I. APPENDIX A: DERIVATION OF THE DD DECOMPOSITION THEOREM

A. Derivation of equation (6)

$$\begin{aligned}\widehat{cov}(\tilde{y}_{it}, \tilde{D}_{it}) &= \frac{1}{NT}\sum_i\sum_t [(y_{it}-\bar{\bar{y}}) - (\bar{y}_i - \bar{\bar{y}}) - (\bar{y}_t - \bar{\bar{y}})]\cdot[(D_{it}-\bar{\bar{D}}) - (\bar{D}_i - \bar{\bar{D}}) - (\bar{D}_t - \bar{\bar{D}})] \\ &= \frac{1}{NT}\sum_i\sum_t [(y_{it}-\bar{\bar{y}})(D_{it}-\bar{\bar{D}}) + (\bar{y}_i-\bar{\bar{y}})(\bar{D}_i-\bar{\bar{D}}) + (\bar{y}_t-\bar{\bar{y}})(\bar{D}_t-\bar{\bar{D}}) \\ &\quad -(y_{it}-\bar{\bar{y}})(\bar{D}_i-\bar{\bar{D}}) - (D_{it}-\bar{\bar{D}})(\bar{y}_i-\bar{\bar{y}}) - (y_{it}-\bar{\bar{y}})(\bar{D}_t-\bar{\bar{D}}) - (D_{it}-\bar{\bar{D}})(\bar{y}_t-\bar{\bar{y}}) \\ &\quad -(\bar{y}_i-\bar{\bar{y}})(\bar{D}_t-\bar{\bar{D}}) - (\bar{D}_i-\bar{\bar{D}})(\bar{y}_t-\bar{\bar{y}})]\end{aligned}$$

Pass the sums through and write as means:

$$\begin{aligned}= &\frac{1}{NT}\sum_i\sum_t (y_{it}-\bar{\bar{y}})(D_{it}-\bar{\bar{D}}) + \frac{1}{N}\sum_i(\bar{y}_i-\bar{\bar{y}})(\bar{D}_i-\bar{\bar{D}}) + \frac{1}{T}\sum_t(\bar{y}_t-\bar{\bar{y}})(\bar{D}_t-\bar{\bar{D}}) \\ &-2\frac{1}{N}\sum_i(\bar{y}_i-\bar{\bar{y}})(\bar{D}_i-\bar{\bar{D}}) - 2\frac{1}{T}\sum_t(\bar{y}_t-\bar{\bar{y}})(\bar{D}_t-\bar{\bar{D}}) \\ &-\frac{1}{NT}\sum_i(\bar{y}_i-\bar{\bar{y}})\underbrace{\sum_t(\bar{D}_t-\bar{\bar{D}})}_{=0} - \frac{1}{NT}\sum_i(\bar{D}_i-\bar{\bar{D}})\underbrace{\sum_t(\bar{y}_t-\bar{\bar{y}})}_{=0}]\end{aligned}$$

This yields equation (6).

B. Proof of the Difference-in-Differences Decomposition Theorem

The DD decomposition theorem follows from solving out the covariance between adjusted variables from equation (6):

$$\widehat{cov}(\tilde{y}_{it}, \tilde{D}_{it}) = \frac{1}{NT}\sum_i\sum_t \left(y_{it}-\bar{\bar{y}}\right)\left(D_{it}-\bar{\bar{D}}\right) - \frac{1}{N}\sum_i\left(\bar{y}_i-\bar{\bar{y}}\right)\left(\bar{D}_i-\bar{\bar{D}}\right) - \frac{1}{T}\sum_t\left(\bar{y}_t-\bar{\bar{y}}\right)\left(\bar{D}_t-\bar{\bar{D}}\right)$$

Because D_{it} and D_i equal zero in the control group, and D_{it} equals zero in all pre-treatment periods, the first two terms in equation (6) collapse to:

$$\frac{1}{NT}\sum_{i\notin U}\sum_{t\geq t_i^*} y_{it} - \frac{1}{N}\sum_{i\notin U}\bar{y}_i\,\bar{D}_i = \frac{1}{N}\sum_{i\notin U}\left[\frac{1}{T}\sum_{t\geq t_i^*} y_{it} - \bar{y}_i\bar{D}_i\right] \qquad (A1)$$

We can write the unit mean of y as an average between a pre- and post-period: $\bar{y}_i = (1 - \bar{D}_j)\bar{y}_i^{PRE(j)} + \bar{D}_j \bar{y}_i^{POST(j)}$. Equation (A1) therefore contains pre/post averages for each unit according to its own treatment timing, and these can be further collapsed to averages for treatment timing groups (rather than individual units):

$$\frac{1}{N}\sum_{i \notin U}\left(\bar{y}_i^{POST(i)} - \bar{D}_i \bar{y}_i^{POST(i)} - (1-\bar{D}_i)\bar{y}_i^{PRE(i)}\right)\bar{D}_i = \sum_k n_k \bar{D}_k (1-\bar{D}_k)\left(\bar{y}_k^{POST(k)} - \bar{y}_k^{PRE(k)}\right) \quad (A2)$$

The third term in equation (6), a sum over t, can be broken up into pieces based on each groups' treatment time: no units are treated before t_1^* ($\bar{D}_t = 0$); only units in group $k = 1$ are treated between t_1^* and $t_2^* - 1$ ($\bar{D}_t = n_1$); only units in groups $k = 1$ and $k = 2$ are treated between and t_2^* and $t_3^* - 1$ ($\bar{D}_t = n_1 + n_2$), etc. The share of all NT observations that are treated equals $\bar{\bar{D}} = \sum_k n_k \bar{D}_k$. This yields:

$$\begin{aligned} & \frac{1}{T}\sum_{t<t_E^*}\left(\bar{y}_t - \bar{\bar{y}}\right)\left(-\sum_k n_k \bar{D}_k\right) \\ + \quad & \frac{1}{T}\sum_{t_1^* \le t < t_2^*}\left(\bar{y}_t - \bar{\bar{y}}\right)\left(n_1(1-\bar{D}_1) - \sum_{k>1} n_k \bar{D}_k\right) \\ + \quad & \frac{1}{T}\sum_{t_2^* \le t < t_3^*}\left(\bar{y}_t - \bar{\bar{y}}\right)\left(\sum_{k\le 2} n_k(1-\bar{D}_k) - \sum_{k>2} n_k \bar{D}_k\right) \\ & \vdots \\ + \quad & \frac{1}{T}\sum_{t_K^* \le t}\left(\bar{y}_t - \bar{\bar{y}}\right)\left(\sum_{k\le 2} n_k(1-\bar{D}_k)\right) \qquad (A3) \end{aligned}$$

Grouping the sums by their weights (the n and $\bar{D}$ terms) leads to an expression in terms of every timing group's pre- and post-period:

$$= \sum_{k \neq U} \frac{1}{T} \left[\underbrace{\sum_{t \geq t_k^*} (\overline{y}_t - \overline{\overline{y}}) n_k (1 - \overline{D}_k)}_{\text{Post}-t_k^*} - \underbrace{\sum_{t < t_E^*} (\overline{y}_t - \overline{\overline{y}}) n_k \overline{D}_k}_{\text{Pre}-t_k^*} \right] \qquad (A4)$$

Each $\overline{y}_t - \overline{\overline{y}}$ is a weighted average group deviations in period t: $\overline{y}_t - \overline{\overline{y}} = \sum_k n_k (\overline{y}_{kt} - \overline{y}_k)$. This shows that every group of cross-sectional units contributes a pre/post difference according to *every* treatment time. Here k indexes each group, including untreated units, and ℓ indexes treatment times that determine the periods in which means are taken. Note that the order of the sum matters because outcome differences for group k according to group ℓ's timing is not the same as outcome differences for group ℓ according to group k's timing:

$$\sum_k \sum_{\ell \neq U} \left[n_\ell n_k \overline{D}_\ell (1 - \overline{D}_\ell) \left(\overline{y}_k^{POST(\ell)} - \overline{y}_k^{PRE(\ell)} \right) \right] = \sum_k \sum_{\ell \neq U} \left[n_\ell n_k \overline{D}_\ell (1 - \overline{D}_\ell) \Delta(y_\ell, t_k^*) \right] \qquad (A5)$$

Where $\Delta(y_a, t_b^*) = \overline{y}_a^{POST(b)} - \overline{y}_a^{PRE(b)}$, the change in average outcomes in group a according to group b's treatment timing.

Combining equations (A2) and (A5) shows that the covariance equals:

$$\sum_{k \neq C} n_k \overline{D}_k (1 - \overline{D}_k) \Delta(y_k, t_k^*) - \sum_\ell \sum_{k \neq C} n_\ell n_k \overline{D}_k (1 - \overline{D}_k) \Delta(y_\ell, t_k^*) \qquad (A6)$$

where ℓ indexes groups of observations (including the control group) so that the double-sum includes all combinations of the cross-sectional units with treatment timings.

