(cohort, 0, 2014)
. keep if keep
(2,278 observations deleted)
. // Depvar
. bysort id (year): generate double dy = logy[2] - logy[1]
. // Treatment variable
. generate double gt = cohort>0 if dy!=.
```

Getting estimates

```
. // Propensity score
. quietly logit gt x1
. predict double px if e(sample)
(option pr assumed; Pr(gt))
.
. // Normalizing means
. summarize gt if dy!=., meanonly
. local mgt = r(mean)
.
. // Propensity score weight
. generate double pxr = px*(1-gt)/(1-px)
. summarize pxr, meanonly
. local mpxr = r(mean)
```

Getting estimates

```
. // atet
. generate double atet = (gt/`mgt´ - pxr/`mpxr´)*dy
. sum atet
```

Variable	Obs	Mean	Std. dev.	Min	Max
atet	992	.1982376	.7244637	-2.467532	3.133791

Getting estimates

```
. csdid logy x1, ivar(id) time(year) gvar(cohort) method(stdipw)
.
Difference-in-difference with Multiple Time Periods
                                                       Number of obs = 992
Outcome model  : weighted mean
Treatment model: stabilized inverse probability
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
g2014        |
 t_2013_2014 |   .1982376   .0271813     7.29   0.000     .1449632     .251512
------------------------------------------------------------------------------
Control: Never Treated
See Callaway and Sant´Anna (2021) for details
```

Full gmm

```
. gmm
Iteration 0:   EE criterion =  2.308e-17
Iteration 1:   EE criterion =  9.533e-32
note: model is exactly identified.
GMM estimation
Number of parameters =  90
Number of moments    =  90
Initial weight matrix: Unadjusted                 Number of obs   =      3,270
                              (Std. err. adjusted for 545 clusters in id)
```

	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
atet1						
_cons	-.0068411	.0263192	-0.26	0.795	-.0584258	.0447435
treat1						
x1	-.0229986	.0472623	-0.49	0.627	-.115631	.0696338
_cons	-.7230127	.5663183	-1.28	0.202	-1.832976	.3869508
gmean1						
_cons	.2701613	.0199381	13.55	0.000	.2310833	.3092393
psmean1						
_cons	.2702069	.0199471	13.55	0.000	.2311112	.3093025
atet2						
_cons	-.0106981	.0280311	-0.38	0.703	-.0656381	.0442419
treat2						
x1	-.0229986	.0472623	-0.49	0.627	-.115631	.0696338
_cons	-.7230127	.5663183	-1.28	0.202	-1.832976	.3869508
gmean2						
_cons	.2701613	.0199381	13.55	0.000	.2310833	.3092393
psmean2						
_cons	.2702069	.0199471	13.55	0.000	.2311112	.3093025
atet3						
_cons	.1982376	.0271813	7.29	0.000	.1449632	.251512

(*output omitted*)

Conclusion

- Our usual tools `gmm` and `margins` help us understand heterogeneous treatment effects
- `gmm` works in all cases but ...
- Our usual estimators work fine (two-way fixed effects is not a broken toy)
- We looked at some Stata tools (`etable`, `collect`, ...)