Econometrics strikes back: GMM and two-way fixed effects

StataCorp LLC

October 12, 2022
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: $1 \ldots T$
- Intervention: $d \in \{0, 1\}$
- Intervention: at time $q$.
  - Pre-intervention $t = 1, \ldots, q - 1$
  - Intervention $t = q, \ldots, T$
- 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 $\tau_t \equiv E[y_t(1) - y_t(0)|d = 1]$
The outcome is $y_t = y_t(0) + d [y_t(1) - y_t(0)]$
Framework: Common intervention period

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

\[
E (y_t|d) = E [y_t(0)|d] + dE (te_t|d) \\
= E [y_t(0)|d] + d [(1 - d)E (te_t|d = 0) + dE (te_t|d = 1)] \\
= E [y_t(0)|d] + d\tau_t
\]
The potential outcome of not receiving treatment is

\[ y_t(0) = y_1(0) + (y_t(0) - y_1(0)) = y_1(0) + g_t(0) \]

Common trends assumption: \( E[g_t(0)|d] = E[g_t(0)] \equiv \theta_t \)
The potential outcome of not receiving treatment is

\[ y_t(0) = y_1(0) + (y_t(0) - y_1(0)) \]
\[ = y_1(0) + g_t(0) \]

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

Because \( d \) is binary \[ E[y_1(0)|d] = \lambda + \zeta d \]
The potential outcome of not receiving treatment is

\[ y_t(0) = y_1(0) + (y_t(0) - y_1(0)) \]
\[ = y_1(0) + g_t(0) \]

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

Because \( d \) is binary \( E[ y_1(0)|d] = \lambda + \zeta d \)

Therefore \( E[ y_t(0)|d] = \lambda + \zeta d + \theta_t \)

\[ E( y_t|d) = E[ y_t(0)|d] + d \tau_t \]
\[ = \lambda + \theta_t + \zeta d + d \tau_t \]
Framework: Common intervention period

\[ E(y_t | d) = \lambda + \theta_t + \zeta d + d\tau_t \]
Framework: Common intervention period

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

- \( E(y_t|d) = \lambda + \theta_t + \zeta d + d\tau_t \)
- 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 $r \in \{q, q+1, ..., T\}$
- 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

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

\[
E (y_t|d) = E [y_t (\infty)|d] + d_q E [te_t(q)|d] + \ldots + d_T E [te_t(T)|d] \\
= E [y_t (\infty)|d] + d_q E [te_t(q)|d_q = 1] + \ldots + d_T E [te_t(T)|d_T = 1]
\]

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

\[
E [y_t (\infty)|d] = E [y_1 (\infty)|d] + E [g_t (\infty)|d] \\
= \eta + \lambda_q d_q + \ldots + \lambda_T d_T + \theta_t
\]
Framework: Staggered intervention

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

\[
E(y_t|d) = E[y_t(\infty)|d] + d_q E[te_t(q)|d] + \ldots + d_T E[te_t(T)|d]
\]

\[
= E[y_t(\infty)|d] + d_q E[te_t(q)|d_q = 1] + \ldots +
\]

\[
d_T E[te_t(T)|d_T = 1]
\]

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

\[
E[y_t(\infty)|d] = E[y_1(\infty)|d] + E[g_t(\infty)|d]
\]

\[
= \eta + \lambda_q d_q + \ldots + \lambda_T d_T + \theta_t
\]

- Our estimating equation can then be written as

\[
E(y_t|d) = \eta + \theta_t + \lambda_q d_q + \ldots + \lambda_T d_T + \tau_{qt} d_q + \ldots + \tau_{Tt} d_T
\]
Framework: Staggered intervention

- Although treatment timing differs we reach analogous conclusions
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 AIPW
Remember: \( E[te_t(r)|d_r = 1] \equiv \tau_{rt} \)

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

The IPW estimator in Callaway and Sant’Anna is given by:

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

Equivalent to `teffects ipw` using \( Y_t - Y_{r-1} \) as the dependent variable

\( p_r(X) \) is an estimate of the probability of belonging to cohort \( r \)
\[
\tau_{rt} = E \left[ \frac{d_r}{E[d_r]} (Y_t - Y_{r-1} - m_{rt}(X)) \right]
\]

\[
m_{rt}(X) = E [Y_t - Y_{r-1}| X, d_{\infty} = 1]
\]

- Equivalent to `teffects ra` using \(Y_t - Y_{r-1}\) as the dependent variable
- \(m_{rt}(X)\) is a regression using the never treated observations
\[ \tau_{rt} = E \left[ \left( \frac{d_r}{E[d_r]} - \frac{p_r(X)d_{r\infty}}{1-p_r(X)} \right) (Y_t - Y_{r-1} - m_{rt}(X)) \right] \]

- Callaway and Sant’Anna in their implementation have that \( p_r(\cdot) \) and \( m_{rt}(X) \) 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
The `gmm` command solves moment conditions of the form: $E[Z'e(X, \theta)] = 0$.