Break out the untreated group from the ℓ part of the second sum (these are the terms that compare changes in outcomes in the untreated group according to *every* possible timing):

$$\sum_{k \neq U} n_k (1 - n_k) \overline{D}_k (1 - \overline{D}_k) \Delta(y_k, t_k^*) - \sum_{k \neq U} n_k n_U \overline{D}_k (1 - \overline{D}_k) \Delta(y_U, t_k^*)$$

$$- \sum_{k \neq U} \sum_{\ell \neq k} n_\ell n_k \overline{D}_k (1 - \overline{D}_k) \Delta(y_\ell, t_k^*) \qquad (A7)$$

For all $k \neq U$, subtract $n_k(1-n_k-n_U)\overline{D}_k(1-\overline{D}_k)\Delta(y_k,t_k^*)$ so that each $n_k(1-n_k)\overline{D}_k\big(1-\overline{D}_k\big)\Delta(y_k,t_k^*)$ becomes $n_k n_U \overline{D}_k\big(1-\overline{D}_k\big)\Delta(y_k,t_k^*)$ and the two sums in the first line of (A7) can be combined. This implies that we must add $n_k \sum_{\substack{\ell\neq U\\ \ell\neq k}} n_\ell\, \overline{D}_k(1-\overline{D}_k)\Delta(y_k,t_k^*)$ for each $k \neq U$, which is the same because $\sum_{\substack{\ell\neq U\\ \ell\neq k}} n_\ell = (1-n_k-n_U)$. This yields an expression only in terms of differences across groups in differences between pre/post periods:

$$\sum_{k\neq U} n_k\, n_U \overline{D}_k\big(1-\overline{D}_k\big)[\Delta(y_k,t_k^*)-\Delta(y_U,t_k^*)]$$

$$+\sum_{k\neq U}\sum_{\ell\neq k} n_k\, n_\ell \overline{D}_k\big(1-\overline{D}_k\big)[\Delta(y_k,t_k^*)-\Delta(y_\ell,t_k^*)] \qquad (A8)$$

Note that every combination of ℓ and k appears twice in the double sum so equation (A8) equals:

$$\sum_{k\neq U} n_k\, n_U \overline{D}_k\big(1-\overline{D}_k\big)\overbrace{[\Delta(y_k,t_k^*)-\Delta(y_U,t_k^*)]}^{\widehat{\beta}_{kU}^{DD}}$$

$$+\sum_{k\neq U}\sum_{\ell>k} n_k\, n_\ell\{\overline{D}_k\big(1-\overline{D}_k\big)[\Delta(y_k,t_k^*)-\Delta(y_\ell,t_k^*)]+\overline{D}_\ell\big(1-\overline{D}_\ell\big)[\Delta(y_\ell,t_\ell^*)-\Delta(y_k,t_\ell^*)]\} \; (A9)$$

The first sum in (A9) contains all the 2x2 DDs that compare one treatment timing group to the untreated group. Substituting the definition of $\Delta(y_a,t_b^*)$ shows that each term in the second line of (A9) equals:

$$= \quad \overline{D}_k\big(1-\overline{D}_k\big)\left[\left(\overline{y}_k^{POST(k)}-\overline{y}_k^{PRE(k)}\right)-\left(\overline{y}_\ell^{POST(k)}-\overline{y}_\ell^{PRE(k)}\right)\right]$$

$$+ \quad \overline{D}_\ell\big(1-\overline{D}_\ell\big)\left[\left(\overline{y}_\ell^{POST(\ell)}-\overline{y}_\ell^{PRE(\ell)}\right)-\left(\overline{y}_k^{POST(\ell)}-\overline{y}_k^{PRE(\ell)}\right)\right]$$

Recall that in my notation group k is always treated earlier than group ℓ, which means that, as shown in figure 1, the period $POST(k)$ includes the middle window where treatment status differs, $MID(k,\ell)$ and the $POST(\ell)$ period. Therefore, any group's mean outcome in group k's post-period are a weighted average of these two periods where the weights are just the share of periods in $MID(k,\ell)$ versus $POST(\ell)$:

$$\overline{y}_j^{POST(k)} = \frac{\overline{D}_k - \overline{D}_\ell}{\overline{D}_k}\overline{y}_j^{MID(k,\ell)} + \frac{\overline{D}_\ell}{\overline{D}_k}\overline{y}_j^{POST(\ell)}$$

Similarly, group ℓ's pre-period includes $PRE(k)$ and $MID(k,\ell)$, so:

$$\overline{y}_j^{PRE(\ell)} = \frac{1-\overline{D}_k}{1-\overline{D}_\ell}\overline{y}_j^{PRE(k)} + \frac{\overline{D}_k - \overline{D}_\ell}{1-\overline{D}_\ell}\overline{y}_j^{MID(k,\ell)}$$

Substituting these identities into (A10) yields:

$$\begin{aligned}\overline{D}_k\left(1-\overline{D}_k\right)&\left[\left(\frac{\overline{D}_k - \overline{D}_\ell}{\overline{D}_k}\overline{y}_k^{MID(k,\ell)} + \frac{\overline{D}_\ell}{\overline{D}_k}\overline{y}_k^{POST(\ell)} - \overline{y}_k^{PRE(k)}\right)\right.\\ &\left.-\left(\frac{\overline{D}_k - \overline{D}_\ell}{\overline{D}_k}\overline{y}_\ell^{MID(k,\ell)} + \frac{\overline{D}_\ell}{\overline{D}_k}\overline{y}_\ell^{POST(\ell)} - \overline{y}_\ell^{PRE(k)}\right)\right]\\ &+\quad \overline{D}_\ell\left(1-\overline{D}_\ell\right)\left[\left(\overline{y}_\ell^{POST(\ell)} - \frac{1-\overline{D}_k}{1-\overline{D}_\ell}\overline{y}_\ell^{PRE(k)} - \frac{\overline{D}_k - \overline{D}_\ell}{1-\overline{D}_\ell}\overline{y}_\ell^{MID(k,\ell)}\right)\right.\\ &\left.-\left(\overline{y}_k^{POST(\ell)} - \frac{1-\overline{D}_k}{1-\overline{D}_\ell}\overline{y}_k^{PRE(k)} - \frac{\overline{D}_k - \overline{D}_\ell}{1-\overline{D}_\ell}\overline{y}_k^{MID(k,\ell)}\right)\right]\end{aligned}$$

Distributing the $\overline{D}_k\left(1-\overline{D}_k\right)$ and $\overline{D}_\ell\left(1-\overline{D}_\ell\right)$ terms yields:

$$\begin{aligned}&\left(\left(1-\overline{D}_k\right)(\overline{D}_k - \overline{D}_\ell)\overline{y}_k^{MID(k,\ell)} + \left(1-\overline{D}_k\right)\overline{D}_\ell\overline{y}_k^{POST(\ell)} - \overline{D}_k\left(1-\overline{D}_k\right)\overline{y}_k^{PRE(k)}\right)\\ &-\left(\left(1-\overline{D}_k\right)(\overline{D}_k - \overline{D}_\ell)\overline{y}_\ell^{MID(k,\ell)} + \left(1-\overline{D}_k\right)\overline{D}_\ell\overline{y}_\ell^{POST(\ell)} - \overline{D}_k\left(1-\overline{D}_k\right)\overline{y}_\ell^{PRE(k)}\right)\\ &+\left(\overline{D}_\ell\left(1-\overline{D}_\ell\right)\overline{y}_\ell^{POST(\ell)} - \overline{D}_\ell\left(1-\overline{D}_k\right)\overline{y}_\ell^{PRE(k)} - \overline{D}_\ell(\overline{D}_k - \overline{D}_\ell)\overline{y}_\ell^{MID(k,\ell)}\right)\\ &-\left(\overline{D}_\ell\left(1-\overline{D}_\ell\right)\overline{y}_k^{POST(\ell)} - \overline{D}_\ell\left(1-\overline{D}_k\right)\overline{y}_k^{PRE(k)} - \overline{D}_\ell(\overline{D}_k - \overline{D}_\ell)\overline{y}_k^{MID(k,\ell)}\right)\end{aligned}$$

Grouping the $\overline{y}^{PRE(k)}$ and $\overline{y}^{POST(\ell)}$ for each group shows that their weights each match some of the $\overline{y}^{MID(k,\ell)}$ terms, and that this expression contains the 2x2 DDs from a timing-only comparison between groups k and ℓ:

$$\left(1-\bar{D}_k\right)(\bar{D}_k-\bar{D}_\ell)\overbrace{\left[\left(\bar{y}_k^{MID(k,\ell)}-\bar{y}_k^{PRE(k)}\right)-\left(\bar{y}_\ell^{MID(k,\ell)}-\bar{y}_\ell^{PRE(k)}\right)\right]}^{\hat{\beta}_{k\ell}^{DD,k}}$$

$$+\bar{D}_\ell(\bar{D}_k-\bar{D}_\ell)\underbrace{\left[\left(\bar{y}_\ell^{POST(\ell)}-\bar{y}_\ell^{MID(k,\ell)}\right)-\left(\bar{y}_k^{POST(\ell)}-\bar{y}_k^{MID(k,\ell)}\right)\right]}_{\hat{\beta}_{k\ell}^{DD,\ell}} \qquad (A11)$$

Since $\left(1-\bar{D}_k\right)(\bar{D}_k-\bar{D}_\ell)+\bar{D}_\ell(\bar{D}_k-\bar{D}_\ell)=(1-(\bar{D}_k-\bar{D}_\ell))(\bar{D}_k-\bar{D}_\ell)$, equation (A11) yields equations (7) and (8) in the paper, which define $\mu_{k\ell}$ and show that a two-group timing estimate weights together the 2x2 DDs on the later group's pre-period and on the early group's post-period:

$$\left(1-(\bar{D}_k-\bar{D}_\ell)\right)(\bar{D}_k-\bar{D}_\ell)\left[\overbrace{\frac{1-\bar{D}_k}{1-(\bar{D}_k-\bar{D}_\ell)}}^{\mu_{k\ell}}\hat{\beta}_{k\ell}^{DD,k}+\overbrace{\frac{\bar{D}_\ell}{1-\bar{D}_k}}^{1-\mu_{k\ell}}\hat{\beta}_{k\ell}^{DD,\ell}\right] \qquad (A12)$$

