

Estimation of DID models using ETWFE

In this section, we describe the Stata command `jwddid` that implements the estimation of DID models using the ETWFE estimator proposed by @Wooldridge2022. One of the main advantages of @Wooldridge2022's estimator is that by being a simple extension of the standard FE estimator, it can be easily modified and implemented to allow for other types of non-linear models. As described in @wooldridge2022, the ETWFE estimator could be used, for example, to model cases where the dependent variable is binary (logit) or count data (poisson). A second advantage of the ETWFE estimator is that the estimation of the baseline model, is transparent as it does not require the use of specialized software, except for the estimation of fixed effects models. This is in contrast with other DID estimators like the ones proposed by @callaway2021, @dechaisemartin2021, or @kirill2022, where the bulk of the model estimation is done in the background, with the user having less control and understanding on what is being estimated.

Thanks to this transparency in model specification, @yotov2023 propose a large set of recommendations for the analysis of DID models in the context of international trade and Gravity models. On this regard, we present the command `jwddid` as a flexible command in Stata that allows to consider @yotov2003 recommendations for the estimation of DID models, in the framework of trade models.

Base line model

As described in @Wooldridge2021, the baseline model for the estimation of the DID model using the ETWFE estimator is the following:

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t} D_{i,g,t} + \xi_i + \xi_t + \varepsilon_{i,t} \quad (1)$$

where $Y_{i,t}$ is the dependent variable, $D_{i,g,t}$ is a dummy that takes the value of 1 if the observation is in the treatment group g , on period t and 0 otherwise. G is a set that indicates at what time treatment started for all observations, and T is the last period of the analysis.

ξ_i and ξ_t are sets of fixed effects for the individual and time dimensions, respectively.¹ In this setup, the $\theta_{g,t}$ coefficients represent the average treatment effect that the treatment group g experiences at time t ($ATT(g, t)$). As described in @Wooldridge2021, allowing for a flexible specification of the $\theta_{g,t}$ avoids the problem of bad controls and negative weights that have been identified in the literature as potential problems in the estimation of DID models using traditional TWFE estimators.

¹Often, one can use group fixed effects instead of individual fixed effects, and would still obtain numerically identical results in the linear model case.

This command can be directly estimated with `jwddid` using the following syntax:

```
jwddid y, ivar(i) tvar(t) gvar(g)
```

Where `y` is the dependent variable, `ivar(i)` is used to identify the individual panel data dimension, `tvar(t)` identifies the time dimension, and `gvar(g)` identifies the treatment group. Specifically, for observation i , g would take the value of zero if the panel observation is never treated (within the window of the analysis), and would take a value different from zero to indicate the year that treatment started for unit i . Following standard assumptions, this specification assumes that the treatment is an absorbing status, meaning that once a unit is treated, it remains treated for the rest of the analysis.

By default, `jwddid` will estimate the baseline model Equation 1 using the `reghdfe` command (@correira), assuming clustered standard errors at `i` level. If other level is desired, the user can specify the `cluster(cvar)` option. While the command does not impose the assumption that the data is a panel, the methodology is designed to work with panel data. In case of repeated crossection, one should instead use the following syntax:

```
jwddid y, tvar(t) gvar(g) [cluster(cvar)]
```

By excluding `ivar(i)`, the command assumes data is a repeated crossection, proceeding to include group fixed effects only. The `cluster(cvar)` option is not required, but can be used to request Standard errors to be clusted at the level `cvar`.

Specifically, this command will estimate the following model:

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t} D_{i,g,t} + \xi_g + \xi_t + \varepsilon_{i,t} \quad (2)$$

This model specification makes the implicit assumption that Parallel trends are satisfied, using all never treated and not-yet treated observations as controls (not included category) for the identification of treatment effects.

If one instead wants to relax this assumption, the user can specify the option `never`:

```
jwddid y, ivar(i) tvar(t) gvar(g) never
```

Which will estimate the following model:

$$Y_{i,t} = \alpha + \sum_{g \in G} \sum_{t=t_0}^{g-1} \theta_{g,t}^{pre} D_{i,g,t} + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t}^{post} D_{i,g,t} + \xi_i + \xi_t + \varepsilon_{i,t} \quad (3)$$

This is in principle the same as strategy as the one proposed by @sunabraham2021, allowing for full heterogeneity across all groups and all relative periods. This specification is also numerically identical to the one proposed by @callaway2021, for the case where there are no covariates. In this case, the only observations that are used as controls are the ones that

were never treated. In this specification, all $\theta_{g,t}^{pre}$ can be used to test for the parallel trends assumption, and all $\theta_{g,t}^{post}$ can be used to estimate the treatment effects.

