

Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Event Study Estimators

Jeff Wooldridge
Department of Economics
Michigan State University

New York Camp Econometrics XV
April 10, 2021

1. Introduction
2. Equivalence of TWFE and the Two-Way Mundlak Regression
3. Interventions with Common Treatment Timing
4. Staggered Event Studies
5. Nonlinear Models
6. Concluding Remarks
7. Stata Output: Staggered Case

1. Introduction

- For panel data, two-way fixed effects (TWFE) is still a staple in empirical research.
 - ▶ Applied to “structural” models – say, production functions.
 - ▶ Applied to policy analysis – event studies.
- Used for all configurations of N and T .
- With small T , large N , time effects often absorbed into covariates.
 - ▶ Analyze as a one-way FE estimator.

- The one-way Mundlak regression has proven useful for many purposes.
 - ▶ Leads to simple, robust, regression-based comparisons between FE and random effects estimation: Arellano (1993).
 - ▶ Produces insight into the pre-testing problem with Hausman tests.
 - ▶ Suggests how to allow heterogeneity to correlate with covariates in nonlinear models: Mundlak-Chamberlain device.

- Wooldridge (2019, Journal of Econometrics): The one-way Mundlak regression applies to unbalanced panels.
 - ▶ In the linear case, Mundlak still produces the complete-cases FE estimator.
 - ▶ Suggests correlated random effects for heterogeneous slopes and nonlinear models.

- Current paper: Shows the equivalence between the TWFE estimator and the obvious two-way Mundlak regression.
 - ▶ In latter case, focus is on pooled OLS, but results also hold for RE.
- Equivalence is simple but useful.
 - ▶ Further reveals the workings of TWFE.
 - ▶ Applications to linear event study estimators with staggered intervention.
 - ▶ Applications to factor models.

- Advantages of TWFE for event studies:

1. We know properties of TWFE when the panel is unbalanced.
2. It is easy to test the null that treatment effects are homogeneous in a robust way.
3. Immediately extensions of the TWFE estimator to removing unit-specific trends can be applied with heterogeneous treatment effects.

- Advantages of POLS for event studies:

- ▶ Given equivalence in the linear case, POLS can be extended to nonlinear models.

2. Equivalence of TWFE and the Two-Way Mundlak Regression

- Motivation for TWFE estimation:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + f_t + u_{it}, \quad t = 1, \dots, T; i = 1, \dots, N$$

- ▶ \mathbf{x}_{it} is $1 \times K$.
- ▶ c_i are the unit-specific effects.
- ▶ f_t are the time-specific effects.

- Results here are pure algebra.
- The two-way dummy variable regression:

y_{it} on \mathbf{x}_{it} , $1, c2_i, \dots, cN_i, f2_t, \dots, fT_t$, $t = 1, \dots, T$; $i = 1, \dots, N$.

- ▶ Coefficients on \mathbf{x}_{it} are $\hat{\boldsymbol{\beta}}_{FE}$ ($K \times 1$).
- \mathbf{x}_{it} only includes variables that have some variation across i and t .

- Baltagi (2001): Two-way within transformation gives $\hat{\beta}_{FE}$.

$$\bar{\mathbf{x}}_{i\cdot} = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}$$

$$\bar{\mathbf{x}}_{\cdot t} = N^{-1} \sum_{i=1}^N \mathbf{x}_{it}$$

$$\bar{\mathbf{x}} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it} = N^{-1} \sum_{i=1}^N \bar{\mathbf{x}}_{i\cdot} = T^{-1} \sum_{t=1}^T \bar{\mathbf{x}}_{\cdot t}$$

$$\ddot{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}) - N^{-1} \sum_{i=1}^N (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}) = \mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}$$

$$\ddot{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$$

- $\hat{\boldsymbol{\beta}}_{FE}$ is also the pooled OLS estimator from

$$\ddot{y}_{it} \text{ on } \ddot{\mathbf{x}}_{it}, t = 1, \dots, T; i = 1, \dots, N.$$

- Alternatively, consider the *two-way Mundlak regression*.
- Pooled OLS of

$$y_{it} \text{ on } 1, \mathbf{x}_{it}, \bar{\mathbf{x}}_{i\cdot}, \bar{\mathbf{x}}_{\cdot t}, t = 1, \dots, T; i = 1, \dots, N.$$

- ▶ Let $\hat{\boldsymbol{\beta}}_M$ be the coefficients \mathbf{x}_{it} .

THEOREM: Provided the $K \times K$ matrix

$$\sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}'_{it} \ddot{\mathbf{x}}_{it}$$

is nonsingular,

$$\hat{\boldsymbol{\beta}}_M = \hat{\boldsymbol{\beta}}_{FE}$$

Moreover, in the extended regression

$$y_{it} \text{ on } 1, \mathbf{x}_{it}, \bar{\mathbf{x}}_{i.}, \bar{\mathbf{x}}_{.t}, \mathbf{z}_i, \mathbf{m}_t, t = 1, \dots, T; i = 1, \dots, N$$

for time-constant variables \mathbf{z}_i and unit-constant variables \mathbf{m}_t , the coefficients on \mathbf{x}_{it} are still $\hat{\boldsymbol{\beta}}_{FE}$. \square

- Proof uses Frisch-Waugh partialling out.
- Coefficients on $\bar{\mathbf{x}}_{i\cdot}$ and $\bar{\mathbf{x}}_{\cdot t}$ do change with the inclusion of $\mathbf{z}_i, \mathbf{m}_t$.
 - Basis of robust, regression-based Hausman tests.
- Random effects-type estimators give the same conclusions.
- Suppose a regressor is an interaction of the form

$$x_{itj} = z_{ij} \cdot m_{tj}$$

► Then

$$\bar{x}_{i\cdot j} = z_{ij} \cdot \bar{m}_j, \quad \bar{x}_{\cdot tj} = \bar{z}_j m_{tj}$$

► Mundlak regression includes z_{ij}, m_{tj} as controls.

3. Interventions with Common Treatment Timing

- T time periods.
 - ▶ $t = 1, \dots, q - 1$ are control periods.
 - ▶ Intervention happens at $t = q$, remains in place.
- Treatment indicator:

$$w_{it} = d_i \cdot p_t$$

$$d_i = 1 \text{ if (eventually) treated}$$

$$p_t = fq_t + \dots + fT_t = 1 \text{ if a post treatment period}$$

- Homogeneous treatment effect.
- Equation that motivates TWFE:

$$y_{it} = \beta w_{it} + c_i + g_t + u_{it}, t = 1, \dots, T; i = 1, 2, \dots, N$$

$$\bar{w}_{i\cdot} = d_i \bar{p}$$

$$\bar{w}_{\cdot t} = \bar{d} p_t$$

- TWM regression is equivalent to the DID regression

$$y_{it} \text{ on } 1, w_{it}, d_i, p_t, t = 1, \dots, T; i = 1, \dots, N$$

- ▶ $\hat{\beta}_{DD} = \hat{\beta}_{FE}$. Enough to control for d_i, p_t .

- The TWFE estimator has the familiar form

$$\hat{\beta}_{FE} = \hat{\beta}_{DD} = (\bar{y}_1^{post} - \bar{y}_0^{post}) - (\bar{y}_1^{pre} - \bar{y}_0^{pre})$$

- Using separate time dummies $f2_t, \dots, fT_t$ in place of p_t has no effect on $\hat{\beta}_{DD}$.
- Adding controls, \mathbf{x}_i or $d_i \cdot \mathbf{x}_i$, has no effect on $\hat{\beta}_{DD}$.

- Now allow TEs to change over treatment period.

$$y_{it} = \beta_q(w_{it} \cdot fq_t) + \cdots + \beta_T(w_{it} \cdot fT_t) + c_i + g_t + u_{it}$$

- Can use TWFE to estimate the β_r .

$$w_{it} \cdot fr_t = d_i \cdot p_t \cdot fr_t = d_i(fq_t + \cdots + fT_t)fr_t = d_i fr_t$$

- Time averages are proportional to d_i .
- Cross-sectional averages proportional to fr_t .
- TWM equation:

$$y_{it} = \alpha + \beta_q(w_{it} \cdot fq_t) + \cdots + \beta_T(w_{it} \cdot fT_t) + \zeta d_i + \theta_q fq_t + \cdots + \theta_T fT_t + e_{it}$$

- POLS and RE give identical estimates.

