Two-Way FE Difference in Difference

By construction, the difference-in-difference estimate of the average treatment effect of X_{it} on Y_{it} in a linear equation is $E[Y_{21} - Y_{11}]$ (difference between treated/control at $t = 1$) minus $E[Y_{22} - Y_{12}]$ (difference between treated/control at $t = 1$). In the two-way FE model of D-in-D, β serves as an estimand for the ATE.

Using the following linear models:

$$
\begin{aligned}
Y_{11} &= \alpha + \beta * X_{11} + \eta_1 + \nu_1 \\
Y_{12} &= \alpha + \beta * X_{12} + \eta_1 + \nu_2 \\
Y_{21} &= \alpha + \beta * X_{21} + \eta_2 + \nu_1 \\
Y_{22} &= \alpha + \beta * X_{22} + \eta_2 + \nu_2
\end{aligned}
$$

and a binary treatment X with $X = 1$ if $i = 2$ and $X = 0$ if $i = 1$. (The treatment variable and the units are perfectly collinear).

Then, the first difference, or the artificial control, is equal to $E[Y_{21} - Y_{11}]$, which from the equations above reduces to:

$$E[Y_{21} - Y_{11}] = \beta * X_{21} + \eta_2 - \beta * X_{11} - \eta_1$$

because the time fixed effects will cancel each other at this time point. Similarly, the second difference, or the artificial treatment, $E[Y_{22} - Y_{12}]$, is equal to

$$E[Y_{21} - Y_{11}] = \beta * X_{22} + \eta_2 - \beta * X_{12} - \eta_1$$

Taking the difference of these differences yields the following expression:

$$E[(Y_{22} - Y_{12})_(Y_{21} - Y_{11}))] = \beta * X_{22} - \beta * X_{12} - \beta * X_{21} + \beta * X_{11}$$

Because the unit fixed effects cancel out in this difference. However, as written above, if $X = 1$ if $i = 2$ and $X = 0$ if $i = 1$, then the expression above reduces to the following:

$$
\begin{aligned}
E[DD_{2FE}] &= \beta * 1 - \beta * 0 - \beta * 1 + \beta * 0 \\
E[DD_{2FE}] &= 0
\end{aligned}
$$

So, for all possible β, the true two-way FE estimate is zero. (Or check my math if I did something wrong). That means that any non-zero values will come from bias in the errors (i.e. they are not in fact mean zero), or if the panels are not balanced (all the βs are not the same).

Interactive Between Effect D-in-D

If, on the other hand, we use the interactive between effects estimator, then we are interested in the effect δ as an estimand for the ATE. As above, $X = 0$ if $i = 1$ and $X = 1$ if $i = 2$, and the equations are as follows:

$$\begin{aligned}
Y_{11} &= \alpha + \beta * X_{11} + \nu_1 + \delta * \nu_1 * X_{11} \\
Y_{12} &= \alpha + \beta * X_{12} + \nu_2 + \delta * \nu_2 * X_{12} \\
Y_{21} &= \alpha + \beta * X_{21} + \nu_1 + \delta * \nu_1 * X_{21} \\
Y_{22} &= \alpha + \beta * X_{22} + \nu_2 + \delta * \nu_2 * X_{22}
\end{aligned}$$

Then, the first difference, or the artificial control, is equal to $E[Y_{21} - Y_{11}]$, which from the equations above reduces to:

$$E[Y_{21} - Y_{11}] = \beta * X_{21} + \delta * \nu_1 * X_{21} - \beta * X_{11} - \delta * \nu_1 * X_{11}$$

Because the time fixed effects will cancel each other at this time point. Similarly, for the second difference, or the artificial treatment, $E[Y_{22} - Y_{12}]$, is equal to

$$E[Y_{21} - Y_{11}] = \beta * X_{22} + \delta * \nu_2 * X_{22} - \beta * X_{12} - \delta * \nu_2 * X_{12}$$

Taking the difference of these differences yields the following expression:

$$\begin{aligned}
E[((Y_{22} - Y_{12}) - (Y_{21} - Y_{11}))] =& \beta * X_{22} + \delta * \nu_2 * X_{22} - \beta * X_{12} - \delta * \nu_2 * X_{12} \\
& - \beta * X_{21} - \delta * \nu_1 * X_{21} + \beta * X_{11} + \delta * \nu_1 * X_{11}
\end{aligned}$$

To plug in the actual values of the variables X and ν, X is either 0 or 1 as before, and $\nu = 0$ for ν_1 and $\nu = 1$ for ν_2 because the first time point will serve as a reference category in the regression. This yields the following expression:

$$\begin{aligned}
E[DD_{IBE}] &= \beta * 1 + \delta * 1 * 1 - \beta * 0 - \delta * 1 * 0 \\
&\quad - \beta * 1 - \delta * 0 * 1 + \beta * 0 + \delta * 0 * 0 \\
E[DD_{IBE}] &= \delta
\end{aligned}$$

So the interactive between effect version of the D-in-D estimator does return δ as an unbiased quantity in expectation, assuming that the errors are mean zero (i.e., the parallel paths assumption).