Tutorial 4: Difference-in-Differences

Alexander Wintzéus*

February 9, 2024

In this tutorial, we will discuss Difference-in-Differences (DiD). To introduce the implementation of related methods in `Stata`, we will be relying on an empirical study of Wolfers (2006). As before, the exercises are collected in the first part of this document. Theoretical concepts are reviewed in the second part.

1 Application

The question posed by Wolfers (2006) is straightforward: Did unilateral divorce laws increase divorce rates in the United States? Between 1968 and 1988, 29 U.S. states abandoned their mutual consent divorce regime in favor of a unilateral divorce regime, allowing one spouse to terminate the marriage without the consent of the other. Over this period, divorce rates in the U.S. increased dramatically, having left researchers and policy makers pondering whether these two trends are connected.

To investigate this question, Friedberg (1998) used comprehensive administrative divorce data from a state-based panel. Previous attempts to answer the question, which largely concluded that divorce rates were left unaffected by the divorce reforms, may have been suffering from an endogeneity problem: States with historically high divorce rates were the first ones to adopt unilateral divorce regimes. By exploiting the panel structure of her data, Friedberg (1998) tried to overcome these concerns. Using a simple Difference-in-Differences design, she finds that the adoption of unilateral divorce laws could explain about one-sixth of the rise in divorce rates since the late sixties. However, as pointed out by Wolfers (2006), the conclusions of Friedberg (1998) may be somewhat misleading as her approach may be confounding the dynamic effects of unilateral divorce regime adoption and preexisting trends. To this end, the author proposes a modified event-study design.

Load the dataset `Wolfers2006-AER.dta` into `Stata` and get to know its contents by using the `browse` and `describe` commands.

1. In a first step, we will plot the average evolution of `divrate` (number of divorcees per 1000 inhabitants) over the length of the panel. The panel starts in 1956 and ends in 1998.

*KU Leuven, Department of Economics. alexander.wintzeus@kuleuven.be

(a) Compute the average of `divrate` for each `year` using the `collpase` command. Supply analytic state population weights collected in the variable `stpop`. Plot the evolution of the mean `divrate` using the `twoway connect` command. Make use of the `xline` option to plot vertical lines at the years 1968 (start of reform period) and 1988 (end of reform period). Discuss the result.

(b) The variable `reform` indicates whether a state ever adopts unilateral divorce laws. In the same graph, plot the evolution of the mean `divrate` using the `twoway connect` command for states that never adopt a unilateral divorce regime and states that ever adopt such a regime, respectively. What do you learn from this graph?

(c) Create a numerical variable `state` from the string variable `st` using the `encode` command. Create a variable `yearsuni` containing the number of in-sample years the state has unilateral divorce laws by using the `egen` command with the `total` function and `bysort` prefix. Based on this variable, create a dummy that indicates if a state has adopted unilateral divorce laws before 1974.

(d) In the same graph, plot the evolution of the mean `divrate` using the `twoway connect` command for states that never adopt a unilateral divorce regime, states that do adopt such a regime before 1974, and states that adopt such a regime from 1974 onwards. Do you think the concern that early adopter of unilateral divorce laws had higher initial divorce rates is valid?

2. Consider the following linear model proposed by Friedberg (1998):[1]

$$y_{s,t} = \sum_{s'=1}^{S} \alpha_{s'} \mathbb{1}\{s = s'\} + \sum_{t'=1}^{T} \gamma_{t'} \mathbb{1}\{t = t'\} + \sum_{s'=1}^{S} \delta_{s'} \mathbb{1}\{s = s'\} t + \beta D_{s,t} + \varepsilon_{s,t} \tag{1}$$

where $y_{s,t}$ denotes the divorce rate in state s in period t and $D_{s,t}$ is a dummy indicating whether state s adopted a unilateral divorce law regime in year t. Here, α_s and γ_t denote state and year fixed effects, respectively. Finally, the δ_s represent state-specific linear time trends.

