## Estimation of DID models using ETWFE

In this section, we describe the Stata command `jwdid` 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 `jwdid` 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 @Wooldrige2021, the baseline model for the estimation of the DID model using the ETWFE estimator is the following:

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

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.

$\xi_i$ and $\xi_t$ 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.

[1] Often, 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 `jwdid` using the following syntax:

```
jwdid 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, `jwdid` 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:

```
jwdid 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:

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

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`:

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

Which will estimate the following model:

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

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 @Wooldrdige2022, 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:

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

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 identify 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 exponetial 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 `jwdid` command allows the user to specify the `method()` option to estimate models described by Equation 4, where one would specificy 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:

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

There is no restrictions on the type `method` one can use with the `jwdid` 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, whenver `method()` is specified `jwdid` 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. `jwdid` 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:

$$
\begin{aligned}
Y_{i,t} = \alpha &+ \sum_{g\in G}\sum_{t=t_0}^{g-1}\theta^{pre}_{g,t}D_{i,g,t} + \sum_{g\in G}\sum_{t=g}^{T}\theta^{post}_{g,t}D_{i,g,t} + \sum_{g\in G}\sum_{t=t_0}^{g-1}D_{i,g,t}x_i'\beta^{pre}_{g,t} + \sum_{g\in G}\sum_{t=g}^{T}D_{i,g,t}x_i'\beta^{post}_{g,t} \\
&+ x_i'\beta + \sum_{t=t_0}^{T}D_{i,t}x_i'\beta_t + \xi_i + \xi_t + \varepsilon_{i,t}
\end{aligned} \tag{5}
$$

Where $D_{i,t}$ is a dummy variable that is equal to 1 if period is equal to $t$, and zero otherwise. This specification is used if `reghdfe` or `ppmlhdfe` are used to estimate the model. Otherwise, if group fixed effects are used, the model also includes an interaction between $x's$ and the group fixed effects Dummies.

From the user persective, `jwdid` would simply need to be called as follows:

```
jwdid y x, ivar(i) tvar(t) gvar(g) never [options]
```

Where `x` is the covariate of interest.

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

### Extensions: Treatment Heterogeneity

As it may be aparent from Equation 5, the number of estimated parameters can grow quickly with the number groups/cohorts, periods of analysis, and covariates, specially if we constrain the analysis to use only never-treated units as controls. This could lead to a large number of parameters to be estimated, increasing the 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.