Extensions: Nonlinear models

As described in @Wooldridge2022, the standard ETWFE model described in Equation 1 or Equation 2 identifies the average treatment effect imposing a linear parallel trends assumption. However, such assumption may not be valid in cases, such as when the dependent variable follows some limited distribution. @CallawayRoth discusses a similar problem, stating that the choice of transformation of the dependent variable is crucial for the identification of the average treatment effect, and only under certain conditions would the ATT be identified for any transformation.

In this regard, @wooldridge2022 proposes that the linear ETWFE models can be adapted to allow for non-linear models, by simply imposing the linear PTA assumption only on the latent variable of the model, but not the outcome itself.

Consider the following transformation of the model defined by Equation 3:

$$E(Y_{i,t}|X, \xi_i, \xi_t) = H \left(\alpha + \sum_{g \in G} \sum_{t=t_0}^{g-1} \theta_{g,t}^{pre} D_{i,g,t} + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t}^{post} D_{i,g,t} + \xi_i + \xi_t \right) \quad (4)$$

This specification focuses on identifying the conditional expected value of the outcome of interest as function of the treatment status, and the individual and time fixed effects. If we assume the $H()$ is the identity function, we would be back to the linear model described by Equation 3. However, if we assume that $H()$ is a non-linear function, like exponential for poisson, or logistic for logit models, we could estimate the average treatment effect under different assumptions, imposing only linear PTA on the latent variable of the model.

The `jwddid` command allows the user to specify the `method()` option to estimate models described by Equation 4, where one would specify the regression model to be estimated, followed by the options associated with that model. For example, if we would be interested in estimating a poisson model, we would use the following syntax:

```
jwddid y, ivar(i) tvar(t) gvar(g) never method(poisson)
```

There are no restrictions on the type `method` one can use with the `jwddid` command, but it has not been tested with all possible models. The user should be aware that the `method()` option is passed directly for the model estimation step, and the user should be familiar with the syntax of the model being estimated.

It should be noted that when estimating non-linear models with a large number of fixed effects, one may face an incidental parameters problem. This is not generally a problem for the linear case, because the parameters of interest can be identified without explicitly estimating the fixed

effects, using, for example the within transformation. However, with the exception of poisson models, fixed effects are generally estimated raising the possibility of incidental parameters. To reduce the impact of this problem, whenever `method()` is specified `jwddid` will incorporate *group* fixed effects, instead of *individual* fixed effects.

For the linear case with balanced data, using *group* instead of *individual* fixed effects provides numerically identical results. If panel is unbalanced the results will not be identical. In such cases, the option `corr` will create additional variables that address the difference. In the case of non-linear models, the best solution is to use *group* fixed effects. However, if one is interested in poisson models, the alterantive to group fixed effects is to use `ppmlhdfe` (@correira2020).² This is the state-of-the-art estimator for poisson models with fixed effects, and it is the recommended estimator for trade analysis.

Extensions: Covariates

As described in @wooldridge, it is possible to include covariates in the model, by simply adding corrections that enable to easily identify the average treatment effect. However, following the literature on DID models, the implicit assumption is that covariates are time-invariant. `jwddid` does not impose any assumption on the covariates, but the user should be aware of the implications.

In general, when covariates are considered, the model of interest is similar to Equation 3, but adjusted for covariates:

$$\begin{aligned}
Y_{i,t} = & \alpha + \sum_{g \in G} \sum_{t=t_0}^{g-1} \theta_{g,t}^{pre} D_{i,g,t} + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t}^{post} D_{i,g,t} + \sum_{g \in G} \sum_{t=t_0}^{g-1} D_{i,g,t} \tilde{x}'_i \beta_{g,t}^{pre} + \sum_{g \in G} \sum_{t=g}^T D_{i,g,t} \tilde{x}'_i \beta_{g,t}^{post} \\
& + x'_i \beta + \sum_{t=t_0}^T D_{i,t} x'_i \beta_t + \sum_{g \in G} D_{i,g} x'_i \beta_g \xi_i + \xi_t + \varepsilon_{i,t}
\end{aligned} \tag{5}$$

Where $D_{i,t}$ is a dummy variable that is equal to 1 if period is equal to t , and $D_{i,g}$ is a dummy variable that is equal to 1 if the group membership is equal to g , and zero otherwise. \tilde{x}_i are the within cohort and period demeaned variables. When using `reghdfe` or `ppmlhdfe` the term $\sum_{g \in G} D_{i,g} x'_i \beta_g$ is omitted if all covariates are time constant.