- $e(X, \theta)$ 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
```

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

<table>
<thead>
<tr>
<th></th>
<th>Unconditional</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>foreign</td>
<td>dy/dx</td>
<td>std. err.</td>
<td>t</td>
<td>P&gt;</td>
</tr>
<tr>
<td>Foreign</td>
<td>-1.880953</td>
<td>1.629301</td>
<td>-1.15</td>
<td>0.252</td>
</tr>
</tbody>
</table>

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()

<table>
<thead>
<tr>
<th></th>
<th>Unconditional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrast</td>
</tr>
<tr>
<td>foreign</td>
<td>-1.880953 1.629301</td>
</tr>
<tr>
<td>(Foreign vs Domestic)</td>
<td></td>
</tr>
</tbody>
</table>

. 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

<table>
<thead>
<tr>
<th></th>
<th>Unconditional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrast</td>
</tr>
<tr>
<td>_at</td>
<td>(2 vs 1)</td>
</tr>
<tr>
<td></td>
<td>-1.880953</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>dydx</th>
<th>contrast atcontrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car origin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>-1.881</td>
<td>(1.629)</td>
</tr>
<tr>
<td>(Foreign vs Domestic)</td>
<td>-1.881</td>
<td>(1.629)</td>
</tr>
<tr>
<td>_at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2 vs 1)</td>
<td>-1.881</td>
<td>(1.629)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>74</td>
<td></td>
</tr>
</tbody>
</table>
```
### Linear regression and contrasts/marginal effects

```plaintext
. 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

<table>
<thead>
<tr>
<th>dydx</th>
<th>contrast</th>
<th>atcontrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car origin</td>
<td>(Foreign vs Domestic)</td>
<td>-1.881</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.629)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>74</td>
<td></td>
</tr>
</tbody>
</table>
```

Number of observations: 74
Effects using \texttt{gmm}

\begin{verbatim}
. 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
\end{verbatim}
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
. *! version 1.0.0 25jun2022
. program _twfe_gmm_fr
1.    version 17
2.    syntax varlist if, at(name) ///
>        [ ///
>        ///
>        ///
>        ///
>        ///
3.    tokenize `varlist´
4.    end
. *! version 1.0.0 25jun2022
. program _twfe_gmm_fr
 1.   version 17
 2.   syntax varlist if, at(name) ///
>       [ ///
>       y(string) ///
>       * ///
>       ] ///
 3.
 4.   tokenize `varlist´
 5.
 6.   tempvar breg bpom bdydx
 7.   tempname beta
 8.
 9.   local reg `1´
10.  local pom0 `2´
11.  local dydx `3´
12.
13. quietly matrix score double `breg´ = `at´ `if´, eq(#1)
14. quietly matrix score double `bpom´ = `at´ `if´, eq(#2)
15. quietly matrix score double `bdydx´ = `at´ `if´, eq(#3)
16.
17. quietly replace `reg´ = `y´ - `breg´ `if´
18. quietly replace `pom0´ = `breg´ - `bpom´ `if´
19. quietly replace `dydx´ = `breg´ - `bpom´ - `bdydx´ `if´
20. 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()`

```stata
. mat list beta
beta[1,7]
          price  __000002  __000003  length  co.__000002
    y1    -0.0002263   0 15.600867   -0.18463724   0

          c.length#
     c.__000003   _cons
    y1    -0.09302183  57.414433
```
Tricking gmm to do at()
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 $w_{it}$ ($\bar{w}$)
### Data

Contains data from `staggered_6.dta`

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Storage type</th>
<th>Display format</th>
<th>Value label</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>int</td>
<td>%9.0g</td>
<td>cross-sectional identifier</td>
</tr>
<tr>
<td>year</td>
<td>int</td>
<td>%9.0g</td>
<td>2011 to 2016</td>
</tr>
<tr>
<td>y</td>
<td>float</td>
<td>%9.0g</td>
<td>outcome, levels</td>
</tr>
<tr>
<td>w</td>
<td>byte</td>
<td>%9.0g</td>
<td>=1 if treated</td>
</tr>
<tr>
<td>x1</td>
<td>byte</td>
<td>%9.0g</td>
<td>time constant control</td>
</tr>
<tr>
<td>x2</td>
<td>byte</td>
<td>%9.0g</td>
<td>time constant control</td>
</tr>
<tr>
<td>logy</td>
<td>float</td>
<td>%9.0g</td>
<td>outcome variable, natural log</td>
</tr>
</tbody>
</table>

Sorted by: id

Note: Dataset has changed since last saved.
. 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