Combining equations (A9) and (A12) shows that:

$$\widehat{cov}\left(\tilde{y}_{it},\tilde{D}_{it}\right)=\sum_{k\neq U} n_k\, n_U \bar{D}_k\left(1-\bar{D}_k\right)\hat{\beta}_{kU}^{DD}$$

$$+\sum_{k\neq U}\sum_{\ell>k} n_k\, n_\ell\left(1-(\bar{D}_k-\bar{D}_\ell)\right)(\bar{D}_k-\bar{D}_\ell)\left[\mu_{k\ell}\,\hat{\beta}_{k\ell}^{DD,k}+(1-\mu_{k\ell})\,\hat{\beta}_{k\ell}^{DD,\ell}\right] \quad (A13)$$

To solve for the denominator of $\hat{\beta}^{DD}$, we just need to evaluate (A13) using D_{it} in place of y_{it}. Averages of D_{it} in pre- and post-periods are, obviously, equal to 1 and 0. The definitions of the non-trivial terms in the denominator:

$$\Delta(D_\ell,t_k^*) \quad = \quad \frac{1}{T_k^*}\left(\sum_{t\geq t_k^*} D_{i\in\ell,t}\right)-0=\frac{\bar{D}_\ell}{\bar{D}_k}$$

$$\Delta(D_k,t_\ell^*) \quad = \quad \frac{1}{T_\ell^*}\left(\sum_{t\geq t_\ell^*} D_{i\in k,t}\right)-\frac{1}{T-T_\ell^*}\left(\sum_{t<t_\ell^*} D_{i\in k,t}\right)=1-\frac{T_k^*-T_\ell^*}{T-T_\ell^*}=\frac{1-\bar{D}_k}{1-\bar{D}_\ell}$$

Plugging in shows that:

$$\widehat{var}\left(\tilde{D}_{it}\right)=$$

$$\sum_{k\neq U} n_k\, n_U \overline{D}_k(1-\overline{D}_k) + \sum_{k\neq U}\sum_{\ell>k} n_k\, n_\ell \left(\overline{D}_k\left(1-\overline{D}_k\right)\left[1-\frac{\overline{D}_\ell}{\overline{D}_k}\right] + \overline{D}_\ell\left(1-\overline{D}_\ell\right)\left[1-\frac{1-\overline{D}_k}{1-\overline{D}_\ell}\right]\right)$$

$$= \sum_{k\neq U} n_k\, n_U \overline{D}_k\left(1-\overline{D}_k\right) + \sum_{k\neq U}\sum_{\ell>k} n_k\, n_\ell\left(\overline{D}_k-\overline{D}_\ell\right)\left(1-\left(\overline{D}_k-\overline{D}_\ell\right)\right) \qquad (A14)$$

This is the sum of the weights on each $\hat{\beta}$ term in (A13), which defines the decomposition weights s_{kU} and $s_{k\ell}$.

II. Appendix B: Derivation of the Decomposition with Unit-Specific Trends

The unit-specific trend specification is:

$$y_{it} \quad = \quad \alpha_i + \alpha_t + A^{k(i)}\alpha_i\big(t - \bar{t}\big) + \beta^{DD}_{trend} D_{it} + e_{it}$$

Before turning to the decomposition, we need to know how unit-specific trends work in general. To apply the Frisch-Waugh theorem we would first fit a linear trend to $\widetilde{D}_{it}$ for every unit: $\widetilde{D}_{it} = A^{k(i)}\big(t - \bar{t}\big) + \widehat{D}_{it}$. The notation $A^{k(i)}$ indicates that the slopes only vary by timing group. The residuals from these detrending regressions ($\widehat{D}_{it} = \widetilde{D}_{it} - A^{k(i)}\big(t - \bar{t}\big)$) are the "new" identifying variation. $\hat{\beta}^{DD}_{trend}$ equals the following univariate regression coefficient:

$$\frac{\widehat{cov}\left(\widehat{D}_{it}, y_{it}\right)}{\widehat{var}\left(\widehat{D}_{it}\right)} = \overbrace{\frac{\widehat{cov}\left(\widetilde{D}_{it}, y_{it}\right)}{\widehat{var}\left(\widetilde{D}_{it}\right)}}^{\widehat{\beta}^{DD}} \overbrace{\frac{\widehat{var}\left(\widetilde{D}_{it}\right)}{\widehat{var}\left(\widehat{D}_{it}\right)}}^{\frac{1}{1-R^2}} - \overbrace{\frac{\widehat{cov}\left(A^{k(i)}(t - \bar{t}), y_{it}\right)}{\widehat{var}\left(A^{k(i)}(t - \bar{t})\right)}}^{\widehat{\beta}^{trend}} \overbrace{\frac{\widehat{var}\left(A^{k(i)}(t - \bar{t})\right)}{\widehat{var}\left(\widehat{D}_{it}\right)}}^{\frac{R^2}{1-R^2}}$$

$$= \widehat{\beta}^{DD} + \frac{R^2}{1 - R^2}\left[\widehat{\beta}^{DD} - \widehat{\beta}^{trend}\right]$$

The detrended estimator equals the unadjusted estimate, $\widehat{\beta}^{DD}$, minus a function of $\widehat{\beta}^{trend}$, the relationship between y and the fitted trends. This is the sense in which trends "take out" underlying trends in the outcome. Moreover, the variance ratios are functions of the R^2 from the regression that fits trends to $\widetilde{D}_{it}$. Unit-specific matter more when they fit $\widetilde{D}_{it}$ better.

In this appendix I first solve for R^2 and $\widehat{\beta}^{trend}$ in the two-group cases with one treated and one untreated group, and with two treated groups (timing only). Next, I solve for the full trend-adjusted estimator and show that it is a weighted average of the two-group detrended estimators.[1] Finally, I apply these expression to a potential outcomes model with a trend.

[1] To solve for the trend adjustment I rely on the algebra of sums, including the following results:

$$\sum_{\bar{t}}^{T} (t - \bar{t}) = \frac{(T + 1)(T - 1)}{8} \qquad (B0.1)$$

C. Fitted Trend through a fixed-effects adjusted treatment variable (A^k)

The Frisch-Waugh expression for the linear trend through $\widetilde{D}_{it}$ for each unit i is:

$$A^k = \frac{\frac{1}{T}\sum_t \widetilde{D}_{it}\left(t-\bar{t}\right)}{\frac{1}{T}\sum_t\left(t-\bar{t}\right)^2} \qquad \forall \;\; i \in k \qquad (B1)$$

i. The group-specific covariance between $\widetilde{D}_{it}$ and a linear trend

The numerator of each of those regressions is:

$$\begin{aligned}\frac{1}{T}\sum_t \widetilde{D}_{it}\,(t-\bar{t}) &= \frac{1}{T}\left[\sum_t \left(D_{it}-\overline{\overline{D}}\right)(t-\bar{t}) - \sum_t \left(\overline{D}_i-\overline{\overline{D}}\right)(t-\bar{t}) - \sum_t \left(\overline{D}_t-\overline{\overline{D}}\right)(t-\bar{t})\right] \\ &= \frac{1}{T}\left[\sum_t D_{it}(t-\bar{t}) - \left(\overline{D}_i-\overline{\overline{D}}\right)\overbrace{\sum_t (t-\bar{t})}^{0} - \sum_t \overline{D}_t(t-\bar{t})\right] \\ &= \frac{1}{T}\sum_t \left(D_{it}-\overline{D}_t\right)(t-\bar{t}) \qquad (B2)\end{aligned}$$

$$\sum_{\bar{t}}^{T}(t-\bar{t})^2 = \frac{(T+1)(T-1)T}{24} \qquad (B0.2)$$

$$\frac{1}{T}\sum_{1}^{T}(t-\bar{t})^2 = \frac{2}{T}\sum_{\bar{t}}^{T}(t-\bar{t})^2 = \frac{(T+1)(T-1)}{12} \qquad (B0.3)$$

$$\sum_{t_k^*}^{T}(t-\bar{t}) = \frac{T^2}{2}\,V_k \qquad (B0.4)$$

$$\sum_{1}^{t_k^*-1}(t-\bar{t}) = -\frac{T^2}{2}\,V_k \qquad (B0.5)$$

$$\sum_{t_k^*}^{t_\ell^*-1}(t-\bar{t}) = \overbrace{\sum_{1}^{t_\ell^*-1}(t-\bar{t})}^{-\frac{T^2}{2}V_\ell} - \overbrace{\sum_{1}^{t_k^*-1}(t-\bar{t})}^{-\frac{T^2}{2}V_k} = \frac{T^2}{2}\,(V_k - V_\ell) \qquad (B0.6)$$

$$\sum_{t_k^*}^{\bar{t}}(\bar{t}-t) = \frac{(T+1)(T-1)}{8} - \frac{T^2}{2}\,V_k \qquad (B0.7)$$

We are essentially fitting a line through the difference between a dummy D_{it} and an increasing step function $\overline{D}_t$.[2] Consider $\overline{D}_t$: no units are treated before t_1^* ($\overline{D}_t = 0$); only units in group $k = 1$ are treated between t_1^* and $t_2^* - 1$ ($\overline{D}_t = n_1$); only units in groups $k = 1$ and $k = 2$ are treated between and t_2^* and $t_3^* - 1$ ($\overline{D}_t = n_1 + n_2$), etc. Therefore, for group k, the covariance between $(D_{it} - \overline{D}_t)$ and the time trend is:

$$\frac{1}{T}\left[\sum_{t_1^*}^{t_2^*-1}(-n_1)(t-\bar{t}) + \sum_{t_2^*}^{t_3^*-1}(-n_1-n_2)(t-\bar{t}) + \cdots + \sum_{t_k^*}^{t_{k+1}^*-1}\left(1-\sum_{j=1}^{k-1}n_j\right)(t-\bar{t}) + \cdots + \sum_{t_K^*}^{T}\left(1-\sum_{j=1}^{K}n_j\right)(t-\bar{t})\right] \qquad (B3)$$