Using \tilde{x} , default option, is not necessity for the estimation of treatment effects. However, if one uses that specification, the parameters $\theta_{g,t}^{pre}$ and $\theta_{g,t}^{post}$ still identify the average treatment effect for group g at time t . One could also use the untransformed covariates in Equation 5,

²The correction implemented with `corr` is not useful to recover the coefficients from `ppmlhdfe` using `poisson` command

and still be able to obtain the same group/time specific treatment effects with the post estimation commands.

From the user perspective, `jwddid` would simply need to be called as follows:

```
jwddid y x, ivar(i) tvar(t) gvar(g) never [xaxis]
```

Where `x` are all covariates of interest. If one uses the option `xaxis`, the command will use the covariates without demeaning them, which may save some computation time.

Extensions: Treatment Heterogeneity

As it may be apparent from Equation 5, the number of estimated parameters can grow quickly with the number groups/cohorts, periods of analysis, and covariates. This could lead to increasing computational burden of the estimation. An alternative, which is already implemented via `xthdidregress` and `hdidregress` in Stata 18, is to estimate the model that reduces the heterogeneity of the treatment effects. Specifically, it allows treatment effects to vary across cohorts, across absolute time, or across relative time.

For the case without covariates, the specification of Equation 3 can be modified to impose the heterogeneity restrictions as follows:

Time heterogeneity:

$$Y_{i,t} = \alpha + \sum_{t=t_0}^T \theta_t^{pre} D_{i,t,pre} + \sum_{t=t_0}^T \theta_t^{post} D_{i,t,post} + \xi_i + \xi_t + \varepsilon_{i,t} \quad (6)$$

Cohort heterogeneity:

$$Y_{i,t} = \alpha + \sum_{g \in G} \theta_g^{pre} D_{i,g,pre} + \sum_{g \in G} \theta_g^{post} D_{i,g,post} + \xi_i + \xi_t + \varepsilon_{i,t} \quad (7)$$

Event (Relative Time) heterogeneity:

$$Y_{i,t} = \alpha + \sum_{e=E_{min}}^{-2} \theta_e D_{i,e} + \sum_{e=0}^{E_{max}} \theta_e D_{i,e} + \xi_i + \xi_t + \varepsilon_{i,t} \quad (8)$$

Where $D_{i,t,pre}$ and $D_{i,t,post}$ are dummies that take the value of 1 if observation i , which is part of an eventually treated group, is not yet treated or is already treated at time t , respectively. $D_{i,g,pre}$ and $D_{i,g,post}$ are dummies that take the value of 1 if observation i belongs to group g and is not yet treated or is already treated at time t , respectively. $D_{i,e}$ is a dummy that takes the value of 1 if observation i is e periods relative to when treatment started. E_{min} and

E_{max} are the minimum and maximum event periods, possible. The *pre* coefficients are only considered when the **never** option is used.

The **jwddid** command allows the user to specify each one of these restrictions using the **hettype()** option.

```
jwddid y, ivar(i) tvar(t) gvar(g) hettype(option)
```

Where **option** can be **time**, **cohort**, or **event**. If no option is selected, the command will estimate the model described by Equation 3 which is the equivalent to allowing for full Cohort-time heterogeneity.

Extensions: Other Options

As described in Section , when covariates are considered in the model, the default option is to interact all covariates (or the demeaned transformations) with the same level of covariate heterogeneity. Some times, however, one may not be interested in estimating the same level of heterogeneity for all covariates. It may be possible, for example, to consider separate sets of covariates that could be interacted only with the time or the group dimensions.