<table>
<thead>
<tr>
<th>cohort</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2,172</td>
<td>66.42</td>
<td>66.42</td>
</tr>
<tr>
<td>2014</td>
<td>804</td>
<td>24.59</td>
<td>91.01</td>
</tr>
<tr>
<td>2015</td>
<td>192</td>
<td>5.87</td>
<td>96.88</td>
</tr>
<tr>
<td>2016</td>
<td>102</td>
<td>3.12</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>3,270</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>
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#2014bn.year#c.x1 ///
> i.w#2015bn.cohort#2015bn.year#c.x1 ///
> i.w#2015bn.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)

| Unconditional Coefficient  | std. err. | t    | P>|t|  | [95% conf. interval] |
|-----------------------------|-----------|------|-------|---------------------|
| 0.w (base outcome)          |           |      |       |                     |
| 1.w                         |           |      |       |                     |
| over#year                   |           |      |       |                     |
| 2014 2014                   | 0.1800395 | 0.0224137 | 8.03 | 0.000 | 0.1360115 0.2240676 |
| 2014 2015                   | 0.1758216 | 0.0229169 | 7.67 | 0.000 | 0.1308052 0.2208379 |
| 2014 2016                   | 0.1849706 | 0.0251249 | 7.36 | 0.000 | 0.135617 0.2343243 |
| 2015 2014                   | 0 (omitted)|      |       |                     |
| 2015 2015                   | 0.0978163 | 0.0414103 | 2.36 | 0.019 | 0.0164725 0.17916 |
| 2015 2016                   | 0.1327046 | 0.0447888 | 2.96 | 0.003 | 0.0447245 0.2206847 |
| 2016 2014                   | 0 (omitted)|      |       |                     |
| 2016 2015                   | 0 (omitted)|      |       |                     |
| 2016 2016                   | 0.092621  | 0.0654263 | 1.42 | 0.157 | -0.0358982 0.2211401 |
### Heterogeneous-treatment-effects regression

**Data type:** Repeated cross-sectional  
**Estimator:** Two-way fixed-effects (Std. err. adjusted for 545 clusters in id)

|                   | ATET       | Robust std. err. | z    | P>|z| | [95% conf. interval] |
|-------------------|------------|------------------|------|------|----------------------|
| ATET              |            |                  |      |      |                      |
| **__cohort# year**|            |                  |      |      |                      |
| 2014 2014         | 0.1800395  | 0.0222936        | 8.08 | 0.000| 0.1363449 0.2237342  |
| 2014 2015         | 0.1758216  | 0.022794         | 7.71 | 0.000| 0.1311461 0.2204971  |
| 2014 2016         | 0.1849706  | 0.0249902        | 7.40 | 0.000| 0.1359907 0.2339505  |
| 2015 2015         | 0.0978163  | 0.0411884        | 2.37 | 0.018| 0.0170885 0.1785441  |
| 2015 2016         | 0.1327046  | 0.0445487        | 2.98 | 0.003| 0.0453907 0.2200185  |
| 2016 2016         | 0.092621   | 0.0650757        | 1.42 | 0.155| -0.034925 0.220167   |
| OME1              |            |                  |      |      |                      |
| **__cohort# year**|            |                  |      |      |                      |
| 2014 2014         | 2.401441   | 0.0765251        | 31.38| 0.000| 2.251454 2.551427    |
| OME0              |            |                  |      |      |                      |
| **__cohort# year**|            |                  |      |      |                      |
| 2014 2014         | 2.221401   | 0.0753911        | 29.47| 0.000| 2.073637 2.369165    |
| OME1              |            |                  |      |      |                      |
| **__cohort# year**|            |                  |      |      |                      |
| 2014 2015         | 2.303018   | 0.0743222        | 30.99| 0.000| 2.157349 2.448687    |
| OME0              |            |                  |      |      |                      |
| **__cohort# year**|            |                  |      |      |                      |
| 2014 2015         | 2.127196   | 0.0753975        | 28.21| 0.000| 1.97942  2.274973    |
| OME1              |            |                  |      |      |                      |
| **__cohort# year**|            |                  |      |      |                      |
| 2014 2015         | 2.303018   | 0.0743222        | 30.99| 0.000| 2.157349 2.448687    |
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
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.
• Obtain group and time cohorts
• Compute effects of interest for each group and time cohort
• Form the moment conditions
• Example for IPW. Remember:

\[ \tau_{rt} = E \left[ \left( \frac{d_r}{E[d_r]} - \frac{p_r(X)d_\infty}{1-p_r(X)} \right) (Y_t - Y_{r-1}) \right] \]
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
. // Treatment variable
. generate double gt = cohort>0 if dy!=.

(StataCorp LLC)
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>atet</td>
<td>992</td>
<td>.1982376</td>
<td>.7244637</td>
<td>-2.467532</td>
<td>3.133791</td>
</tr>
</tbody>
</table>
```
## Getting estimates

```
csid logy x1, ivar(id) time(year) gvar(cohort) method(stdipw)
```

Difference-in-difference with Multiple Time Periods

Outcome model : weighted mean
Treatment model: stabilized inverse probability

| Coefficient | Std. err. | z      | P>|z| | [95% conf. interval] |
|-------------|-----------|--------|------|----------------------|
| g2014       | 0.1982376 | 0.0271813 | 7.29 | 0.000 | 0.1449632 0.251512 |
| t_2013_2014 | 0.0271813 | 0.0271813 | 0.000 | 0.000 |

Control: Never Treated
See Callaway and Sant´Anna (2021) for details

Number of obs = 992
. 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

|                  | Robust Coefficient | 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             |
|                  | -.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             |
|                  | -.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, ...)