Pull out a $(1 - n_k)$ from every sum starting at t_k^* or later, and then all other n_j's enter negatively in all sums starting at t_j^* or later, which yields:

$$\sum_t (D_{it} - \overline{D}_t)\frac{(t-\bar{t})}{T} = \begin{cases} \left[(1-n_k)\sum_{t=t_k^*}^{T}\frac{(t-\bar{t})}{T} - \sum_{j\neq k,U}\left\{n_j\sum_{t=t_j^*}^{T}\frac{(t-\bar{t})}{T}\right\}\right] & for\ i \in k \neq U \\ \left[-\sum_{j\neq U}\left\{n_j\sum_{t=t_j^*}^{T}\frac{(t-\bar{t})}{T}\right\}\right] & for\ i \in U \end{cases} \qquad (B4)$$

[2] The step function part is the same for all units because it equals the cumulative share of the sample that has $D_{it} = 1$. The dummy turns on earlier or later across timing groups (and not at all for untreated units). If it turns on right away, most of the time periods just have one added to them, so most period-to-period variation comes from the declining $-\overline{D}_t$ part. If it turns on only at the very end of the panel, the same thing is true: most period-to-period variation comes from the declining $-\overline{D}_t$ part. If it turns on in the middle, though, the earliest periods are zero or negative and the latest periods are all positive, and the slope will be relatively more positive. Therefore, it fits a more positive trend in the middle treated unit than earlier/later treated units.

Note that every term is negative for untreated units because their adjusted treatment variable is weakly falling. It is straightforward, if tedious, to use the properties of sums of integers to simplify these expressions. We know that every sum from t_k^* to T is positive because it only includes some, if any, of the negative $t-\bar{t}$ terms:

$$\sum_{t=t_k^*}^{T}(t-\bar{t}) = \left(\sum_{t=t_k^*}^{T} t\right) - \bar{t}\big(T-\ (t_k^*-1)\big)$$

The sum can be rewritten as running from 1 to $T-(t_k^*-1)$:

$$= \left(\sum_{j=1}^{T-(t_k^*-1)} j + (t_k^*-1)\right) - \bar{t}\big(T-\ (t_k^*-1)\big)$$

$$= \frac{\big(T-(t_k^*-1)\big)(T-(t_k^*-1)+1)}{2} + (t_k^*-1)(T-(t_k^*-1)) - \bar{t}\big(T-\ (t_k^*-1)\big)$$

$$= \left[\frac{(t_k^*-1)(T-(t_k^*-1))}{2}\right]$$

$$= \frac{T^2}{2}\bar{D}_k(1-\bar{D}_k) \equiv \frac{T^2}{2}V_k \qquad (B5)$$

Where to save space I define the variance of the treatment dummy in group k as $V_k = \bar{D}_k(1-\bar{D}_k)$, so the numerator equals:

$$\sum_t \big(D_{it}-\bar{D}_t\big)\frac{(t-\bar{t})}{T} = \begin{cases} \frac{T}{2}\left[(1-n_k)V_k - \sum_{j\neq k,U} n_j V_j\right] & for\ i \in k \neq U \\ \frac{T}{2}\left[-\sum_{j\neq U} n_j V_j\right] & for\ i \in U \end{cases} \qquad (B6)$$

ii. The variance of a linear trend

The denominator of the Frisch-Waugh expression is the variance of $t-\bar{t}$. We can rewrite the sums over demeaned time not as sums from 1 to T, but from $1-\bar{t} = 1-\frac{T+1}{2} = -\frac{T-1}{2} = -(\bar{t}-1)$ to

$T - \bar{t} = \frac{2T}{2} - \frac{T+1}{2} = \frac{T-1}{2} = (\bar{t} - 1)$. The variance then is just the sum of squared integers, so the negative and the positive portions are equal:

$$\frac{1}{T}\sum_{j=-(\bar{t}-1)}^{(\bar{t}-1)} j^2 = \frac{2}{T}\sum_{j=1}^{(\bar{t}-1)} j^2 = 2\left[\frac{\bar{t}(\bar{t}-1)(2\bar{t}-1)}{6T}\right] = \left[\frac{\bar{t}(\bar{t}-1)}{3}\right] = \frac{(T+1)(T-1)}{12} \qquad (B7)$$

Where the last equality follows because our assumption that T is odd implies that $\left(2\bar{t} - 1\right) = T$, and $\bar{t}\left(\bar{t} - 1\right) = \left(\frac{T+1}{2}\right)\left(\frac{T-1}{2}\right) = \frac{(T+1)(T-1)}{4}$.

iii. The group-specific linear trend in $\widetilde{D}_{it}$

Combining (B6) and (B7) shows that the coefficient on a linear trend fit through group k's treatment variable equals:

$$A^k = \frac{6T}{(T+1)(T-1)}\left[(1-n_k)V_k - \sum_{j\neq k,U} n_j V_j\right] \qquad (B8)$$

D. Solving for two-group trend-adjusted estimators

To write a two-group detrended estimator in the form of (27), we need to know the value of the fitted trends for sub-samples with just two groups. The only way that equation (B8) changes in the two-group case is that the sample shares refer to the two-group sample and not the full estimation sample. For example, in the two group estimator that uses groups k and U, group k's share is $\frac{n_k}{n_k+n_U} = \frac{n_k}{n_k+n_U}$, which implies:

$$A_{kU}^k = \frac{6T}{(T+1)(T-1)}\overbrace{\left(1 - \frac{n_k}{n_k+n_U}\right)}^{\frac{n_U}{n_k+n_U}} V_k = \frac{6T}{(T+1)(T-1)}\left(\frac{n_U}{n_k+n_U}\right)V_k \qquad (B9)$$

And

$$A_{kU}^U = -\frac{6T}{(T+1)(T-1)}\left(\frac{n_k}{n_k+n_U}\right)V_k \qquad (B10)$$

In a two-group timing-only estimator, we can calculate the

$$A_{k\ell}^{k} = \frac{6T}{(T+1)(T-1)}\left[\overbrace{\left(1-\frac{n_k}{n_k+n_\ell}\right)}^{\frac{n_\ell}{n_k+n_\ell}} V_k - \frac{n_\ell}{n_k+n_\ell} V_\ell\right]$$

$$= \frac{6T}{(T+1)(T-1)}\left(\frac{n_\ell}{n_k+n_\ell}\right)(V_k - V_\ell) \qquad (B11)$$

and

$$A_{k\ell}^{\ell} = -\frac{6T}{(T+1)(T-1)}\left(\frac{n_k}{n_k+n_\ell}\right)(V_k - V_\ell) \qquad (B12)$$

i. What do linear trends partial out of each two-group estimator?

With the definitions of these trend coefficients, it is straightforward to calculate the regression coefficient that relates the outcome to the fitted trends. For 2x2 estimators that compare a treated group to the untreated group, we have

$$\hat{\beta}_{kU}^{trend} = \frac{\frac{1}{NT}\sum_{i\in k,U}\sum_t A_{kU}^{i}(t-\bar{t})\, y_{it}}{\frac{1}{NT}\sum_i \sum_t \left[A_{kU}^{i}(t-\bar{t})\right]^2} \qquad (B13)$$

$$= \frac{\frac{1}{T} n_k A_{kU}^{k}\sum_t (t-\bar{t})\,\bar{y}_{kt} + \frac{1}{T} n_U A_{kU}^{U}\sum_t (t-\bar{t})\,\bar{y}_{Ut}}{\left(n_k {A_{kU}^{k}}^2 + n_U {A_{kU}^{U}}^2\right)\frac{1}{T}\sum_t[(t-\bar{t})]^2} \qquad (B14)$$

Where the second line follows because the sums over units, i, collapse to weighted sums over groups, k.

The sum of each group's time means, $\bar{y}_{kt}$, times $(t-\bar{t})$ has a positive part when $t > \bar{t}$, and a negative part when $t < \bar{t}$. Each portion equals a mean of y before or after $\bar{t}$ that use $|t-\bar{t}|$ as weights. We have:

$$\sum_t (t-\bar{t})\,\bar{y}_{jt} = \sum_{t=\bar{t}}^{T}(t-\bar{t})\left[\overbrace{\frac{\sum_{t=\bar{t}}^{T}(t-\bar{t})\,\bar{y}_{jt}}{\sum_{t=\bar{t}}^{T}(t-\bar{t})}}^{post-\bar{t}} - \overbrace{\frac{\sum_{t=1}^{\bar{t}}(\bar{t}-t)\,\bar{y}_{kt}}{\sum_{t=1}^{\bar{t}}(\bar{t}-t)}}^{pre-\bar{t}}\right]$$

$$= \frac{(T+1)(T-1)}{8}\Delta^*\bar{y}_j \qquad (B15)$$

Where the second line uses $\Delta^*\bar{y}_j$ to denote the weighted differences in y before and after $\bar{t}$, and the fact that the sum of integers $\sum_{j=1}^{\bar{t}-1} j = \frac{\bar{t}(\bar{t}-1)}{2} = \frac{(T+1)(T-1)}{8}$. The denominator also contains the variance of t, which was defined in equation (B7). Substituting (B7) and (B15) into (B14) yields

$$\hat{\beta}_{kU}^{trend} = \frac{\frac{(T+1)(T-1)}{8T}\left[\frac{n_k}{n_k+n_U}A_{kU}^{k}\Delta^*\bar{y}_k + \frac{n_U}{n_k+n_U}A_{kU}^{U}\Delta^*\bar{y}_U\right]}{\frac{(T+1)(T-1)}{12}\left(\frac{n_k}{n_k+n_U}{A_{kU}^{k}}^2 + \frac{n_U}{n_k+n_U}{A_{kU}^{U}}^2\right)} \qquad (B16)$$