(a) Create a dummy variable `window` equal to one if `year` is in the reform window from 1968 to 1988. Estimate equation (1) by pooled OLS using the `regress` command on the sample restricted by the `window` variable. You can ignore the state-specific time trends for now. Supply analytic state population weights collected in the variable `stpop`. Cluster standard errors at the state level by providing the `vce(cluster state)` option. What do the results suggest?

(b) Estimate equation (1) by pooled OLS using the `regress` command on the sample restricted by the `window` variable. This time, include the state-specific time trends. Supply

[1] This model does not entirely correspond to the one proposed by Friedberg (1998). In her specification, she also includes state-specific quadratic trends. However, for sake of simplicity, we will omit these.

analytic state population weights collected in the variable `stpop`. Cluster standard errors at the state level by providing the `vce(cluster state)` option. Why may it be important to include state-specific time trends? How do the results change compared to the previous exercise?

(c) The pooled OLS estimate of the parameter β obtained in the previous exercises corresponds to the associated Least Squares Dummy Variables (LSDV) estimator, which, in turn, is equivalent to the Two-Way Fixed Effects (TWFE) estimator. To verify this, one could in principle use `Stata`'s built-in `xtreg` command after communicating to `Stata` through the `xtset` command that the data in memory should be interpreted as panel data. A downside of this command is that it can only handle one-way fixed effects, so one still has to include year dummies. On top of this, `xtreg` cannot handle weights that vary across states. Hence, for the purpose of this tutorial, we cannot rely on `xtreg`.

3. To disentangle the dynamic effects of unilateral divorce regime adoption and preexisting time trends in divorce rates, Wolfers (2006) proposes the following event-study specification:

$$y_{s,t} = \sum_{s'=1}^{S} \alpha_{s'} \mathbb{1}\{s = s'\} + \sum_{t'=1}^{T} \gamma_{t'} \mathbb{1}\{t = t'\} + \sum_{s'=1}^{S} \delta_{s'} \mathbb{1}\{s = s'\} t + \sum_{\ell=-K, \ell \neq 0}^{L} \beta_\ell \mathbb{1}\{t = F_s - 1 + \ell\} + \varepsilon_{s,t}. \quad (2)$$

In this specification, β_ℓ is supposed to capture the cumulative effect on the divorce rate of having adopted unilateral divorce laws for ℓ periods. As such, it aims to capture the dynamic effects of divorce law adoption.

(a) Assuming that $K = 9$ and $L = 16$, construct relative-time indicators `rel_time_l` for each period ℓ in the event window. Make use of endpoint binning to allow for constant effects outside the event window. To this end, start by constructing a variable `rel_divlaw` using the `generate` command denoting the number of periods relative to the period prior to treatment onset for each state-year observation.

(b) Estimate equation (2) by pooled OLS using the `regress` command.[2] Supply analytic state population weights collected in the variable `stpop`. Cluster standard errors at the state level by providing the `vce(cluster state)` option. Store the results using the `estimates store` command. Interpret the results.

(c) Construct a joint test for the no-anticipation and parallel-trends assumption using the `testparm` command. This corresponds to testing whether the placebo estimates are jointly significant. What do the results suggest?

(d) Use the `coefplot` command to make an event study plot based on stored regression coefficients. Make use of the `keep(rel_time*)` option to only plot the event study coefficients. Supply the `vertical` option to plot the coefficients from left to right rather than top to bottom.

[2] Do not forget to exclude the relative-time indicator for $\ell = 0$.

2 Theoretical summary

In this section, we will summarize the main theoretical concepts regarding Difference-in-Differences (DiD). In the first part, we will briefly discuss panel methods from the perspective of the traditional model-based approach. In the second part, we shift our perspective towards the design-based approach. We will adapt our potential outcomes notation to allow for possible dynamic treatment effects and look if and how our objects of interest can be estimated using the methods discussed in the first part. To shape the discussion, we will consider two prominent research designs with binary treatments.

2.1 Panel methods

A panel data set has multiple observations, typically over time, for a number of cross-sectional units. Panel data thus have two dimensions, one for the cross-sectional units i and one for observations or time periods t. Often, cross-sectional units are referred to as groups; as will be the case in the dynamic potential outcomes framework introduced below.