- Add time-constant covariates, \mathbf{x}_i
- Coefficients on change with time period.
- Treatment effects change with t , \mathbf{x}_i , and possibly both:

$$\begin{aligned}
y_{it} = & \beta_q(w_{it} \cdot fq_t) + \cdots + \beta_T(w_{it} \cdot fT_t) + [w_{it} \cdot fq_t \cdot (\mathbf{x}_i - \boldsymbol{\mu}_1)]\boldsymbol{\gamma}_q \\
& + \cdots + [w_{it} \cdot fT_t \cdot (\mathbf{x}_i - \boldsymbol{\mu}_1)]\boldsymbol{\gamma}_T \\
& + (fq_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_q + \cdots + (fT_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_T + c_i + g_t + u_{it}
\end{aligned}$$

$$\boldsymbol{\mu}_1 \equiv E(\mathbf{x}_i | d_i = 1)$$

- Can estimate by TWFE or TWM.

- TWM includes the time-constant variables $d_i, \mathbf{x}_i, d_i \cdot \mathbf{x}_i$.
- Need time dummies for fq_t, \dots, fT_t .

$$\begin{aligned}
y_{it} = & \alpha + \beta_q(w_{it} \cdot fq_t) + \dots + \beta_T(w_{it} \cdot fT_t) + [w_{it} \cdot fq_t \cdot (\mathbf{x}_i - \boldsymbol{\mu}_1)]\boldsymbol{\gamma}_q \\
& + \dots + [w_{it} \cdot fT_t \cdot (\mathbf{x}_i - \boldsymbol{\mu}_1)]\boldsymbol{\gamma}_T \\
& + (fq_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_q + \dots + (fT_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_T + \zeta d_i + \mathbf{x}_i\boldsymbol{\xi} + (d_i \cdot \mathbf{x}_i)\boldsymbol{\lambda} \\
& + \theta_q fq_t + \dots + \theta_T fT_t + e_{it}
\end{aligned}$$

- Replace $\boldsymbol{\mu}_1$ with $\bar{\mathbf{x}}_1 = N_1^{-1} \sum_{i=1}^N d_i \cdot \mathbf{x}_i$.
 - In principle, account for sampling error. See did_4.do.

- Can connect this to period-by-period regression adjustment.
- Use periods $1, \dots, q - 1$ as the control and then each $r \in \{q, q + 1, \dots, T\}$ in turn:
- Can estimate

$$\begin{aligned}
y_{it} = & \alpha + \beta_r w_{it} + w_{it} \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_1) \boldsymbol{\gamma}_r + (fr_t \cdot \mathbf{x}_i) \boldsymbol{\delta}_r \\
& + \zeta d_i + d_i \cdot \mathbf{x}_i \boldsymbol{\lambda} + \theta_r fr_t + error_{it} \\
t \in & \{1, \dots, q - 1, r\}, i = 1, \dots, N
\end{aligned}$$

by POLS.

- Parametric version of Callaway and Sant'Anna (2020), but regression in levels.
- Estimates are identical to pooling all time periods.
- Doing separately has advantage of using doubly robust estimators (IPW and RA).

- In using a canned package – such as Stata and its `margins` option – need to be careful about how to specify interactions.
- Regression should look like

$$y_{it} \text{ on } 1, d_i, \mathbf{x}_i, d_i \cdot \mathbf{x}_i, fq_t, \dots, fT_t, fq_t \cdot \mathbf{x}_i, \dots, fT_t \cdot \mathbf{x}_i, \\ w_{it} \cdot fq_t, \dots, w_{it} \cdot fT_t, w_{it} \cdot fq_t \cdot \mathbf{x}_i, \dots, w_{it} \cdot fT_t \cdot \mathbf{x}_i$$

- Then use

```
margins, dydx(w) at(d = 1 fq = 1 fqp1 = 0
... fT = 0), subpop(if d == 1) vce(uncon)
```

```
margins, dydx(w) at(d = 1 fq = 0 fqp1 = 1
... fT = 0), subpop(if d == 1) vce(uncon)
```

```
margins, dydx(w) at(d = 1 fq = 0 fqp1 = 0
... fT = 1), subpop(if d == 1) vce(uncon)
```

- Same with TWFE.
- See `did_4.do`.

What is Being Estimated?

- Potential outcomes, $y_t(0)$ and $y_t(1)$.
- Treatment effect for a generic unit:

$$te_t = y_t(1) - y_t(0)$$

- We can identify the ATT in each treated period:

$$\tau_t \equiv E[y_t(1) - y_t(0)|d = 1], t = q, q + 1, \dots, T$$

Assumption CT (Common Trend): With the (eventually) treated indicator d ,

$$E[y_t(0) - y_1(0)|d] = E[y_t(0) - y_1(0)] \equiv \theta_t, t = 2, \dots, T. \quad \square$$

- With covariates:

Assumption CCT (Conditional Common Trends): For treatment indicator d and covariates \mathbf{x} ,

$$E[y_t(0) - y_1(0)|d, \mathbf{x}] = E[y_t(0) - y_1(0)|\mathbf{x}], \quad t = 2, \dots, T. \quad \square$$

- Abadie (2005) uses this with $T = 2$.
- Similar to Callaway and Sant'Anna (2020); Sun and Abraham (2021); and others.

4. Staggered Treatments

- Wooldridge (2005): Studied robustness of FE to heterogeneous slopes.
 - ▶ Suggested some strategies to allow heterogeneity based on intensity.
- TWFE under recent scrutiny for staggered (and more general) interventions.

- de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2018), Callaway and Sant'Anna (2021), Sun and Abraham (2021).
- Previous analysis shows TWFE is fine in common intervention time case provided TEs are allowed to change across t and with \mathbf{x} .
- Can use TWFE or TWM in staggered case with lots of heterogeneity.

- First intervention period is $t = q$.
- Some units treated in each period after q , up to T .
- No reversibility.
- Define treatment cohorts by dummies: d_q, \dots, d_T .
 - ▶ $d_r = 1$ if unit enters treatment in period r .
 - ▶ Essentially intensity indicators.

- At least two way to define treatment effects.

1. Have two potential outcomes again, but the TEs are

$$E[y_t(1) - y_t(0)|d_r = 1], r = q, \dots, T; t = r, \dots, T$$

2. Think of the intensities defining different potential outcomes:

$$\tau_{rt} \equiv E[y_t(r) - y_t(0)|d_r = 1], r = q, \dots, T; t = r, \dots, T$$

- Second is easier to deal with.
 - Both lead to the same estimating equation.

Assumption CTS (Common Trend, Staggered): With the exposure dummies d_q, \dots, d_T ,

$$E[y_t(0) - y_1(0)|d_q, \dots, d_T] = E[y_t(0) - y_1(0)] \equiv \theta_t, \quad t = 2, \dots, T. \quad \square$$

- Again, similar to Callaway and Sant'Anna; Sun and Abraham; others.
- The paper derives the following for all t :

$$E(y_{it}|\mathbf{d}_i) = \eta + \lambda_q d_{iq} + \dots + \lambda_T d_{iT} + \sum_{s=2}^T \theta_s f s_t + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (w_{it} \cdot d_{ir} \cdot f s_t)$$

- This is the Mundlak equation.
 - ▶ Time dummies for $t < q$ are redundant.
- Mundlak regression:

$$y_{it} \text{ on } 1, d_{iq}, \dots, d_{iT}, fq_t, \dots, fT_t, \\ w_{it} \cdot d_{iq} \cdot fq_t, \dots, w_{it} \cdot d_{iq} \cdot fT_t, \dots, w_{it} \cdot d_{iT} \cdot fT_t$$

- Equivalently, can use TWFE and drop everything except

$$w_{it} \cdot d_{iq} \cdot fq_t, \dots, w_{it} \cdot d_{iq} \cdot fT_t, \dots, w_{it} \cdot d_{iT} \cdot fT_t$$

- In practice, may need to impose commonality across cohort/time.

- Add covariates.