Substituting the definitions of fitted trends from equations (B9) and (B10) shows that $\hat{\beta}_{kU}^{trend}$ is:

$$\frac{\frac{3}{4}\left(\frac{1}{n_k+n_U}\right)^2\left[n_k n_U V_k\,\Delta^*\bar{y}_k - n_U n_k V_k \Delta^*\bar{y}_U\right]}{\frac{(T+1)(T-1)}{12}\frac{36T^2}{(T+1)^2(T-1)^2}\left(\frac{n_k}{n_k+n_U}\left(\frac{n_U}{n_k+n_U}\right)^2 V_k^2 + \frac{n_U}{n_k+n_U}\left(\frac{n_k}{n_k+n_U}\right)^2 V_k^2\right)} \qquad (B17)$$

$$= \frac{\frac{(T+1)}{T}\frac{(T-1)}{T}\frac{1}{4}n_k n_U V_k\left(\frac{1}{n_k+n_U}\right)^2(\Delta^*\bar{y}_k - \Delta^*\bar{y}_U)}{V_k^2 n_k n_U\left(\frac{1}{n_k+n_U}\right)^2\left(\underbrace{\frac{n_U}{n_k+n_U}+\frac{n_k}{n_k+n_U}}_{1}\right)}$$

$$\hat{\beta}_{kU}^{trend} = \frac{(T+1)}{T}\frac{(T-1)}{T}\frac{V^*}{V_k}(\Delta^*\bar{y}_k - \Delta^*\bar{y}_U) \qquad (B18)$$

I replace the $\frac{1}{4}$ with V^*, which I define as the variance of a dummy that equals one after $\bar{t}$.

A similar derivation for a two-group timing-only estimator yields:

$$\hat{\beta}_{k\ell}^{trend} = \frac{(T+1)}{T}\frac{(T-1)}{T}\frac{V^*}{V_k - V_\ell}(\Delta^*\bar{y}_k - \Delta^*\bar{y}_\ell) \qquad (B19)$$

Equations (B18) and (B19) show that the coefficient that relates outcomes and a trend line fitted through each group's treatment variable equals the difference in a time-weighted average before and after time $\bar{t}$, scaled by a term measuring how close to the middle of the panel they are treated.

ii. *What is the R^2 from a two-group regression of $\widetilde{D}_{it}$ on $A^k(t-\bar{t})$?*

To calculate a trend-adjusted estimator, in this case for two groups, we need $\frac{R^2}{1-R^2}$ from the regression of the fixed-effects-adjusted treatment variable on the unit-specific trends, which equals $\frac{\widehat{var}(A^{j(i)}(t-\bar{t}))}{\widehat{var}(\widetilde{D}_{it})-\widehat{var}(A^{j(i)}(t-\bar{t}))}$. For the 2x2 DD comparing groups k and U we already have:

$$\widehat{var}\left(A^{j(i)}(t-\bar{t})\right)=\frac{(T+1)(T-1)}{12}\left(n_k {A_{kU}^{k}}^2+n_U {A_{kU}^{U}}^2\right)$$

$$=3\frac{T}{T+1}\frac{T}{T-1}n_k n_U V_k^2\left(\frac{1}{n_k+n_U}\right)^2$$

$$\widehat{var}\left(\widetilde{D}_{it}\right)=\left(\frac{1}{n_k+n_U}\right)^2 n_k n_U V_k$$

$$\frac{R_{kU}^2}{1-R_{kU}^2}=\frac{\widehat{var}\,(A^{j(i)}(t-\bar{t}))}{\widehat{var}\left(\widetilde{D}_{it}\right)-\widehat{var}\,(A^{j(i)}(t-\bar{t}))}=\frac{\overbrace{3\frac{T}{T+1}\frac{T}{T-1}V_k}^{R_{kU}^2}}{1-3\frac{T}{T+1}\frac{T}{T-1}V_k}\qquad (B20)$$

For a timing-only comparison between groups k and ℓ we have:

$$\frac{R_{k\ell}^2}{1-R_{k\ell}^2}=\frac{\overbrace{3\frac{T}{T+1}\frac{T}{T-1}\frac{(V_k-V_\ell)^2}{(\bar{D}_k-\bar{D}_\ell)\left(1-(\bar{D}_k-\bar{D}_\ell)\right)}}^{R_{k\ell}^2}}{1-3\frac{T}{T+1}\frac{T}{T-1}\frac{(V_k-V_\ell)^2}{(\bar{D}_k-\bar{D}_\ell)\left(1-(\bar{D}_k-\bar{D}_\ell)\right)}}\qquad (B21)$$

iii. *What are the detrended 2x2 estimators?*

Equation (27) already shows what each 2x2 detrended estimator is in terms of $\hat{\beta}_{ab}^{2x2}$, $\hat{\beta}_{ab}^{trend}$, and R_{ab}^2, but equations (B16), (B17), (B20), and (B21) allow us to write it explicitly in terms of sample means and shares:

$$\hat{\beta}_{kU}^{2x2,trend} = \left(\overline{y}_k^{POST(k)} - \overline{y}_k^{PRE(k)}\right) - \left(\overline{y}_U^{POST(k)} - \overline{y}_U^{PRE(k)}\right)$$

$$+ \frac{3\frac{T}{T+1}\frac{T}{T-1}V_k}{1-3\frac{T}{T+1}\frac{T}{T-1}V_k}\left[\left(\left(\overline{y}_k^{POST(k)} - \overline{y}_k^{PRE(k)}\right) - \left(\overline{y}_U^{POST(k)} - \overline{y}_U^{PRE(k)}\right)\right)\right.$$

$$\left. - \frac{(T+1)}{T}\frac{(T-1)}{T}\frac{V^*}{V_k}(\Delta^*\overline{y}_k - \Delta^*\overline{y}_U)\right] \qquad (B22)$$

$$\hat{\beta}_{k\ell}^{2x2,trend} = \left[\frac{1-\overline{D}_k}{1-(\overline{D}_k - \overline{D}_\ell)}\right]\left[\left(\overline{y}_k^{MID(k,\ell)} - \overline{y}_k^{PRE(k)}\right) - \left(\overline{y}_\ell^{MID(k,\ell)} - \overline{y}_\ell^{PRE(k)}\right)\right]$$

$$+ \left[\frac{\overline{D}_\ell}{1-(\overline{D}_k - \overline{D}_\ell)}\right]\left[\left(\overline{y}_\ell^{POST(\ell)} - \overline{y}_\ell^{MID(k,\ell)}\right) - \left(\overline{y}_k^{POST(\ell)} - \overline{y}_k^{MID(k,\ell)}\right)\right]$$

$$+ \frac{3\frac{T}{T+1}\frac{T}{T-1}\frac{(V_k - V_\ell)^2}{(\overline{D}_k - \overline{D}_\ell)\left(1-(\overline{D}_k - \overline{D}_\ell)\right)}}{1-3\frac{T}{T+1}\frac{T}{T-1}\frac{(V_k - V_\ell)^2}{(\overline{D}_k - \overline{D}_\ell)\left(1-(\overline{D}_k - \overline{D}_\ell)\right)}}\left[\left[\frac{1-\overline{D}_k}{1-(\overline{D}_k - \overline{D}_\ell)}\right]\left[\left(\overline{y}_k^{MID(k,\ell)}\right.\right.\right.$$

$$\left.\left. - \overline{y}_k^{PRE(k)}\right) - \left(\overline{y}_\ell^{MID(k,\ell)} - \overline{y}_\ell^{PRE(k)}\right)\right]$$

$$+ \left[\frac{\overline{D}_\ell}{1-(\overline{D}_k - \overline{D}_\ell)}\right]\left[\left(\overline{y}_\ell^{POST(\ell)} - \overline{y}_\ell^{MID(k,\ell)}\right) - \left(\overline{y}_k^{POST(\ell)} - \overline{y}_k^{MID(k,\ell)}\right)\right]$$

$$\left. - \frac{(T+1)}{T}\frac{(T-1)}{T}\frac{V^*}{V_k - V_\ell}(\Delta^*\overline{y}_k - \Delta^*\overline{y}_\ell)\right] \qquad (B23)$$

Note that because $R_{k\ell}^2$ has $V_k - V_\ell$ in the numerator, it equals zero when the variance of treatment is the same for two timing groups. This happens when they are treated equally close to the ends the panel, in which case unit-specific linear trends have no effect on the 2x2 point estimate.