Traditionally, the discussion surrounding estimation in panel data settings revolves around the *error-components model*. As before, we will assume that the relationship between the dependent variable y_{it} and the independent variables $\mathbf{x}_{it}$ is linear. That is, for all $i = 1, \dots, n$ and all $t = 1, \dots, T$:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \tag{3}$$

In the error-components model, the error term is usually thought of as consisting of two parts:

$$\varepsilon_{it} = \alpha_i + \eta_{it}. \tag{4}$$

The first unobservable component α_i is called the individual fixed effect as it is a component common to individual i across all periods t. In this particular setting, the usual orthogonality condition that the regressors $\mathbf{x}_{it}$ are orthogonal to the errors ε_{it} will be satisfied if both

$$E[\mathbf{x}_{it}\alpha_i] = 0 \tag{5}$$

and

$$E[\mathbf{x}_{it}\eta_{it}] = 0 \tag{6}$$

hold for all i and t. However, in most panel data settings, assuming that the regressors are orthogonal to the individual fixed effects is unreasonable. Hence, pooled OLS estimation would generally fail to provide unbiased and consistent estimates of the parameter vector $\boldsymbol{\beta}$. A popular alternative estimator that is robust to failures of the orthogonality condition with respect to the individual fixed effects is known as the One-Way Fixed Effects (OWFE) estimator. The One-Way Fixed Effects estimator is a pooled OLS estimator defined on transformed data:

$$\boldsymbol{\beta}_{OWFE} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} \tag{7}$$

where $\tilde{\mathbf{y}}$ is the $nT \times 1$ pooled vector of demeaned values for the dependent variable and $\tilde{\mathbf{X}}$ is the $nT \times K$ pooled matrix of demeaned values for the independent variables. For example, the transformed dependent variable is defined as

$$\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}. \tag{8}$$

This estimator, although partly robust to failure of the orthogonality condition, is not able to identify some of the unknown parameters of the model. That is, any unit-specific time-invariant characteristic will be removed from the regression specification after transforming the data. As such, the OWFE cannot identify the effect of such characteristics on the dependent variable.

The traditional error-component model, as it was defined above, can more accurately be referred to as the one-way error-component model. As this name suggests, it is possible to think of even more involved error structures. For example, the error term could consists of three parts:

$$\varepsilon_{it} = \alpha_i + \gamma_t + \eta_{it}. \tag{9}$$

This error structure corresponds to the two-way error-components model. As before, α_i is an individual fixed effect. However, the error-structure now includes γ_t – a time or period fixed effect. In this setting, the usual orthogonality condition implies the following *additional* condition:

$$E[\mathbf{x}_{it}\gamma_t] = 0 \tag{10}$$

for all i and t. Again, there may be a number of panel data settings where we believe this assumption to be unreasonable. Nevertheless, an alternative estimator, known as the Two-Way Fixed Effects (TWFE) estimator, exists that is also robust against failure of this additional assumption:

$$\boldsymbol{\beta}_{TWFE} = (\check{\mathbf{X}}'\check{\mathbf{X}})^{-1}\check{\mathbf{X}}'\check{\mathbf{y}} \tag{11}$$

where $\check{\mathbf{y}}$ is the $nT \times 1$ pooled vector of transformed values for the dependent variable and $\check{\mathbf{X}}$ is the $nT \times K$ pooled matrix of transformed values for the independent variables. For example, the transformed dependent variable is defined as

$$\check{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} - \frac{1}{n}\sum_{i=1}^{n} y_{it} + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} y_{it}. \tag{12}$$

As with the OWFE estimator, the effect of any unit-specific time-invariant characteristic on the dependent variable cannot be identified by TWFE estimator. On top of this, the TWFE estimator is not able to capture the effect of any period-specific unit-invariant characteristic. Still, it is the most commonly used estimator in Difference-in-Differences (DiD) designs.