Assumption CCTS (Conditional Common Trends, Staggered):

For exposure indicators d_r and covariates \mathbf{x} ,

$$E[y_t(0) - y_1(0)|d_q, \dots, d_T, \mathbf{x}] = E[y_t(0) - y_1(0)|\mathbf{x}], \quad t = 2, \dots, T. \quad \square$$

- Assume all conditional expectations are linear in \mathbf{x} .
- This means linearity conditional on each $d_r = 1, r = q, \dots, T$.

$$\dot{\mathbf{x}}_r \equiv \mathbf{x} - E(\mathbf{x}|d_r = 1), \quad r = q, \dots, T$$

- Paper derives

$$\begin{aligned}
E(y_t|d_q, \dots, d_T, \mathbf{x}) = & \eta + \sum_{r=q}^T \lambda_r d_r + \mathbf{x}\boldsymbol{\kappa} + \sum_{r=q}^T (d_r \cdot \mathbf{x}) \zeta_r \\
& + \sum_{s=2}^T \theta_s f s_t + \sum_{s=2}^T (f s_t \cdot \mathbf{x}) \pi_t \\
& + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (d_r \cdot f s_t) + \sum_{r=q}^T \sum_{s=r}^T (d_r \cdot f s_t \cdot \dot{\mathbf{x}}_r) \rho_{rs}.
\end{aligned}$$

- Słoczyński (forthcoming, REStat) in cross-sectional case with unconfoundness: Can be important to allow for slope heterogeneity.

- The regression is

$$\begin{aligned}
& y_{it} \text{ on } 1, d_{iq}, \dots, d_{iT}, \mathbf{x}_i, d_{iq} \cdot \mathbf{x}_i, \dots, d_{iT} \cdot \mathbf{x}_i, \\
& \quad fq_t, \dots, fT_t, fq_t \cdot \mathbf{x}_i, \dots, fT_t \cdot \mathbf{x}_i, \\
& \quad w_{it} \cdot d_{iq} \cdot fq_t, \dots, w_{it} \cdot d_{iq} \cdot fT_t, \dots, \\
& \quad w_{it} \cdot d_{i,q+1} \cdot f(q+1)_t, \dots, w_{it} \cdot d_{i,q+1} \cdot fT_t, \dots, w_{it} \cdot d_{iT} \cdot fT_t, \\
& \quad w_{it} \cdot d_{iq} \cdot fq_t \cdot \dot{\mathbf{x}}_{iq}, \dots, w_{it} \cdot d_{iq} \cdot fT_t \cdot \dot{\mathbf{x}}_{iT}, \\
& \quad w_{it} \cdot d_{i,q+1} \cdot f(q+1)_t \cdot \dot{\mathbf{x}}_{iq}, \dots, w_{it} \cdot d_{i,q+1} \cdot fT_t \cdot \dot{\mathbf{x}}_{iT}, \dots, \\
& \quad w_{it} \cdot d_{iT} \cdot fT_t \cdot \dot{\mathbf{x}}_{iT}
\end{aligned}$$

$$\dot{\mathbf{x}}_{ir} = \mathbf{x}_i - \bar{\mathbf{x}}_r = \mathbf{x}_i - N_r^{-1} \sum_{h=1}^N d_{hr} \mathbf{x}_h.$$

- RE gives identical estimates.
 - ▶ Improving over POLS requires allowing more general patterns of serial correlation and maybe time-varying variances.
- Or drop everything in the first line except $f q_t \cdot \mathbf{x}_i, \dots, f T_t \cdot \mathbf{x}_i$ and use TWFE.
 - ▶ Numerically identical.
- Can use Stata and `margins` to account for sampling variation in $\bar{\mathbf{x}}_r$.
 - ▶ See `staggered_6.do`.

5. Nonlinear Models

- With small T , TWFE cannot be applied to most nonlinear models.
 - ▶ Exception is Poisson regression.
- The pooled OLS methods extend to any nonlinear model.
- Suppose $y_{it} \geq 0$.

- Exponential mean is natural:

$$E(y_{it}|d_{iq}, \dots, d_{iT}) = \exp \left[\eta + \lambda_q d_{iq} + \dots + \lambda_T d_{iT} + \sum_{s=2}^T \theta_s f_{s_t} + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (w_{it} \cdot d_{ir} \cdot f_{s_t}) \right]$$

- Can be estimated by pooled Poisson.
 - Estimator is fully robust; cluster standard errors (with T small).

- Or, can drop the d_{ir} and introduce multiplicative heterogeneity:

$$E(y_{it}|d_{iq}, \dots, d_{iT}, c_i) = c_i \exp \left[\sum_{s=2}^T \theta_s f s_t + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (w_{it} \cdot d_{ir} \cdot f s_t) \right].$$

- Estimate by FE Poisson.
 - Estimates are identical to POLS – as in linear case.

- No longer equivalent with covariates.
- Can estimate

$$\begin{aligned}
 E(y_{it}|d_{iq}, \dots, d_{iT}, \mathbf{x}_i, c_i) = & c_i \exp \left[\sum_{s=2}^T \theta_s f s_t + \sum_{s=2}^T (f s_t \cdot \mathbf{x}_i) \boldsymbol{\pi}_s \right. \\
 & + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (w_{it} \cdot d_{ir} \cdot f s_t) \\
 & \left. + \sum_{r=q}^T \sum_{s=r}^T (w_{it} \cdot d_{ir} \cdot f s_t \cdot \dot{\mathbf{x}}_{ir}) \boldsymbol{\rho}_{rs} \right]
 \end{aligned}$$

by FE Poisson.

- Fully robust, allows unbalanced panels.

- Underlying common trends assumption to identify ATTs:

$$\frac{E[y_t(0)|d_q, \dots, d_T, \mathbf{x}]}{E[y_1(0)|d_q, \dots, d_T, \mathbf{x}]} = \frac{E[y_t(0)|\mathbf{x}]}{E[y_1(0)|\mathbf{x}]}$$

- For other nonlinear response functions – logit, probit – CT assumption is on underlying latent variable.

6. Concluding Remarks

- Equivalence between TWFE and TWM has applications to event study estimators.
 - ▶ Show TWFE can be made much more flexible than constant effect.
 - ▶ TWFE some resilience to unbalanced panels.
 - ▶ TWFE can be used for exponential models.
 - ▶ TWM extends to general nonlinear models (logit, fractional logit).
- Equivalence may also suggest different strategies for factor models.

7. Stata Output: Staggered Case

```
. use staggered_6, clear

.
. xtset id year
      panel variable:  id (strongly balanced)
      time variable:  year, 2011 to 2016
                  delta:  1 unit

. gen f2014 = year == 2014

. gen f2015 = year == 2015

. gen f2016 = year == 2016

. egen wsum = sum(w), by(id)

. gen d4 = wsum == 3

. gen d5 = wsum == 2

. gen d6 = wsum == 1
```

•

R-sq:

```
corr(u_i, Xb) = -0.0655
```

Obs per group:

F(4 , 544)	=	77.56
Prob > F	=	0.0000

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
w	.1633485	.0148916	10.97	0.000	.1340964	.1926007
f2014	.0677731	.010606	6.39	0.000	.0468972	.0885649
f2015	-.0314376	.0109958	-2.86	0.004	-.0530371	-.0098381
f2016	.061815	.0109547	5.64	0.000	.0402962	.0833337
_cons	2.412875	.0035134	686.77	0.000	2.405973	2.419776
sigma_u	.97499069					
sigma_e	.19979414					
rho	.95970045	(fraction of variance due to u_i)				

```
. reg logy w f2014 f2015 f2016 d4 d5 d6, vce(cluster id)
```

Linear regression	Number of obs	=	3,270
	F(7, 544)	=	46.52
	Prob > F	=	0.0000
	R-squared	=	0.0164
	Root MSE	=	.98206