E. Solving for a general trend-adjusted estimator

We know that:

$$\hat{\beta}_{trend}^{DD} = \frac{\widehat{cov}\left(\widetilde{D}_{it}, y_{it}\right) - \widehat{cov}\left(A^{j(i)}(t-\overline{t}), y_{it}\right)}{\widehat{var}\left(\widetilde{D}_{it}\right) - \widehat{var}\left(A^{j(i)}\left(t-\overline{t}\right)\right)}$$

The first covariance comes from the typical two-way fixed effects estimator. We can write the second covariance as a function of the $\Delta^*\bar{y}_j$ terms:

$$\frac{1}{NT}\sum_i A^{j(i)}\sum_t (t-\bar{t})y_{it} = \frac{n_U A^U}{T}\sum_t (t-\bar{t})\bar{y}_{Ut} + \sum_k \frac{n_k A^k}{T}\sum_t (t-\bar{t})\bar{y}_{kt}$$

$$= \frac{(T+1)(T-1)}{8T}\left[n_U A^U \Delta^*\bar{y}_U + \sum_k n_k A^k \Delta^*\bar{y}_k\right] \qquad (B24)$$

Then using equation (B7) we can substitute for A^U and A^k:

$$= \frac{3}{4}\left[n_U\left(-\sum_{k\neq U} n_j V_j\right)\Delta^*\bar{y}_U + \sum_k n_k\left(\overbrace{(1-n_k)}^{n_U+\sum_{j\neq k,U} n_j} V_k - \sum_{j\neq k,U} n_j V_j\right)\Delta^*\bar{y}_k\right]$$

$$= \frac{3}{4}\left[\sum_{k\neq U} n_U n_j V_j\left(\Delta^*\bar{y}_j - \Delta^*\bar{y}_U\right) + \overbrace{\sum_k n_k \Delta^*\bar{y}_k \sum_{j\neq k,U} n_j\left(V_j - V_k\right)}^{each\ pair\ appears\ twice}\right]$$

$$= 3\left[\sum_{k\neq U} n_U n_j V^* V_j\left(\Delta^*\bar{y}_j - \Delta^*\bar{y}_U\right) + \sum_{k\neq U}\sum_{\ell>k} n_k n_\ell V^*(V_k - V_\ell)(\Delta^*\bar{y}_k - \Delta^*\bar{y}_\ell)\right] \qquad (B25)$$

Where again, I replace $\frac{1}{4}$ with V^*, the variance of a dummy that turns on at $\bar{t}$. We can multiply each term by $\frac{V_j}{V_j}$ (or $\frac{V_k-V_\ell}{V_k-V_\ell}$) as well as $\frac{(T+1)}{T}\frac{(T-1)}{T}\frac{T}{T+1}\frac{T}{T-1}$, and write equation (B25) as a function of the two-group $\hat{\beta}^{trend}$ terms and the R^2 values:

$$= \sum_{k\neq U} n_U n_j V_j \overbrace{\left(3\frac{T}{T+1}\frac{T}{T-1}V_j\right)}^{R^2_{jU}} \overbrace{\left(\frac{(T+1)}{T}\frac{(T-1)}{T}\frac{V^*}{V_j}\left(\Delta^*\bar{y}_j - \Delta^*\bar{y}_U\right)\right)}^{\hat{\beta}^{trend}_{jU}}$$

$$+\sum_{k\neq U}\sum_{\ell>k} n_k n_\ell (\bar{D}_k-\bar{D}_\ell)\big(1$$

$$-(\bar{D}_k-\bar{D}_\ell)\big)\overbrace{\left(3\frac{T}{T+1}\frac{T}{T-1}\frac{(V_k-V_\ell)^2}{(\bar{D}_k-\bar{D}_\ell)\big(1-(\bar{D}_k-\bar{D}_\ell)\big)}\right)}^{R^2_{k\ell}}\left(\frac{(T+1)}{T}\frac{(T-1)}{T}\frac{V^*}{(V_k-V_\ell)}(\Delta^*\bar{y}_k\right.$$

$$\left.-\Delta^*\bar{y}_\ell)\right) \qquad (B26)$$

So

$$\sum_{k\neq U} n_U n_j V_j R^2_{jU}\hat{\beta}^{trend}_{jU} + \sum_{k\neq U}\sum_{\ell>k} n_k n_\ell (\bar{D}_k-\bar{D}_\ell)\big(1-(\bar{D}_k-\bar{D}_\ell)\big)R^2_{k\ell}\hat{\beta}^{trend}_{k\ell} \qquad (B27)$$

Note that the weights on each $R^2_{jU}\hat{\beta}^{trend}_{jU}$ term are the *same* as in the DD decomposition theorem, so the whole numerator is:

$$\sum_{k\neq U} n_U n_j V_j\big(\hat{\beta}^{2x2}_{jU}-R^2_{jU}\hat{\beta}^{trend}_{jU}\big) + \sum_{k\neq U}\sum_{\ell>k} n_k n_\ell (\bar{D}_k-\bar{D}_\ell)\big(1-(\bar{D}_k-\bar{D}_\ell)\big)\big(\hat{\beta}^{2x2}_{k\ell}-R^2_{k\ell}\hat{\beta}^{trend}_{k\ell}\big)$$

$$=\sum_{k\neq U} n_U n_j V_j\big(1-R^2_{jU}\big)\left(\hat{\beta}^{2x2}_{jU}+\frac{R^2_{jU}}{1-R^2_{jU}}\big[\hat{\beta}^{2x2}_{jU}-\hat{\beta}^{trend}_{jU}\big]\right)$$

$$+\sum_{k\neq U}\sum_{\ell>k} n_k n_\ell (\bar{D}_k-\bar{D}_\ell)\big(1-(\bar{D}_k-\bar{D}_\ell)\big)(1-R^2_{k\ell})\left(\hat{\beta}^{2x2}_{k\ell}\right.$$

$$\left.+\frac{R^2_{k\ell}}{1-R^2_{k\ell}}\big[\hat{\beta}^{2x2}_{k\ell}-\hat{\beta}^{trend}_{k\ell}\big]\right)$$

$$\sum_{k\neq U} n_U n_j V_j \left(1-R_{jU}^2\right)\hat{\beta}_{jU}^{2x2,trend}$$

$$+\sum_{k\neq U}\sum_{\ell>k} n_k n_\ell (\overline{D}_k-\overline{D}_\ell)\left(1-(\overline{D}_k-\overline{D}_\ell)\right)(1-R_{k\ell}^2)\hat{\beta}_{k\ell}^{2x2,trend} \quad (B28)$$

F. The detrended denominator

Substituting D for y in (B28) shows that the denominator of $\hat{\beta}_{trend}^{DD}$ equals the sum of the terms that multiply each $\hat{\beta}$ in (B28):

$$\widehat{var}\left(\widehat{D}_{it}\right)=\sum_{k\neq U} n_U n_j V_j\left(1-R_{jU}^2\right)+\sum_{k\neq U}\sum_{\ell>k} n_k n_\ell(\overline{D}_k-\overline{D}_\ell)\left(1-(\overline{D}_k-\overline{D}_\ell)\right)(1-R_{k\ell}^2) \quad (B29)$$

Combining (B29) and (B28) proves the following decomposition theorem for the detrended estimator:

Theorem B1. Detrended Difference-in-Differences Decomposition Theorem

The OLS estimate of $\widehat{\beta}^{DD}$ in a two-way fixed-effects model with unit-specific linear time trends is a weighted average of all possible two-group detrended DD estimators.

$$\widehat{\beta}_{trend}^{DD}=\sum_{k\neq U}\sigma_{kU}\overbrace{\left(\widehat{\beta}_{kU}^{2x2}+\frac{R_{kU}^2}{1-R_{kU}^2}\left[\widehat{\beta}_{kU}^{2x2}-\hat{\beta}_{kU}^{trend}\right]\right)}^{\hat{\beta}_{kU}^{2x2,trend}}+\sum_{k\neq U}\sum_{\ell>k}\sigma_{k\ell}\overbrace{\left(\hat{\beta}_{k\ell}+\frac{R_{k\ell}^2}{1-R_{k\ell}^2}\left[\hat{\beta}_{k\ell}-\hat{\beta}_{k\ell}^{trend}\right]\right)}^{\hat{\beta}_{k\ell}^{2x2,trend}} \quad (B30)$$

$\widehat{\beta}_{kU}^{2x2}$ is defined in theorem 1 and $\hat{\beta}_{k\ell}$ is defined in equation (8). The two-group trend terms are:

$$\hat{\beta}_{kU}^{trend}=\frac{(T^2-1)}{4T^2}\frac{(\Delta^*\bar{y}_k-\Delta^*\bar{y}_U)}{\overline{D}_k(1-\overline{D}_k)} \quad (B31)$$

$$\hat{\beta}_{k\ell}^{trend}=\frac{(T^2-1)}{4T^2}\frac{(\Delta^*\bar{y}_k-\Delta^*\bar{y}_\ell)}{\overline{D}_k(1-\overline{D}_k)-\overline{D}_\ell(1-\overline{D}_\ell)} \quad (B32)$$

$\Delta^*\bar{y}_j$ is difference in the average of y before and after $\bar{t}$ weighted by $|t-\bar{t}|$. The weights on the two-group terms are:

$$\sigma_{kU}=\frac{n_U n_k\overline{D}_k(1-\overline{D}_k)(1-R_{kU}^2)}{\widehat{var}\,(\widehat{D}_{it})} \quad (B33)$$

$$\sigma_{k\ell} = \frac{n_k n_\ell(\bar{D}_k - \bar{D}_\ell)(1-(\bar{D}_k - \bar{D}_\ell))(1-R^2_{k\ell})}{\widehat{var}\,(\hat{D}_{it})} \qquad (B34)$$

$$R^2_{kU} = \frac{3T^2}{(T^2-1)}\bar{D}_k(1-\bar{D}_k) \qquad (B35)$$

$$R^2_{k\ell} = \frac{3T^2}{(T^2-1)}\frac{(\bar{D}_k(1-\bar{D}_k) - \bar{D}_\ell(1-\bar{D}_\ell))^2}{(\bar{D}_k - \bar{D}_\ell)\big(1-(\bar{D}_k - \bar{D}_\ell)\big)} \qquad (B36)$$

and $\sum_{k\neq U}\sigma_{kU} + \sum_{k\neq U}\sum_{\ell>k}\sigma_{k\ell} = 1.$

The weights closely resemble the weights in the main DD decomposition theorem, but also incorporate how well linear trends fit the treatment variable in each pair via the $\left(1-R^2_{jU}\right)$. Trends thus change a DD estimate in two ways. First, they change the value of each 2x2 component by netting out a trend. Second, they alter the weights placed on each two-group comparison. Because it is a function of the variance of the treatment dummy, $R^2_{jU} = 3\frac{T}{T+1}\frac{T}{T-1}V_j$, will be highest for groups treated toward the middle of the panel. Trends thus downweight these terms relative to an unadjusted estimator.