Finally, it is worthwhile to mention that both the One-Way Fixed Effects (OWFE) and Two-Way Fixed Effects (TWFE) estimator are equivalent to a particular Least Squares Dummy Variables (LSDV) estimator. The OWFE estimate of the parameter vector $\boldsymbol{\beta}$ is equivalent to the one obtained by performing pooled OLS on the following equation:

$$y_{it} = \sum_{j=1}^{n} \alpha_j \mathbb{1}\{i = j\} + \mathbf{x}_{it}'\boldsymbol{\beta} + \eta_{it} \tag{13}$$

where $\mathbb{1}\{.\}$ denotes the indicator function. Similarly, the TWFE estimate of the parameter vector $\boldsymbol{\beta}$ is equivalent to the one obtained by performing pooled OLS on the following equation:

$$y_{it} = \sum_{j=1}^{n} \alpha_j \mathbb{1}\{i = j\} + \sum_{r=1}^{T} \gamma_r \mathbb{1}\{t = r\} + \mathbf{x}_{it}'\boldsymbol{\beta} + \eta_{it}. \tag{14}$$

Computationally, LSDV is generally more intensive as the dimensionality increases fast with the number of included fixed effects.

2.2 Binary designs

For the exposition in this section, we will closely follow the notation and content presented de Chaisemartin and D'Haultfoeuille (2023). We are interested in estimating the effect of a treatment on an outcome. To this end, we have at our disposal a panel of G groups observed at T periods, indexed by g and t respectively.[3] Throughout, we will assume that the panel data set at hand is balanced.

Let $D_{g,t}$ denote the treatment of group g at period t. We will assume that treatment is assigned at the (g,t) level and is binary $D_{g,t} \in \{0,1\}$. Let $\mathbf{D}_g = (D_{g,1}, \ldots, D_{g,T})$ be the $1 \times T$ vector of treatments for group g from period 1 to T and let $\mathbf{D} = (\mathbf{D}_1, \ldots, \mathbf{D}_G)$ be the vector of treatments of all groups in all time periods. $\mathbf{D}$ is then referred to as the design of a study. As in the static potential outcomes framework, we will assume that the potential outcomes of a given group g do not depend on the treatments of the other groups. Hence, if $(d_1, \ldots, d_T) \in \{0,1\}^T$ denotes a particular sequence of treatments for group g in every time period t, then $Y_{g,t}(d_1, \ldots, d_T)$ will denote its associated potential outcomes. Finally, let $Y_{g,t}$ denote the observed outcome of group g at time t.

In general, the dynamic potential outcomes framework thus explicitly allows a group's outcome in

[3] In principle, one could think of the groups g to be units such as individuals or firms. However, the setup with groups or locations is slightly more general.

a given period t to depend not only on its current treatment, but on its past and future treatments as well. However, in order to achieve identification of treatment effects, we will have to make some restrictions on how current outcomes can be affected by past and future treatments:

1. **No anticipation:** For all groups g and all possible values of treatments from period 1 to T $(d_1, \ldots, d_T) \in \{0,1\}^T$:

$$Y_{g,t}(d_1, \ldots, d_T) = Y_{g,t}(d_1, \ldots, d_t). \tag{15}$$

   This so-called no-anticipation hypothesis requires that a group's current outcome does not depend on its future treatments – an assumption that we will maintain throughout the rest of this exposition.

2. **No dynamic effects:** For all groups g and all possible values of treatments from period 1 to t $(d_1, \ldots, d_t) \in \{0,1\}^t$:

$$Y_{g,t}(d_1, \ldots, d_t) = Y_{g,t}(d_t). \tag{16}$$

   This assumption requires that a group's current outcome does not depend on its past treatments. Together with the no-anticipation assumption, this implies that the dynamic potential outcomes framework with binary treatments reduces to the standard static potential outcomes framework. However, the results presented below do not rely on imposing this no-dynamic-effects assumption. Hence, we will not maintain it, but simply highlight how things change or simplify in case we would.

3. **Parallel trends:** For all time periods $t \geq 2$

$$E\big[Y_{g,t}(\mathbf{0}_t) - Y_{g,t-1}(\mathbf{0}_{t-1})\big] \tag{17}$$

   does not vary across g. The parallel-trends assumption requires that every group experiences the same expected evolution in its never-treated potential outcome. Note that, if the no-dynamic-effects assumption holds, then the parallel trends assumption reduces to:

$$E\big[Y_{g,t}(0) - Y_{g,t-1}(0)\big] \tag{18}$$

   does not vary across g for all $t \geq 2$. This is the standard parallel-trends assumption in classic Difference-in-Differences settings.