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
w	.1633485	.0148985	10.96	0.000	.1340829	.1926142
f2014	.067731	.0106109	6.38	0.000	.0468876	.0885744
f2015	-.0314376	.0110009	-2.86	0.004	-.053047	-.0098282
f2016	.061815	.0109598	5.64	0.000	.0402863	.0833436
d4	-.3287467	.0910431	-3.61	0.000	-.5075857	-.1499077
d5	-.0187403	.2060319	-0.09	0.928	-.4234558	.3859753
d6	.0190121	.2663328	0.07	0.943	-.5041546	.5421789
_cons	2.494212	.0521605	47.82	0.000	2.391751	2.596672

```
. * TE varies by period:
```

```
.
. xtreg logy c.w#c.f2014 c.w#c.f2015 c.w#c.f2016 ///
>      f2014 f2015 f2016, fe vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.w#c.f2014	.1700551	.0220181	7.72	0.000	.1268043	.2133059
c.w#c.f2015	.1569093	.021125	7.43	0.000	.1154127	.198406
c.w#c.f2016	.1639358	.0221439	7.40	0.000	.1204379	.2074338
f2014	.0660821	.0114382	5.78	0.000	.0436136	.0885506
f2015	-.0294763	.0118883	-2.48	0.013	-.0528288	-.0061237
f2016	.0616178	.0119632	5.15	0.000	.0381181	.0851174
_cons	2.412875	.0035142	686.61	0.000	2.405972	2.419778
sigma_u	.97500299					
sigma_e	.19985916					
rho	.95967624	(fraction of variance due to u_i)				

```
. reg logy c.w#c.f2014 c.w#c.f2015 c.w#c.f2016 ///
> f2014 f2015 f2016 d4 d5 d6, vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.w#c.f2014	.1700551	.0220282	7.72	0.000	.1267844	.2133258
c.w#c.f2015	.1569093	.0211348	7.42	0.000	.1153936	.1984251
c.w#c.f2016	.1639358	.022154	7.40	0.000	.1204179	.2074538
f2014	.0660821	.0114435	5.77	0.000	.0436032	.0885609
f2015	-.0294763	.0118937	-2.48	0.014	-.0528396	-.006113
f2016	.0616178	.0119687	5.15	0.000	.0381073	.0851283
d4	-.3288891	.0910915	-3.61	0.000	-.5078233	-.149955
d5	-.017765	.2060861	-0.09	0.931	-.422587	.387057
d6	.0189143	.2664067	0.07	0.943	-.5043976	.5422262
_cons	2.494192	.0521756	47.80	0.000	2.391702	2.596683

```
. xtreg logy c.w#c.f2014 c.w#c.f2015 c.w#c.f2016 ///
> f2014 f2015 f2016 d4 d5 d6, re vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
c.w#c.f2014	.1700551	.0220282	7.72	0.000	.1268807	.2132296
c.w#c.f2015	.1569093	.0211348	7.42	0.000	.115486	.1983327
c.w#c.f2016	.1639358	.022154	7.40	0.000	.1205147	.207357
f2014	.0660821	.0114435	5.77	0.000	.0436532	.0885109
f2015	-.0294763	.0118937	-2.48	0.013	-.0527876	-.006165
f2016	.0616178	.0119687	5.15	0.000	.0381596	.085076
d4	-.3288891	.0910915	-3.61	0.000	-.5074252	-.1503531
d5	-.017765	.2060861	-0.09	0.931	-.4216863	.3861564
d6	.0189143	.2664067	0.07	0.943	-.5032334	.5410619
_cons	2.494192	.0521756	47.80	0.000	2.39193	2.596455
sigma_u	.9638903					
sigma_e	.19985916					
rho	.95877965	(fraction of variance due to u_i)				

```
. * TE varies by entry cohort (intensity):
.
. xtreg logy c.w#c.d4 c.w#c.d5 c.w#c.d6 ///
>      f2014 f2015 f2016, fe vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.w#c.d4	.1800497	.0163678	11.00	0.000	.147898	.2122015
c.w#c.d5	.1133156	.0314756	3.60	0.000	.0514869	.1751442
c.w#c.d6	.0919859	.0637986	1.44	0.150	-.0333359	.2173078
f2014	.0636247	.0107827	5.90	0.000	.0424438	.0848055
f2015	-.0326062	.0110059	-2.96	0.003	-.0542254	-.010987
f2016	.0628723	.0109928	5.72	0.000	.0412789	.0844658
_cons	2.412875	.0035181	685.84	0.000	2.405964	2.419786
sigma_u	.97563289					
sigma_e	.19965948					
rho	.9598034	(fraction of variance due to u_i)				


```
. reg logy c.w#c.d4 c.w#c.d5 c.w#c.d6 ///
> f2014 f2015 f2016 d4 d5 d6, vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.w#c.d4	.1800497	.0163753	11.00	0.000	.1478832	.2122163
c.w#c.d5	.1133156	.0314901	3.60	0.000	.0514584	.1751727
c.w#c.d6	.0919859	.063828	1.44	0.150	-.0333936	.2173654
f2014	.0636247	.0107877	5.90	0.000	.0424341	.0848153
f2015	-.0326062	.0110109	-2.96	0.003	-.0542354	-.010977
f2016	.0628723	.0109978	5.72	0.000	.0412689	.0844757
d4	-.3370973	.0910864	-3.70	0.000	-.5160215	-.1581731
d5	-.0020626	.2058754	-0.01	0.992	-.4064708	.4023456
d6	.0309059	.2665283	0.12	0.908	-.4926449	.5544567
_cons	2.494915	.0521823	47.81	0.000	2.392411	2.597418

```
. xtreg logy c.w#c.d4 c.w#c.d5 c.w#c.d6 ///
> f2014 f2015 f2016 d4 d5 d6, re vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
c.w#c.d4	.1800497	.0163753	11.00	0.000	.1479547	.2121447
c.w#c.d5	.1133156	.0314901	3.60	0.000	.0515961	.1750351
c.w#c.d6	.0919859	.063828	1.44	0.150	-.0331146	.2170865
f2014	.0636247	.0107877	5.90	0.000	.0424812	.0847681
f2015	-.0326062	.0110109	-2.96	0.003	-.0541872	-.0110252
f2016	.0628723	.0109978	5.72	0.000	.041317	.0844277
d4	-.3370973	.0910864	-3.70	0.000	-.5156234	-.1585712
d5	-.0020626	.2058754	-0.01	0.992	-.4055711	.4014458
d6	.0309059	.2665283	0.12	0.908	-.49148	.5532919
_cons	2.494915	.0521823	47.81	0.000	2.392639	2.59719
sigma_u	.9638972					
sigma_e	.19965948					
rho	.95885915	(fraction of variance due to u_i)				

```

. * TE varies by intensity (cohort) and calendar year:
.
. xtreg logy c.d4#c.f2014 c.d4#c.f2015 c.d4#c.f2016 ///
>       c.d5#c.f2015 c.d5#c.f2016 ///
>       c.d6#c.f2016 ///
>       f2014 f2015 f2016, fe vce(cluster id)

```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.d4#c.f2014	.1799818	.0223482	8.05	0.000	.1360826	.2238811
c.d4#c.f2015	.175937	.0228629	7.70	0.000	.1310266	.2208473
c.d4#c.f2016	.1843389	.0249739	7.38	0.000	.1352819	.2333959
c.d5#c.f2015	.0986968	.0411317	2.40	0.017	.0179005	.1794931
c.d5#c.f2016	.1280679	.0439805	2.91	0.004	.0416755	.2144603
c.d6#c.f2016	.0943579	.0640588	1.47	0.141	-.031475	.2201909
f2014	.0636414	.0114978	5.54	0.000	.0410559	.0862268
f2015	-.0307366	.0118934	-2.58	0.010	-.0540993	-.007374
f2016	.0608776	.0119708	5.09	0.000	.037363	.0843921
_cons	2.412875	.0035195	685.57	0.000	2.405961	2.419788
sigma_u	.97563252					
sigma_e	.19975631					
rho	.95976594	(fraction of variance due to u_i)				

```
. reg logy c.d4#c.f2014 c.d4#c.f2015 c.d4#c.f2016 ///
>      c.d5#c.f2015 c.d5#c.f2016 ///
>      c.d6#c.f2016 ///
>      f2014 f2015 f2016 d4 d5 d6, vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.d4#c.f2014	.1799818	.0223584	8.05	0.000	.1360624	.2239013
c.d4#c.f2015	.175937	.0228734	7.69	0.000	.131006	.220868
c.d4#c.f2016	.1843389	.0249854	7.38	0.000	.1352593	.2334185
c.d5#c.f2015	.0986968	.0411506	2.40	0.017	.0178633	.1795303
c.d5#c.f2016	.1280679	.0440007	2.91	0.004	.0416357	.2145001
c.d6#c.f2016	.0943579	.0640883	1.47	0.142	-.0315329	.2202488
f2014	.0636414	.0115031	5.53	0.000	.0410455	.0862372
f2015	-.0307366	.0118989	-2.58	0.010	-.0541101	-.0073632
f2016	.0608776	.0119763	5.08	0.000	.0373522	.084403
d4	-.3371154	.0911247	-3.70	0.000	-.5161148	-.158116
d5	-.0020849	.205955	-0.01	0.992	-.4066494	.4024797
d6	.0305106	.2666447	0.11	0.909	-.4932688	.55429
_cons	2.494933	.0522037	47.79	0.000	2.392387	2.597478