Note that the detrended decomposition is in terms of *two-group* DD's not 2x2 DD. I did not attempt to decompose $\hat{\beta}^{trend}_{k\ell}$ into pieces that would be subtracted from the two components of a given timing comparison, $\mu_{k\ell}\,\widehat{\beta}^{2x2,k}_{k\ell}$ and $(1-\mu_{k\ell})\,\widehat{\beta}^{2x2,\ell}_{k\ell}$.

G. When do unit-specific trends work and when do they fail?

This detrended decomposition result makes it straightforward to learn what problems unit-specific time-trends can fix and what problems they create. Rewrite counterfactual outcomes as a constant, c^k, plus a linear component $a^k(t-\bar{t})$, plus mean-zero deviations from that trend, dY^0_{kt}; and group-time ATTs as a time average as in equation (9) plus deviations, $dATT_{kt}(W)$:[3]

[3] These deviations have a zero mean during period W, but not in other post-treatment windows. When $W = POST(k)$, for example, $ATT_k(W) = ATT_k(POST(k))$ and the mean of $dATT_k(POST(k))$ does equal 0.

$$\bar{y}_{kt} = \overbrace{\left[c^k + a^k(t-\bar{t}) + dY^0_{kt}\right]}^{Y^0_{kt}} + D_{kt}\overbrace{\left[ATT_k(W) \quad + dATT_{kt}(W)\right]}^{ATT_k(t)} \qquad (B37)$$

Here I analyze consistency for a two-group DD comparing treated to untreated units by substituting this definition of potential outcomes into the two-group version.[4] For comparison, rewrite β^{2x2}_{kU} from equation (11a):

$$\beta^{2x2}_{kU} = ATT_k(POST(k)) + \frac{(a^k - a^U)\,T}{2} + \Delta_{kU}dY^0 \qquad (B38)$$

Bias comes from differential linear trends, $\frac{(a^k-a^U)\,T}{2}$, and differential deviations from those trends, $\Delta_{kU}dY^0 = \left(\overline{dY}_k^{0,POST(k)} - \overline{dY}_k^{0,PRE(k)}\right) - \left(\overline{dY}_U^{0,POST(k)} - \overline{dY}_U^{0,PRE(k)}\right)$. The corresponding estimator with unit-specific trends equals:

$$\beta^{2x2,trend}_{kU} = ATT_k(POST(k)) - \frac{3}{4}\frac{\Delta^* dATT_k}{1-R^2_{kU}} + \frac{\left[\Delta_{kU}dY^0 - \frac{3}{4}\Delta^*_{kU}dY^0\right]}{1-R^2_{kU}} \qquad (B39)$$

Equation (B39) shows that unit-specific trends successfully eliminate bias from differential linear trends—no terms involving $a^k - a^U$ appear—but can introduce bias from two other sources.

Unit-specific trends over control when treatment effects grow over time. Equation (B39) subtracts $\Delta^* dATT_k = \frac{\sum_{t=\bar{t}}^{T}(t-\bar{t})D_{kt}dATT_{kt}(W)}{\sum_{t=\bar{t}}^{T}(t-\bar{t})} - \frac{\sum_{t=1}^{\bar{t}}(\bar{t}-t)D_{kt}dATT_{kt}(W)}{\sum_{t=1}^{\bar{t}}(\bar{t}-t)}$, the change in time-weighted mean of treatment effect deviations before and after $\bar{t}$. $\Delta^* dATT_k$ would put the most weight on the largest effects at T, making $\Delta^* dATT_k > 0$ and biasing the treatment effect toward zero. This shows

[4] $\hat{\beta}^{DD}_{trend}$ is a proper weighed average of two-group estimators like (27), but because treatment variances and the scaling factor $\frac{T^2}{(T^2-1)}$ appear in both $\hat{\beta}^{trend}_{gh}$ and R^2_{gh}, the detrended terms simplify to:

$$\hat{\beta}^{2x2,trend}_{kU} = \hat{\beta}^{2x2}_{kU} - \frac{3}{4}\left(\Delta^*\bar{y}_j - \Delta^*\bar{y}_U\right)$$

$$\hat{\beta}^{2x2,trend}_{k\ell} = \hat{\beta}^{2x2}_{k\ell} - \frac{3}{4}\frac{(V_k - V_\ell)}{(\bar{D}_k - \bar{D}_\ell)\left(1-(\bar{D}_k - \bar{D}_\ell)\right)}\left(\Delta^*\bar{y}_k - \Delta^*\bar{y}_\ell\right)$$

formally that unit-specific trends conflate time-varying treatment effects with underlying trends and take out "too much."

Unit-specific trends are also sensitive to deviations in unobservables at the beginning or end of the panel. Equation (34) also subtracts $\Delta_{kU}^{*} dY^0 = \left[\frac{\sum_{t=\bar{t}}^{T}(t-\bar{t})dY_{kt}^0}{\sum_{t=\bar{t}}^{T}(t-\bar{t})} - \frac{\sum_{t=1}^{\bar{t}}(\bar{t}-t)dY_{kt}^0}{\sum_{t=1}^{\bar{t}}(\bar{t}-t)}\right] - \left[\frac{\sum_{t=\bar{t}}^{T}(t-\bar{t})dY_{Ut}^0}{\sum_{t=\bar{t}}^{T}(t-\bar{t})} - \frac{\sum_{t=1}^{\bar{t}}(\bar{t}-t)dY_{Ut}^0}{\sum_{t=1}^{\bar{t}}(\bar{t}-t)}\right]$, the change in time-weighted means of trend deviations in counterfactual outcomes before and after $\bar{t}$. If these deviations lead to bias in the unadjusted estimator ($\Delta_{kU} dY^0 \neq 0$), unit-specific trends can help if the time-weighted changes, $\Delta_{kU}^{*} dY^0$, are of a similar magnitude. Unfortunately, this term may also induce bias. If these deviations average out before and after t_k^*, but the time-weighted averages do not, then unit-specific trends will generate bias by emphasizing the end points.[5]

[5] We cannot draw firm conclusions about whether unit-specific trends lead to more or less bias without additional assumptions on the dY_{kt}^0 terms. It is clear, though, that these estimates are "noisier" because $\widehat{var}\left(\widehat{D}_{it}\right) < \widehat{var}\left(\widetilde{D}_{it}\right)$ so the trend-adjusted estimator exploits less variation and has higher standard errors. In terms of point estimates, however, the imprecision of trend-adjusted estimates increases the probability that a given sample's result is spurious. Assume that there are no linear trends ($a^k = 0$) and that dY_{kt}^0 is a classical error term so that in repeated sampling it averages to zero on any subset of time with or without the time weights. In repeated sampling, the trend-adjusted estimator in this case is unbiased. But by putting disproportionate weight on errors at the ends of the panel, the probability that the trend adjusted estimator yields a biased estimate in a given sample is high.

III. APPENDIX C: GROUP-SPECIFIC LINEAR PRE-TRENDS

Using data from before t_1^*, estimate a pre-trend in y_{it} for each timing group.[6] The potential outcome model in (B37) shows that this slope will equal the linear component of unobservables plus a linear approximation to trend deviations before before t_1^*:

$$\frac{\widehat{cov}(\bar{y}_{kt}, t-\bar{t}|t<t_1^*)}{\widehat{var}(t-\bar{t}|t<t_1^*)} = a^k + \overbrace{\frac{\widehat{cov}\left(\overline{dY}_{kt}^0, t-\bar{t}|t<t_1^*,\right)}{\widehat{var}(t-\bar{t}|t<t_1^*)}}^{da^k} \qquad (C1)$$

The fitted pre-trends equal the linear component of unobservables, a^k, plus any non-linear deviations that vary systematically with time in the pre-period, da^k. Removing this trend from the full panel means that the outcome variable will be:

$$\bar{y}_{kt}^* = -da^k(t-\bar{t}) + \overline{dY}_{kt}^0 + D_{kt}ATT_k(t) \qquad (C2)$$

Where $\overline{da}^k$ is the amount by which group k's pre-trend estimate deviates from the true linear component, a^k. The two-group estimate then equals:

$$\hat{\beta}_{kU}^{2x2,pretrend} = ATT_k(POST(k)) - \frac{\left(\overline{da}^k - \overline{da}^U\right)T}{2} + \Delta_{kU}dY^0\left(POST(k), PRE(k)\right) \quad (C3)$$

Bias from non-linear unobservables remains because a linear detrending strategy does not attempt to control for it. (Controlling for unit-specific linear trends suffers from this source of bias as well, so partialling out pre-trends is not obviously any worse in this regard.) Bias from linear pre-trends, however, is addressed. Unless pre-treatment nonlinearities lead to the wrong slope estimates (which becomes less likely with more pre-tratement periods), $\frac{(\overline{da}^k - \overline{da}^U)\,T}{2}$ should be smaller than $\frac{(a^k - a^U)\,T}{2}$.

[6] This detrending, and actually the strategy of adding linear trends as a control, only needs to happen at the group level, not the individual level. Point estimates are the same while standard errors are generally smaller, especially errors from misspecification of the trend during the extrapolated period which well tend to be smaller for groups of units compared to individual units.