2.2.1 No variation in treatment timing

The classical Difference-in-Differences (DiD) setting corresponds to a design with a binary treatment and no variation in treatment timing. Formally, in this design there exists a time period $F \geq 2$

such that for all groups g:

$$D_{g,t} = \mathbb{1}\{t \geq F\} T_g \tag{19}$$

where $T_g \in \{0,1\}$ is an indicator equal to one for treatment groups and zero for control groups. We require that there is at least one treatment group and one control group. Clearly, as F does not depend on g, all treated groups receive treatment from the same time period onwards.

Suppose that we propose the following linear two-way fixed effects model:

$$Y_{g,t} = \sum_{g'=1}^{G} \alpha_{g'} \mathbb{1}\{g = g'\} + \sum_{t'=1}^{T} \gamma_t \mathbb{1}\{t = t'\} + \beta D_{g,t} + \varepsilon_{g,t}. \tag{20}$$

As noted above, the pooled OLS estimate of the parameter β corresponds to the Two-way Fixed Effects (TWFE) estimate. In a design with a binary treatment $D_{g,t}$ and no variation in treatment timing, it can be shown that this TWFE estimator β_{TWFE} is a simple Difference-in-Differences estimator:

$$\beta_{TWFE} = \left(\frac{1}{G_1 T_1} \sum_{g:T_g=1, t \geq F}^{T} Y_{g,t} - \frac{1}{G_1 T_0} \sum_{g:T_g=1, t<F}^{T} Y_{g,t} \right) - \left(\frac{1}{G_0 T_1} \sum_{g:T_g=0, t \geq F}^{T} Y_{g,t} - \frac{1}{G_0 T_0} \sum_{g:T_g=0, t<F}^{T} Y_{g,t} \right) \tag{21}$$

where G_0 and G_1 denote the number of control and treatment groups, and T_0 and T_1 denote the number of untreated and treated periods, respectively. A natural target parameter in this setting, is the Average Treatment Effect on the Treated (ATT):

$$ATT = \frac{1}{G_1 T_1} \sum_{(g,t):D_{g,t}=1} E[Y_{g,t}(\mathbf{0}_{F-1}, \mathbf{1}_{t-F+1}) - Y_{g,t}(\mathbf{0}_t)]. \tag{22}$$

That is, the average expected effect of having been treated rather than untreated for $t - F + 1$ periods, across all treated groups. If the no-dynamic-effects assumption holds, the term within the summation simplifies to the expected effect of having been treated rather than untreated in period t. Finally, if the no-anticipation and parallel-trends assumptions hold, it can readily be verified that the TWFE estimator is unbiased for the Average Treatment Effect on the Treated (ATT):

$$E[\beta_{TWFE}] = ATT. \tag{23}$$

In designs with a binary treatment and no variation in treatment timing, researches have often estimated so-called two-way fixed effects event-study regressions:

$$Y_{g,t} = \alpha_0 + \alpha_1 T_g + \sum_{t'=1, t' \neq F-1}^{T} \gamma_t \mathbb{1}\{t = t'\} + \sum_{\ell=-F+2, \ell \neq 0}^{T-F+1} \beta^{\ell} \mathbb{1}\{t = F - 1 + \ell\} T_g + \varepsilon_{g,t}. \tag{24}$$

Again, it can readily be verified that for all $\ell \neq 0$, the TWFE estimator β^{ℓ}_{TWFE} is a simple DiD estimator

$$\beta^{\ell}_{TWFE} = \frac{1}{G_1}\sum_{g:T_g=1}^{T}(Y_{g,F-1+\ell} - Y_{g,F-1}) - \frac{1}{G_0}\sum_{g:T_g=0}^{T}(Y_{g,F-1+\ell} - Y_{g,F-1}). \tag{25}$$

comparing the outcome evolution from period $F-1$ to period $F-1+\ell$ in treatment and control groups. Note that, for all ℓ, all DiD estimators are relative to period $F-1$; the period before treatment onset. For all $\ell \geq 1$, the DiD estimator compares future periods to past periods, whereas for all $\ell \leq -1$, the DiD estimator compares past to future periods. For $\ell \leq -1$, the TWFE estimator β^{ℓ}_{TWFE} is often referred to as the pre-trends or placebo estimator.