```
. xtreg logy c.d4#c.f2014 c.d4#c.f2015 c.d4#c.f2016 ///
>      c.d5#c.f2015 c.d5#c.f2016 ///
>      c.d6#c.f2016 ///
>      f2014 f2015 f2016 d4 d5 d6, re vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
c.d4#c.f2014	.1799818	.0223584	8.05	0.000	.1361601	.2238036
c.d4#c.f2015	.175937	.0228734	7.69	0.000	.1311059	.220768
c.d4#c.f2016	.1843389	.0249854	7.38	0.000	.1353685	.2333093
c.d5#c.f2015	.0986968	.0411506	2.40	0.016	.0180431	.1793505
c.d5#c.f2016	.1280679	.0440007	2.91	0.004	.041828	.2143078
c.d6#c.f2016	.0943579	.0640883	1.47	0.141	-.0312529	.2199687
f2014	.0636414	.0115031	5.53	0.000	.0410958	.086187
f2015	-.0307366	.0118989	-2.58	0.010	-.0540581	-.0074152
f2016	.0608776	.0119763	5.08	0.000	.0374045	.0843506
d4	-.3371154	.0911247	-3.70	0.000	-.5157165	-.1585142
d5	-.0020849	.205955	-0.01	0.992	-.4057493	.4015796
d6	.0305106	.2666447	0.11	0.909	-.4921035	.5531247
_cons	2.494933	.0522037	47.79	0.000	2.392615	2.59725
sigma_u	.96389386					

sigma_e		.19975631	
rho		.9588206	(fraction of variance due to u_i)

```
. sum x1 if d4
```

Variable		Obs	Mean	Std. Dev.	Min	Max
x1		804	11.73881	1.197314	8	15

```
. gen x1_dm4 = x1 - r(mean)
```

```
. sum x1 if d5
```

Variable		Obs	Mean	Std. Dev.	Min	Max
x1		192	11.28125	1.529488	6	14

```
. gen x1_dm5 = x1 - r(mean)
```

```
. sum x1 if d6
```

Variable		Obs	Mean	Std. Dev.	Min	Max
x1		102	12	1.980099	9	15

```
. gen x1_dm6 = x1 - r(mean)
```

```
. xtreg logy c.w#c.d4#c.f2014 c.w#c.d4#c.f2015 c.w#c.d4#c.f2016 ///
> c.w#c.d5#c.f2015 c.w#c.d5#c.f2016 ///
> c.w#c.d6#c.f2016 ///
> c.w#c.d4#c.f2014#c.x1_dm4 c.w#c.d4#c.f2015#c.x1_dm4 c.w#c.d4#c.f2016#c.x1_dm4 /
> c.w#c.d5#c.f2015#c.x1_dm5 c.w#c.d5#c.f2016#c.x1_dm5 ///
> c.w#c.d6#c.f2016#c.x1_dm6 ///
> f2014 f2015 f2016 c.f2014#c.x1 c.f2015#c.x1 c.f2016#c.x1, fe vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval	
c.w#c.d4#c.f2014	.1800395	.0222895	8.08	0.000	.1362555	.2238236
c.w#c.d4#c.f2015	.1758216	.0225127	7.81	0.000	.1315991	.220044
c.w#c.d4#c.f2016	.1849706	.0249513	7.41	0.000	.1359579	.2339833
c.w#c.d5#c.f2015	.0978163	.0413402	2.37	0.018	.0166103	.1790222
c.w#c.d5#c.f2016	.1327046	.0434009	3.06	0.002	.0474506	.2179586
c.w#c.d6#c.f2016	.092621	.062385	1.48	0.138	-.0299239	.2151659
c.w#c.d4#c.f2014#c.x1_dm4	-.0189186	.0161854	-1.17	0.243	-.0507122	.012875
c.w#c.d4#c.f2015#c.x1_dm4	-.0392521	.0176217	-2.23	0.026	-.0738671	-.0046372
c.w#c.d4#c.f2016#c.x1_dm4	-.0246808	.0215333	-1.15	0.252	-.0669795	.0176178
c.w#c.d5#c.f2015#c.x1_dm5	.0001442	.0237887	0.01	0.995	-.0465848	.0468732

c.w#c.d5#c.f2016#c.x1_dm5	-.039724	.0212825	-1.87	0.063	-.0815299	.0020819
c.w#c.d6#c.f2016#c.x1_dm6	-.0403321	.0322018	-1.25	0.211	-.1035872	.022923
f2014	.0454511	.0707414	0.64	0.521	-.0935088	.1844109
f2015	-.0158952	.0783568	-0.20	0.839	-.1698141	.1380238
f2016	-.0474077	.0795173	-0.60	0.551	-.2036063	.1087908
c.f2014#c.x1	.0015447	.0060751	0.25	0.799	-.0103889	.0134783
c.f2015#c.x1	-.0012545	.0065546	-0.19	0.848	-.0141299	.0116209
c.f2016#c.x1	.0091707	.0066265	1.38	0.167	-.0038459	.0221874
_cons	2.412875	.0035054	688.34	0.000	2.405989	2.419761

sigma_u	.97594894					
sigma_e	.1996671					
rho	.95982544	(fraction of variance due to u_i)				

```
. reg logy c.w#c.d4#c.f2014 c.w#c.d4#c.f2015 c.w#c.d4#c.f2016 ///
> c.w#c.d5#c.f2015 c.w#c.d5#c.f2016 ///
> c.w#c.d6#c.f2016 ///
> c.w#c.d4#c.f2014#c.x1_dm4 c.w#c.d4#c.f2015#c.x1_dm4 c.w#c.d4#c.f2016#c.x1_dm4 /
> c.w#c.d5#c.f2015#c.x1_dm5 c.w#c.d5#c.f2016#c.x1_dm5 ///
> c.w#c.d6#c.f2016#c.x1_dm6 ///
> f2014 f2015 f2016 c.f2014#c.x1 c.f2015#c.x1 c.f2016#c.x1 ///
> d4 d5 d6 x1 c.d4#c.x1 c.d5#c.x1 c.d6#c.x1, vce(cluster id)
```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval	
c.w#c.d4#c.f2014	.1800395	.0223136	8.07	0.000	.1362082	.2238708
c.w#c.d4#c.f2015	.1758216	.022537	7.80	0.000	.1315514	.2200917
c.w#c.d4#c.f2016	.1849706	.0249782	7.41	0.000	.1359051	.2340361
c.w#c.d5#c.f2015	.0978163	.0413848	2.36	0.018	.0165227	.1791098
c.w#c.d5#c.f2016	.1327046	.0434477	3.05	0.002	.0473587	.2180505
c.w#c.d6#c.f2016	.092621	.0624522	1.48	0.139	-.0300561	.2152981
c.w#c.d4#c.f2014#c.x1_dm4	-.0189186	.0162029	-1.17	0.243	-.0507465	.0129092
c.w#c.d4#c.f2015#c.x1_dm4	-.0392521	.0176407	-2.23	0.026	-.0739044	-.0045999
c.w#c.d4#c.f2016#c.x1_dm4	-.0246808	.0215565	-1.14	0.253	-.0670251	.0176634
c.w#c.d5#c.f2015#c.x1_dm5	.0001442	.0238144	0.01	0.995	-.0466352	.0469236

c.w#c.d5#c.f2016#c.x1_dm5	-.039724	.0213054	-1.86	0.063	-.081575	.002127
c.w#c.d6#c.f2016#c.x1_dm6	-.0403321	.0322365	-1.25	0.211	-.1036554	.0229912
f2014	.0454511	.0708177	0.64	0.521	-.0936586	.1845608
f2015	-.0158952	.0784413	-0.20	0.839	-.1699801	.1381897
f2016	-.0474077	.079603	-0.60	0.552	-.2037747	.1089592
c.f2014#c.x1	.0015447	.0060817	0.25	0.800	-.0104018	.0134911
c.f2015#c.x1	-.0012545	.0065617	-0.19	0.848	-.0141438	.0116348
c.f2016#c.x1	.0091707	.0066336	1.38	0.167	-.0038599	.0222014
d4	-1.356329	.7810326	-1.74	0.083	-2.890538	.17788
d5	1.191999	1.144603	1.04	0.298	-1.056385	3.440383
d6	1.690494	1.679344	1.01	0.315	-1.6083	4.989288
x1	.0682135	.025351	2.69	0.007	.0184158	.1180113
c.d4#c.x1	.0872358	.0672249	1.30	0.195	-.0448164	.2192881
c.d5#c.x1	-.1026349	.1014854	-1.01	0.312	-.3019862	.0967163
c.d6#c.x1	-.1394164	.1293121	-1.08	0.281	-.3934286	.1145958
_cons	1.689357	.3034252	5.57	0.000	1.093328	2.285385