IV. APPENDIX D: DERIVATION OF THE DECOMPOSITION WITH DISAGGREGATED TIME FIXED EFFECTS

With disaggregated time fixed effects, the only difference is that there are now R-specific year effects taken out, so the covariance equals:

$$\widehat{cov}\left(\tilde{y}_{it}, \tilde{D}_{it}\right) = \sum_R \frac{N_R}{N}[\frac{1}{N_R T}\sum_{i\in R}\sum_t \left(y_{it} - \bar{\bar{y}}_R\right)\left(D_{it} - \bar{\bar{D}}_R\right) - \frac{1}{N_R}\sum_{i\in R}\left(\bar{y}_i - \bar{\bar{y}}_R\right)\left(\bar{D}_i - \bar{\bar{D}}_R\right)$$

$$-\frac{1}{T}\sum_t \left(\bar{y}_t^R - \bar{\bar{y}}_R\right)\left(\bar{D}_t^R - \bar{\bar{D}}_R\right)] \qquad (D1)$$

The simplification does not change, but now it happens within each R. The first two terms are:

$$\sum_R n^R \left[\sum_k n_k^R \bar{D}_k\left(1 - \bar{D}_k\right)\left(\bar{y}_{k,R}^{POST(k)} - \bar{y}_{k,R}^{PRE(k)}\right)\right] \qquad (D2)$$

And the third term is:

$$\sum_R n^R \left[\sum_k \sum_{\ell \neq U} n_\ell^R n_k^R \bar{D}_\ell\left(1 - \bar{D}_\ell\right)\Delta(y_{\ell,R}, t_k^*)\right] \qquad (D3)$$

Combining equations (D2) and (D3) shows that the covariance equals:

$$\sum_R n^R \left[\sum_{k\neq U} n_k^R \bar{D}_k(1 - \bar{D}_k)\Delta(y_{k,R}, t_k^*) - \sum_\ell \sum_{k\neq C} n_\ell^R n_k^R \bar{D}_k(1 - \bar{D}_k)\Delta(y_{\ell,R}, t_k^*)\right] \qquad \text{(D4)}$$

Following the derivation in appendix A, we know the covariance within each region essentially follows the decomposition theorem and that the whole estimate is divided by $\hat{V}(\tilde{D}_{it}^{/R})$, the variance of D conditional on unit and region-by-year fixed effects:

$$\sum_R \frac{n^R}{\hat{V}(\tilde{D}_{it}^{/R})}[\sum_{k\neq U} \bar{D}_k(1 - \bar{D}_k) n_k^R n_U^R \hat{\beta}_{kU,R}^{2x2}$$

$$+ \sum_{k\neq U}\sum_{\ell > k}(\bar{D}_k - \bar{D}_\ell)\left(1 - (\bar{D}_k - \bar{D}_\ell)\right) n_k^R n_\ell^R \left[\mu_{k\ell}\hat{\beta}_{k\ell,R}^{2x2,k} + (1 - \mu_{k\ell})\hat{\beta}_{k\ell,R}^{2x2,\ell}\right] \qquad \text{(D5)}$$

Consider the weight on the terms that compare group k to group U within each region:

$$\frac{\overline{D}_k(1-\overline{D}_k)}{\hat{V}(\widetilde{D}_{it}^{/R})}\sum_R n^R n_k^R n_U^R\,\widehat{\beta}_{kU,R}^{2x2} = \frac{\overline{D}_k(1-\overline{D}_k)\sum_R n^R n_k^R n_U^R}{\hat{V}(\widetilde{D}_{it}^{/R})}\sum_R \frac{n^R n_k^R n_U^R}{\sum_R n^R n_k^R n_U^R}\widehat{\beta}_{kU,R}^{2x2}$$

$$= \overbrace{\frac{\overline{D}_k(1-\overline{D}_k)\sum_R n^R n_k^R n_U^R}{\hat{V}(\widetilde{D}_{it}^{/R})}}^{s_{kU}^{R\times t}}\overbrace{\sum_R \rho_{kU}^R\,\widehat{\beta}_{kU,R}^{2x2}}^{\widehat{\beta}_{kU,R\times t}^{2x2}} \qquad \text{(D6)}$$

The same type of expression holds for the other two-group terms. The analogy of equation (7) for the disaggregated fixed effects specification is:

$$\widehat{\beta}_{R\times t}^{DD} = \sum_{k\neq U} s_{kU}^{R\times t}\,\widehat{\beta}_{kU}^{2x2,R\times t} + \sum_{k\neq U}\sum_{\ell>k} s_{k\ell}^{R\times t}\,[\mu_{k\ell}\,\widehat{\beta}_{k\ell,R\times t}^{2x2,k} + (1-\mu_{k\ell})\,\widehat{\beta}_{k\ell,R\times t}^{2x2,\ell}] \qquad (D7)$$

V. APPENDIX E: TREATMENTS THAT TURN OFF

Many treatments turn on *and* off. To characterize how these models work, write the treatment dummy as the difference between a dummy for whether the treatment has ever turned on and a dummy for whether the treatment has ever turned off: $W_{it} = D_{it}^{ON} - D_{it}^{OFF}$. The fixed-effects adjusted version of W_{it} is just $\widetilde{D}_{it}^{ON} - \widetilde{D}_{it}^{OFF}$, so the DD estimate equals:

$$\hat{\beta}_{ON,OFF}^{DD} = \frac{\widehat{cov}\left(\widetilde{y}_{it}, \widetilde{D}_{it}^{ON}\right) - \widehat{cov}\left(\widetilde{y}_{it}, \widetilde{D}_{it}^{OFF}\right)}{\widehat{var}\,(\widetilde{W}_{it})} \qquad (E1)$$

The fact that the decomposition theorem applies to both terms in the numerator shows that DD estimates combine estimated effects when treatments turn on and when they turn off. To show how these comparisons are made, I consider a 2x2 comparison between a group, j, whose treatment turns on and off and another, U, that never receives treatment. Both covariances have the DD form, and are multiplied by the product of the sample shares and the treatment variance:

$$n_j n_U \overline{D}_j^{ON}\left(1 - \overline{D}_j^{ON}\right)\left[\left(\overline{y}_j^{POST-ON(j)} - \overline{y}_j^{PRE(j)}\right) - \left(\overline{y}_U^{POST-ON(j)} - \overline{y}_U^{PRE(j)}\right)\right] +$$
$$n_j n_U \overline{D}_j^{OFF}\left(1 - \overline{D}_j^{OFF}\right)\left[\left(\overline{y}_j^{OFF(j)} - \overline{y}_j^{PRE-OFF(j)}\right) - \left(\overline{y}_U^{OFF(j)} - \overline{y}_U^{PRE-OFF(j)}\right)\right] \qquad (E2)$$

The time periods used in these comparisons overlap, though. A portion $\frac{\overline{D}_j^{ON} - \overline{D}_j^{OFF}}{\overline{D}_j^{ON}}$ the period when $D_{it}^{ON} = 1$ contains the "middle window" when $W_{it} = D_{it}^{ON} - D_{it}^{OFF} = 1$ and a portion $\frac{\overline{D}_j^{OFF}}{\overline{D}_j^{ON}}$ contains the period after treatment turns off, $W_{it} = 1 - 1 = 0$. Similarly, the period when $D_{it}^{OFF} = 0$ contains the same "middle window" and the period before treatment starts and $W_{it} = 0 - 0 = 0$. This matches the two-group timing-only estimator, and equation (E2) can be written in the same way:

$$\hat{\beta}^{2x2,kU}_{ON,OFF} = \mu^*_{jU} \overbrace{\left[\left(\overline{y}_j^{ON(j)} - \overline{y}_j^{PRE(j)}\right) - \left(\overline{y}_U^{ON(j)} - \overline{y}_U^{PRE(j)}\right)\right]}^{\hat{\beta}^{2x2,ON}_{jU}} +$$

$$(1 - \mu^*_{jU}) \overbrace{\left[\left(\overline{y}_j^{ON(j)} - \overline{y}_j^{OFF(j)}\right) - \left(\overline{y}_U^{ON(j)} - \overline{y}_U^{OFF(j)}\right)\right]}^{\hat{\beta}^{2x2,OFF}_{jU}} \qquad (E3)$$

$$\mu^*_{jU} = \frac{1 - \overline{D}_j^{ON}}{1 - \left(\overline{D}_j^{ON} - \overline{D}_j^{OFF}\right)} \qquad (E4)$$

Equations (E3) and (E4) show that a DD model with a treatment that turns off averages the 2x2 DD that compares the middle window to the pre-treatment period and the 2x2 DD that compares the middle window to the period when treatment has turned off. The weights come from how close to the middle of the panel is the treatment's start date versus its end date. The analogous version with two treated groups would be more complicated and depend on the overlap between the periods when each unit's treatment has not yet turned on, is on, or has already turned off.

Note that while $\hat{\beta}^{2x2,ON}_{jU}$ is a typical 2x2 estimator, $\hat{\beta}^{2x2,OFF}_{jU}$ compares outcomes in the period when treatment is turned on to a period when treatment is turned off but, by definition, has been on in the past. If treatment effects persist after treatment stops, we do not observe the counterfactual outcome in the $OFF(j)$ period. Outcomes may not actually change when treatment turns off, making $\hat{\beta}^{2x2,OFF}_{jU}$ and therefore the overall DD estimate too small. A version of the DD decomposition theorem and plots like figure 4 could be used to analyze the extent to which the estimates based on treatments turning off differ systematically from estimates based on treatments turning on.