An advantage of using event-study regressions is that it allows for the estimation of dynamic treatment effects. Moreover, the pre-trends or placebo estimators allow us to partly test the no-anticipation and parallel-trends assumptions. In this setting, one can consider a set of target parameters. For all $\ell = 1, \ldots, T-F+1$

$$ATT_{\ell} = \frac{1}{G_1}\sum_{g:T_g=1} E[Y_{g,F-1+\ell}(\mathbf{0}_{F-1}, \mathbf{1}_{\ell}) - Y_{g,F-1+\ell}(\mathbf{0}_{F-1})]. \tag{26}$$

That is, we can consider the average expected effect of having been treated for ℓ periods at period $F-1+\ell$, across all treated groups. Furthermore, note that

$$ATT = \frac{1}{T-F+1}\sum_{\ell=1}^{T-F+1} ATT_{\ell}. \tag{27}$$

In this design with a binary treatment and no variation in treatment timing, a very powerful theorem holds if the no-anticipation and parallel-trends assumptions are satisified: For all $\ell = 1, \ldots, T-F+1$

$$E[\beta^{\ell}_{TWFE}] = ATT_{\ell} \tag{28}$$

and for all $\ell = -1, \ldots, -F+2$ with $F \geq 3$

$$E[\beta^{\ell}_{TWFE}] = 0. \tag{29}$$

If we have at least one pre-period, the second part of this theorem constitutes a test of the null hypothesis that the no-anticipation and parallel-trends assumptions hold. If we can reject the null hypothesis that the coefficients β^{ℓ}_{TWFE} are jointly equal to zero for $\ell \leq -1$, then we can reject that the no-anticipation assumption and parallel-trends assumption hold together. Note that rejection does not necessarily imply that both assumptions are violated. Rather, it strongly suggests that at least one assumption is is. Finally, note that separately testing whether each coefficient β^{ℓ}_{TWFE} is zero would give rise to a multiple testing problem.

2.2.2 Staggered adoption designs

As discussed in the previous subsection, the TWFE estimator is a simple DiD estimator that is unbiased for the Average Treatment Effect on the Treated (ATT) if we have a binary treatment and no variation in treatment timing. Motivated by this fact, researchers have often used TWFE estimators in more complex designs. In this subsection, we will consider a design with a binary treatment and variation in treatment timing. Formally, in such a design, there exists a time period F_g for each group g so that:

$$D_{g,t} = \mathbb{1}\{t \geq F_g\}. \tag{30}$$

The period of treatment onset F_g may vary across groups and is simply set strictly greater than T for untreated groups. We require that among groups that are not treated in the first period, not all of them become treated in exactly the same period.

As it turns out, in a design with variation in treatment timing, the TWFE estimator may no longer be unbiased for the ATT even if no-anticipation and parallel-trends hold. Furthermore, TWFE event-study regressions are not robust to treatment effects being heterogenous across groups and may suffer for contamination bias. In this case, it even becomes impossible to use the placebo estimates to test for no-anticipation and parallel trends. Therefore, in a design with variation in treatment timing one should be careful with simply running a TWFE regression. Instead, one could consult one of the heterogeneity-robust DiD estimators proposed by Callaway and Sant'Anna (2021) and Sun and Abraham (2021).

References

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200-230.

de Chaisemartin, C., & D'Haultfoeuille, X. (2023). *Difference-in-differences for simple and complex natural experiments*. Available at SSRN 4487202.

Friedberg, L. (1998). Did unilateral divorce raise divorce rates? evidence from panel data. *American Economic Review*, *88*(3), 608–627.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, *225*(2), 175-199.

Wolfers, J. (2006). Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *American Economic Review*, *96*(5), 1802-1820.