```

. xtreg logy c.w#c.d4#c.f2014 c.w#c.d4#c.f2015 c.w#c.d4#c.f2016 ///
> c.w#c.d5#c.f2015 c.w#c.d5#c.f2016 ///
> c.w#c.d6#c.f2016 ///
> c.w#c.d4#c.f2014#c.x1_dm4 c.w#c.d4#c.f2015#c.x1_dm4 c.w#c.d4#c.f2016#c.x1_dm4 /
> c.w#c.d5#c.f2015#c.x1_dm5 c.w#c.d5#c.f2016#c.x1_dm5 ///
> c.w#c.d6#c.f2016#c.x1_dm6 ///
> f2014 f2015 f2016 c.f2014#c.x1 c.f2015#c.x1 c.f2016#c.x1 ///
> d4 d5 d6 x1 c.d4#c.x1 c.d5#c.x1 c.d6#c.x1, re vce(cluster id)

```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval	
c.w#c.d4#c.f2014	.1800395	.0223136	8.07	0.000	.1363058	.2237733
c.w#c.d4#c.f2015	.1758216	.022537	7.80	0.000	.1316499	.2199932
c.w#c.d4#c.f2016	.1849706	.0249782	7.41	0.000	.1360143	.233927
c.w#c.d5#c.f2015	.0978163	.0413848	2.36	0.018	.0167036	.1789289
c.w#c.d5#c.f2016	.1327046	.0434477	3.05	0.002	.0475486	.2178606
c.w#c.d6#c.f2016	.092621	.0624522	1.48	0.138	-.0297831	.2150251
c.w#c.d4#c.f2014#c.x1_dm4	-.0189186	.0162029	-1.17	0.243	-.0506757	.0128384
c.w#c.d4#c.f2015#c.x1_dm4	-.0392521	.0176407	-2.23	0.026	-.0738273	-.004677
c.w#c.d4#c.f2016#c.x1_dm4	-.0246808	.0215565	-1.14	0.252	-.0669309	.0175692
c.w#c.d5#c.f2015#c.x1_dm5	.0001442	.0238144	0.01	0.995	-.0465311	.0468195

c.w#c.d5#c.f2016#c.x1_dm5	-.039724	.0213054	-1.86	0.062	-.0814819	.0020339
c.w#c.d6#c.f2016#c.x1_dm6	-.0403321	.0322365	-1.25	0.211	-.1035145	.0228503
f2014	.0454511	.0708177	0.64	0.521	-.0933491	.1842513
f2015	-.0158952	.0784413	-0.20	0.839	-.1696373	.1378469
f2016	-.0474077	.079603	-0.60	0.551	-.2034268	.1086113
c.f2014#c.x1	.0015447	.0060817	0.25	0.800	-.0103752	.0134646
c.f2015#c.x1	-.0012545	.0065617	-0.19	0.848	-.0141151	.0116062
c.f2016#c.x1	.0091707	.0066336	1.38	0.167	-.003831	.0221724
d4	-1.356329	.7810326	-1.74	0.082	-2.887125	.1744666
d5	1.191999	1.144603	1.04	0.298	-1.051383	3.43538
d6	1.690494	1.679344	1.01	0.314	-1.60096	4.981949
x1	.0682135	.025351	2.69	0.007	.0185266	.1179005
c.d4#c.x1	.0872358	.0672249	1.30	0.194	-.0445226	.2189943
c.d5#c.x1	-.1026349	.1014854	-1.01	0.312	-.3015426	.0962728
c.d6#c.x1	-.1394164	.1293121	-1.08	0.281	-.3928634	.1140306
_cons	1.689357	.3034252	5.57	0.000	1.094654	2.284059

sigma_u	.95699311					
sigma_e	.1996671					
rho	.9582852	(fraction of variance due to u_i)				

```

. * Now use margins for ATTs to account for sampling error in the covariate means.
. * The changes in standard errors tend to be minor:
.
. reg logy c.w#c.d4#c.f2014 c.w#c.d4#c.f2015 c.w#c.d4#c.f2016 ///
>      c.w#c.d5#c.f2015 c.w#c.d5#c.f2016 ///
>      c.w#c.d6#c.f2016 ///
>      c.w#c.d4#c.f2014#c.x1 c.w#c.d4#c.f2015#c.x1 c.w#c.d4#c.f2016#c.x1 ///
>      c.w#c.d5#c.f2015#c.x1 c.w#c.d5#c.f2016#c.x1 ///
>      c.w#c.d6#c.f2016#c.x1 ///
>      f2014 f2015 f2016 c.f2014#c.x1 c.f2015#c.x1 c.f2016#c.x1 ///
>      d4 d5 d6 x1 c.d4#c.x1 c.d5#c.x1 c.d6#c.x1, vce(cluster id)

```

(Std. Err. adjusted for 545 clusters in id)

logy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
c.w#c.d4#c.f2014	.4021219	.1899911	2.12	0.035	.0289157	.775328
c.w#c.d4#c.f2015	.6365949	.2113629	3.01	0.003	.2214076	1.051782
c.w#c.d4#c.f2016	.4746943	.2530328	1.88	0.061	-.0223467	.9717353
c.w#c.d5#c.f2015	.0961897	.2600982	0.37	0.712	-.4147302	.6071095
c.w#c.d5#c.f2016	.5808412	.2470379	2.35	0.019	.0955762	1.066106
c.w#c.d6#c.f2016	.5766064	.3640648	1.58	0.114	-.1385386	1.291751
c.w#c.d4#c.f2014#c.x1	-.0189186	.0162029	-1.17	0.243	-.0507465	.0129092
c.w#c.d4#c.f2015#c.x1	-.0392521	.0176407	-2.23	0.026	-.0739044	-.0045999

c.w#c.d4#c.f2016#c.x1	-.0246808	.0215565	-1.14	0.253	-.0670251	.0176634
c.w#c.d5#c.f2015#c.x1	.0001442	.0238144	0.01	0.995	-.0466352	.0469236
c.w#c.d5#c.f2016#c.x1	-.039724	.0213054	-1.86	0.063	-.081575	.002127
c.w#c.d6#c.f2016#c.x1	-.0403321	.0322365	-1.25	0.211	-.1036554	.0229912
f2014	.0454511	.0708177	0.64	0.521	-.0936586	.1845607
f2015	-.0158952	.0784413	-0.20	0.839	-.1699801	.1381897
f2016	-.0474077	.079603	-0.60	0.552	-.2037747	.1089592
c.f2014#c.x1	.0015447	.0060817	0.25	0.800	-.0104018	.0134911
c.f2015#c.x1	-.0012545	.0065617	-0.19	0.848	-.0141438	.0116348
c.f2016#c.x1	.0091707	.0066336	1.38	0.167	-.0038599	.0222014
d4	-1.356329	.7810326	-1.74	0.083	-2.890538	.17788
d5	1.191999	1.144603	1.04	0.298	-1.056385	3.440383
d6	1.690494	1.679344	1.01	0.315	-1.6083	4.989288
x1	.0682135	.025351	2.69	0.007	.0184158	.1180113
c.d4#c.x1	.0872358	.0672249	1.30	0.195	-.0448164	.2192881
c.d5#c.x1	-.1026349	.1014854	-1.01	0.312	-.3019862	.0967163
c.d6#c.x1	-.1394164	.1293121	-1.08	0.281	-.3934286	.1145958
_cons	1.689357	.3034252	5.57	0.000	1.093328	2.285385