Specifically, assume there are no variables we wish to consider for the treatment heterogeneity, but instead consider three sets of covariates: x^{EX} or variables we wish to incorporate without further interactions, x^T or variables that would be interacted with the Time variables only, and x^G or variables that would be interacted with group indicators only. In this case the setup would be:

$$\begin{aligned}
Y_{i,t} = & \alpha + \sum_{g \in G} \sum_{t=t_0}^{g-1} \theta_{g,t}^{pre} D_{i,g,t} + \sum_{g \in G} \sum_{t=g}^T \theta_{g,t}^{post} D_{i,g,t} \\
& + x_i'^{EX} \beta + \sum_{t=t_0}^T D_{i,t} x_i'^T \beta_t + \sum_{g \in G} D_{i,g} x_i'^G \beta_g \xi_i + \xi_t + \varepsilon_{i,t}
\end{aligned} \tag{9}$$

This specification can be estimated using the following syntax:

```
jwddid y , ivar(i) tvar(t) gvar(g) never exogvar(x_ex) xtvar(x_t) xgvar(x_g)
```

If covariates are included after y, they would still be treated following the specification of Equation 5, or following any of the heterogeneity restrictions described in Section .

An advanced version of this option is the inclusion of high order fixed effects (and interactions with fixed effects) that are different from the individual and time fixed effects. It is possible to request the inclusion of those types of fixed effects using the option **fevar()**, which is only valid if one is using the default estimator method **reghdfe** or **ppmlhdfe**. In both cases, the additional fixed effects (or interactions) are included in the estimation of the model without further interactions.

Post estimation: Aggregated treatment effects

After the estimation of the model, under the default options, one can use the coefficients $\theta_{g,t}^{post}$ as direct estimates of the group and time specific average treatment effects on the treated. However, one may also be interested in estimating aggregated ATTs for the overall data, across groups or periods, or dynamic effects. Furthermore, when the underlying method is a non-linear model, these coefficients cannot be directly interpreted as the average treatment effect on the outcome, but only on the latent variable.

`jwddid` comes along with the post estimation command `jwddid_estat` that can be used for that purpose. Internally, it uses the `margins` command to identify average treatment effects under the following algorithm:

1. Using the model estimates, predict the outcome of interest for all observations given the observed covariates and fixed effects. Call this $\hat{Y}(obs)$ or predicted outcome under the observed covariates. The model prediction could be the linear prediction, or the predicted probability in the case of logit models, or the predicted count in the case of poisson models.
2. Consider the specification `@eq_e2`, and assume that all $\theta_{g,t}^{post}$ (and $\theta_{g,t}^{pre}$ if `never` is used), as well as all β^{post} and β^{pre} are zero, and predict the outcome of interest. Call this $\hat{Y}(0)$ or predicted outcome under the counterfactual scenario of no treatment.

In this case, the predicted Average Treatment Effect on the Treated for observation i is given by:

$$\widehat{ATT}_i = \hat{Y}(obs) - \hat{Y}(0)$$

This is zero for observations that were never treated, and nonzero for the treated-post treatment observations.³

From here, any aggregated average treatment effects can be calculated as follows:

$$AGGTE_r = \frac{\sum_i^N \widehat{ATT}_i \times w_{i,t} \times R_{i,t}}{\sum_i^N w_{i,t} \times R_{i,t}}$$

where $R_{i,t}$ takes the value of one whenever observation i fulfills the required conditions, and $w_{i,t}$ is the weight of the observation i at time t used in for the estimation model.⁴ $AGGTE$ is the aggregated average treatment effect on the treated given the conditions R .

In general, there are four types of aggregations that are implemented in the `jwddid_estat/estat` command:

³The case of pre-treatment treated observations may be assumed to be zero if the `never` option is not used. When `never` is used, pre-treatment ATTs can be used for testing the parallel trends assumption.

⁴While this is the default option, `estat` command also allow you to provide other weights for aggregation

1. **estat simple**: This option calculates the average treatment effect on the treated for all observations that were treated at some point in time. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t \geq g \text{ for observation } i \in g$$

$$R_{i,t} = 0 \text{ otherwise}$$

2. **estat group**: This option calculates the average treatment effect for observations that were treated at time g . The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t \geq g_c \text{ for observation } i$$

$$R_{i,t} = 0 \text{ otherwise}$$

where g_c is a particular group/cohort of interest. **estat group** estimates this for all groups g in G .

3. **estat time**: This option calculates the average treatment effect at time t for all observations that were effectively treated at that point. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t_c \geq g \text{ for observation } i$$

$$R_{i,t} = 0 \text{ otherwise}$$

where t_c is a particular time of interest. **estat time** estimates this for all times t in T that has at least one unit that was treated.