```
. margins, dydx(w) at(d4 = 1 d5 = 0 d6 = 0 f2014 = 1 f2015 = 0 f2016 = 0) ///
> subpop(if d4 == 1) vce(uncond)
```

```
Average marginal effects      Number of obs      =      3,270
                               Subpop. no. obs      =      804
```

```
Expression   : Linear prediction, predict()
dy/dx w.r.t. : w
at           : d4              =          1
              f2014            =          1
              f2015            =          0
              f2016            =          0
              d5               =          0
              d6               =          0
```

(Std. Err. adjusted for 545 clusters in id)

		Unconditional				
	dy/dx	Std. Err.	t	P> t	[95% Conf. Interval]	
w	.1800395	.0223999	8.04	0.000	.1360386	.2240404

```
. margins, dydx(w) at(d4 = 1 d5 = 0 d6 = 0 f2014 = 0 f2015 = 1 f2016 = 0) ///
> subpop(if d4 == 1) vce(uncond)
```

```
Average marginal effects      Number of obs      =      3,270
                               Subpop. no. obs      =      804
```

```
Expression   : Linear prediction, predict()
dy/dx w.r.t. : w
at           : d4              =          1
              f2014            =          0
```

f2015	=	1
f2016	=	0
d5	=	0
d6	=	0

(Std. Err. adjusted for 545 clusters in id)

	Unconditional					
	dy/dx	Std. Err.	t	P> t	[95% Conf. Interval]	
w	.1758216	.0229027	7.68	0.000	.1308329	.2208102

```
. margins, dydx(w) at(d4 = 1 d5 = 0 d6 = 0 f2014 = 0 f2015 = 0 f2016 = 1) ///
> subpop(if d4 == 1) vce(uncond)
```

Average marginal effects

Number of obs	=	3,270
Subpop. no. obs	=	804

```
Expression : Linear prediction, predict()
dy/dx w.r.t. : w
at : d4 = 1
    f2014 = 0
    f2015 = 0
    f2016 = 1
    d5 = 0
    d6 = 0
```

(Std. Err. adjusted for 545 clusters in id)

	Unconditional					
	dy/dx	Std. Err.	t	P> t	[95% Conf. Interval]	

w		.1849706	.0251094	7.37	0.000	.1356474	.2342938
---	--	----------	----------	------	-------	----------	----------

```
. margins, dydx(w) at(d4 = 0 d5 = 1 d6 = 0 f2014 = 0 f2015 = 1 f2016 = 0) ///
> subpop(if d5 == 1) vce(uncond)
```

Average marginal effects	Number of obs	=	3,270
	Subpop. no. obs	=	192

Expression : Linear prediction, predict()

dy/dx w.r.t. : w

at	:	d4	=	0
		f2014	=	0
		f2015	=	1
		f2016	=	0
		d5	=	1
		d6	=	0

(Std. Err. adjusted for 545 clusters in id)

		Unconditional				[95% Conf. Interval]	
		dy/dx	Std. Err.	t	P> t		
w		.0978163	.0413848	2.36	0.018	.0165227	.1791098

```
. margins, dydx(w) at(d4 = 0 d5 = 1 d6 = 0 f2014 = 0 f2015 = 0 f2016 = 1) ///
> subpop(if d5 == 1) vce(uncond)
```

Average marginal effects	Number of obs	=	3,270
	Subpop. no. obs	=	192

Expression : Linear prediction, predict()

```

dy/dx w.r.t. : w
at           : d4           =           0
              f2014         =           0
              f2015         =           0
              f2016         =           1
              d5             =           1
              d6             =           0

```

(Std. Err. adjusted for 545 clusters in id)

	Unconditional				
	dy/dx	Std. Err.	t	P> t	[95% Conf. Interval]
w	.1327046	.0447612	2.96	0.003	.0447787 .2206305

```

. margins, dydx(w) at(d4 = 0 d5 = 0 d6 = 1 f2014 = 0 f2015 = 0 f2016 = 1) ///
>      subpop(if d6 == 1) vce(uncond)

```

```

Average marginal effects      Number of obs      =      3,270
                               Subpop. no. obs     =      102

```

```

Expression : Linear prediction, predict()
dy/dx w.r.t. : w
at           : d4           =           0
              f2014         =           0
              f2015         =           0
              f2016         =           1
              d5             =           0
              d6             =           1

```

(Std. Err. adjusted for 545 clusters in id)

		Unconditional				
	dy/dx	Std. Err.	t	P> t	[95% Conf. Interval]	
w	.092621	.065386	1.42	0.157	-.0358189	.2210609

```

. * Now exponential model for y. Without covariates, FE Poisson and
. * pooled Poisson are identical. Adding the full set of year dummies
. * does not change the estimates.
.
. xtpoisson y c.d4#c.f2014 c.d4#c.f2015 c.d4#c.f2016 ///
>      c.d5#c.f2015 c.d5#c.f2016 ///
>      c.d6#c.f2016 ///
>      f2014 f2015 f2016, fe vce(robust)
note: you are responsible for interpretation of non-count dep. variable

```

(Std. Err. adjusted for clustering on id)

y	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
c.d4#c.f2014	.1942707	.0320979	6.05	0.000	.1313599	.2571814
c.d4#c.f2015	.1708469	.0339736	5.03	0.000	.1042598	.237434
c.d4#c.f2016	.2224609	.0299002	7.44	0.000	.1638575	.2810643
c.d5#c.f2015	.0589306	.0440037	1.34	0.180	-.0273152	.1451763
c.d5#c.f2016	.2043645	.0548574	3.73	0.000	.0968459	.3118831
c.d6#c.f2016	.1179841	.1654163	0.71	0.476	-.2062258	.4421941
f2014	.045545	.0198769	2.29	0.022	.006587	.084503
f2015	-.0547305	.0170008	-3.22	0.001	-.0880515	-.0214096
f2016	.040442	.0188927	2.14	0.032	.0034131	.077471

```
. poisson y c.d4#c.f2014 c.d4#c.f2015 c.d4#c.f2016 ///
>          c.d5#c.f2015 c.d5#c.f2016 ///
>          c.d6#c.f2016 ///
>          f2014 f2015 f2016 d4 d5 d6, vce(cluster id)
note: you are responsible for interpretation of noncount dep. variable
```

(Std. Err. adjusted for 545 clusters in id)

y	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
c.d4#c.f2014	.1942707	.0321274	6.05	0.000	.1313022	.2572392
c.d4#c.f2015	.1708469	.0340048	5.02	0.000	.1041986	.2374952
c.d4#c.f2016	.2224609	.0299277	7.43	0.000	.1638036	.2811181
c.d5#c.f2015	.0589306	.0440442	1.34	0.181	-.0273944	.1452556
c.d5#c.f2016	.2043645	.0549078	3.72	0.000	.0967471	.3119818
c.d6#c.f2016	.1179841	.1655683	0.71	0.476	-.2065237	.4424919
f2014	.045545	.0198951	2.29	0.022	.0065512	.0845388
f2015	-.0547305	.0170164	-3.22	0.001	-.0880821	-.021379
f2016	.040442	.01891	2.14	0.032	.003379	.077505
d4	-.4453415	.1198097	-3.72	0.000	-.6801642	-.2105188
d5	.0809309	.201272	0.40	0.688	-.313555	.4754169
d6	.2727917	.3909351	0.70	0.485	-.493427	1.03901
_cons	3.000529	.0623774	48.10	0.000	2.878272	3.122787

```
. xtpoisson y c.d4#c.f2014 c.d4#c.f2015 c.d4#c.f2016 ///
> c.d5#c.f2015 c.d5#c.f2016 ///
> c.d6#c.f2016 ///
> i.year, fe vce(robust)
note: you are responsible for interpretation of non-count dep. variable
```

(Std. Err. adjusted for clustering on id)

y	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
c.d4#c.f2014	.1942707	.0320979	6.05	0.000	.1313599	.2571814
c.d4#c.f2015	.1708469	.0339736	5.03	0.000	.1042598	.237434
c.d4#c.f2016	.2224609	.0299002	7.44	0.000	.1638575	.2810643
c.d5#c.f2015	.0589306	.0440037	1.34	0.180	-.0273152	.1451763
c.d5#c.f2016	.2043645	.0548574	3.73	0.000	.0968459	.3118831
c.d6#c.f2016	.1179841	.1654163	0.71	0.476	-.2062258	.4421941
year						
2012	.0306881	.0156213	1.96	0.049	.000071	.0613052
2013	.0580666	.0168517	3.45	0.001	.0250379	.0910953
2014	.075411	.0206204	3.66	0.000	.0349958	.1158263
2015	-.0248645	.0203301	-1.22	0.221	-.0647108	.0149818
2016	.070308	.0187326	3.75	0.000	.0335928	.1070233

```

. * Now with a covariate. The results are not the same, but the differences
. * in the estimated ATTs are minor in this application.
. * Also, the results change with full time dummies and full
. * of time dummy interactions with covariates.
.
. xtpoisson y c.w#c.d4#c.f2014 c.w#c.d4#c.f2015 c.w#c.d4#c.f2016 ///
> c.w#c.d5#c.f2015 c.w#c.d5#c.f2016 ///
> c.w#c.d6#c.f2016 ///
> c.w#c.d4#c.f2014#c.x1_dm4 c.w#c.d4#c.f2015#c.x1_dm4 c.w#c.d4#c.f2016#c.x1_dm4 /
> c.w#c.d5#c.f2015#c.x1_dm5 c.w#c.d5#c.f2016#c.x1_dm5 ///
> c.w#c.d6#c.f2016#c.x1_dm6 ///
> f2014 f2015 f2016 c.f2014#c.x1 c.f2015#c.x1 c.f2016#c.x1, fe vce(robust)
note: you are responsible for interpretation of non-count dep. variable

```

(Std. Err. adjusted for clustering on id)

y	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval	
c.w#c.d4#c.f2014	.1966912	.0320517	6.14	0.000	.1338709	.2595114
c.w#c.d4#c.f2015	.181199	.0330198	5.49	0.000	.1164814	.2459165
c.w#c.d4#c.f2016	.2307054	.0297962	7.74	0.000	.172306	.2891048
c.w#c.d5#c.f2015	.0550393	.0431379	1.28	0.202	-.0295093	.139588
c.w#c.d5#c.f2016	.2142023	.0527788	4.06	0.000	.1107577	.3176469
c.w#c.d6#c.f2016	.0219704	.1270021	0.17	0.863	-.2269492	.27089
c.w#c.d4#c.f2014#c.x1_dm4	-.012003	.0214177	-0.56	0.575	-.0539809	.0299749

c.w#c.d4#c.f2015#c.x1_dm4	-.0602142	.0202287	-2.98	0.003	-.0998616	-.0205667
c.w#c.d4#c.f2016#c.x1_dm4	-.0347854	.029333	-1.19	0.236	-.0922771	.0227062
c.w#c.d5#c.f2015#c.x1_dm5	-.0194663	.0249223	-0.78	0.435	-.0683132	.0293806
c.w#c.d5#c.f2016#c.x1_dm5	-.0460565	.0245187	-1.88	0.060	-.0941123	.0019993
c.w#c.d6#c.f2016#c.x1_dm6	-.1263765	.0500253	-2.53	0.012	-.2244244	-.0283287
f2014	.041844	.1302967	0.32	0.748	-.2135328	.2972208
f2015	-.001671	.1172639	-0.01	0.989	-.231504	.2281621
f2016	-.1416546	.117782	-1.20	0.229	-.372503	.0891939
c.f2014#c.x1	.0003109	.0107946	0.03	0.977	-.0208461	.0214678
c.f2015#c.x1	-.0044377	.009647	-0.46	0.646	-.0233455	.01447
c.f2016#c.x1	.0151421	.0094951	1.59	0.111	-.0034679	.0337521

```
. poisson y c.w#c.d4#c.f2014 c.w#c.d4#c.f2015 c.w#c.d4#c.f2016 ///
> c.w#c.d5#c.f2015 c.w#c.d5#c.f2016 ///
> c.w#c.d6#c.f2016 ///
> c.w#c.d4#c.f2014#c.x1_dm4 c.w#c.d4#c.f2015#c.x1_dm4 c.w#c.d4#c.f2016#c.x1_dm4 /
> c.w#c.d5#c.f2015#c.x1_dm5 c.w#c.d5#c.f2016#c.x1_dm5 ///
> c.w#c.d6#c.f2016#c.x1_dm6 ///
> f2014 f2015 f2016 c.f2014#c.x1 c.f2015#c.x1 c.f2016#c.x1 ///
> d4 d5 d6 x1 c.d4#c.x1 c.d5#c.x1 c.d6#c.x1, vce(cluster id)
note: you are responsible for interpretation of noncount dep. variable
```

(Std. Err. adjusted for 545 clusters in id)

y	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval	
c.w#c.d4#c.f2014	.1960423	.0321189	6.10	0.000	.1330905	.2589942
c.w#c.d4#c.f2015	.1782379	.0335015	5.32	0.000	.1125762	.2438995
c.w#c.d4#c.f2016	.2291282	.0297649	7.70	0.000	.1707902	.2874663
c.w#c.d5#c.f2015	.0552965	.0429822	1.29	0.198	-.0289471	.1395402
c.w#c.d5#c.f2016	.2127154	.0526125	4.04	0.000	.1095968	.3158339
c.w#c.d6#c.f2016	.0250988	.1058216	0.24	0.813	-.1823077	.2325053
c.w#c.d4#c.f2014#c.x1_dm4	-.0087947	.0166791	-0.53	0.598	-.0414852	.0238957
c.w#c.d4#c.f2015#c.x1_dm4	-.043178	.017645	-2.45	0.014	-.0777615	-.0085944
c.w#c.d4#c.f2016#c.x1_dm4	-.0276426	.0224272	-1.23	0.218	-.0715992	.0163139

c.w#c.d5#c.f2015#c.x1_dm5	-.0226903	.0298553	-0.76	0.447	-.0812055	.035825
c.w#c.d5#c.f2016#c.x1_dm5	-.0476805	.0244321	-1.95	0.051	-.0955665	.0002055
c.w#c.d6#c.f2016#c.x1_dm6	-.1207353	.0572943	-2.11	0.035	-.2330301	-.0084405
f2014	.0422137	.117699	0.36	0.720	-.1884721	.2728996
f2015	-.0077869	.1054262	-0.07	0.941	-.2144186	.1988447
f2016	-.1199528	.1051667	-1.14	0.254	-.3260757	.08617
c.f2014#c.x1	.0002798	.0097223	0.03	0.977	-.0187755	.0193351
c.f2015#c.x1	-.0039262	.0086462	-0.45	0.650	-.0208726	.0130201
c.f2016#c.x1	.0133375	.0084839	1.57	0.116	-.0032906	.0299656
d4	-1.525272	.6487306	-2.35	0.019	-2.79676	-.2537829
d5	.751651	.9702689	0.77	0.439	-1.150041	2.653343
d6	3.010725	1.848229	1.63	0.103	-.6117372	6.633187
x1	.0549107	.026275	2.09	0.037	.0034127	.1064087
c.d4#c.x1	.0914792	.0539856	1.69	0.090	-.0143307	.1972891
c.d5#c.x1	-.0564077	.0864516	-0.65	0.514	-.2258498	.1130344
c.d6#c.x1	-.2337497	.1524969	-1.53	0.125	-.532638	.0651387
_cons	2.34674	.3263905	7.19	0.000	1.707027	2.986454