4. **estat event**: This option calculates dynamic treatment effects, also known as event studies, using the period before treatment as the reference. When the condition ‘never’ is used, this approach can be used to estimate pre-treatment ATT’s, which could be used for testing PTA. The condition R is defined as:

$$R_{i,t} = 1 \text{ if } t - g = e_c \text{ \& } e \neq -1 \text{ for observation } i$$

$$R_{i,t} = 0 \text{ otherwise}$$

where e_c is a particular event of interest. In contrast with the previous aggregations, if option **never** was used for estimation, one could also add the option **pretrend** to run a simple PTA test with the following null hypothesis:

$$H_0 : AGGTE_e = 0 \text{ for all } e < -1 \text{ vs } H_1 : AGGTE_e \neq 0 \text{ for some } e < -1$$

Failure to reject this hypothesis is evidence in support of parallel trends assumption. Otherwise, one can use `test` command to test for the significance of specific pre-treatment ATT's.⁵

Post estimation: Other options

The `jwddid_estat/estat` command allows for further options that may be of interest for the user. In this section we provide a brief description of those options:

Weights

The default option for the estimation of the aggregated ATTs is to use the weights $w_{i,t}$ that were used in the estimation of the model. However, if the user wants to use different weights, it is possible to do using the following syntax:

```
estat [aggregation] [pw = weight]
```

Standard Errors

@wooldridge2022 suggests that when one estimates standard errors for the aggregated ATTs, one should use `vce(unconditional)` option in `Stata`, to allow for uncertainty in the explanatory variables. `jwddid_estat/estat` does not use this approach by default, because it requires that the underlying command is able to produce Scores for the estimated model. For example, if the model was estimated using `method(regress)`, the Scores will be available, and unconditional Standard errors for the aggregated ATTs can be estimated as follows:

```
estat [aggregation], [vce(unconditional)]
```

This is not possible if one uses the default methodology `reghdfe`.

Other Aggregation restrictions

As described in Section , the default aggregation considers all observations treated observations, imposing restrictions only in terms of time, group or event dimensions. However, one may be interested in imposing further restrictions that could leverage on the use of covariates in the model specification. For example, say that one estimates a DID model with covariates using the following syntax:

```
jwddid y i.dx, ivar(i) tvar(t) gvar(g) never
```

⁵It should be noticed that this test is different from the test proposed by @callaway2021, which is based on testing *all* group/time specific ATT's, instead of the event aggregated ones.

As usual, one could request the estimation of the aggregated ATTs for the whole sample as follows:

```
estat [aggregation]
```

However, one would also be able to make the same estimation imposing the added restriction that the covariate `dx` is zero or one, using the option `orestriction()`:

```
estat [aggregation], orestriction(dx==0)
```

```
estat [aggregation], orestriction(dx==1)
```

The expression inside the parenthesis should be a valid Stata expression that is used when calculating the aggregated ATTs.

Storing, and saving results

After aggregate effects have been estimated, the user may want to store the results for further analysis or reporting. Because `estat` uses `margins` in the background, the default option is to store the results in memory as `r()` elements. Alternatively, `jwddid_estat/estat` allows the user to store the output of the command using three different options:

1. `estat [aggregation], post`: As with `margins`, option `post` will “post” the results of the command to be the current estimations in memory `e()`, which can be saved as usual for further analysis.
2. `estat [aggregation], estore(name)`: This option stores the results from the aggregation in memory as `name`. This is similar to using `estimation store name` after a regression command. The previously estimated results from `jwddid` are not overwritten.
3. `estat [aggregation], esave(filename)`: This option saves the results from the aggregation in a file `filename`, as a `ster` file, which can be used at a later point.

Plotting

After `time`, `group` or `event` aggregations are estimated, it is possible to request plotting those results using `estat plot`. The basic syntax is to type it after the aggregation command, with only minimal command specific options:

1. `estat plot, style(style)`: The option `style` allows the user to select the style of the plot. The default is using a `rspike` style, but `rarea`, `rcap` and `rbar` are also available. See `help twoway` for more information on the styles.
2. `estat plot, pstyle1(str) color1(str) pstyle2(str) color2(str) lwidth1(str) lwidth2(str) barwidth1(str) barwidth2(str)`: The options `pstyle#`, `color#`, `lwidth#` and `barwidth#` can be used to alter the color and style of the lines in the plot for the pre and post treatment periods. Only `event` aggregation allows for `#2` options.

3. `estat plot, twoway_options`: Most other twoway graph option can be used after the `estat plot` command.