1. 1. If $\phi(0) \neq \phi(1)$ then $\phi(X_i)$ represents a difference in the growth of Y_{it} across units with $X_i = 1/0$. λ can be interpreted as the cumulative ATT induced by D_i switching on from time g.
   2. Denote $P_\ell^t = 1[t = g+\ell]$ for $\ell \in \{-g+1, \ldots, T-g\}$ and set $P_{-1}^t = 0$ to avoid multicollinearity. $Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \in L} D_i P_\ell^t \tau_\ell + \epsilon_{it}$, under SUTVA, no anticipation $Y_{it}(0) = Y_{it}(1) \quad \forall t < g, i$ with $D_i = 1$ and parallel trends $E[Y_{it'}(0) - Y_{it}(0)|D_i = 1] = E[Y_{it'}(0) - Y_{it}(0)|D_i = 0], \forall t \neq t'$, can identify λ as τ_ℓ for $\ell \geq 0$, i.e. cumulative ATT $E[Y_{i,g+\ell}(1) - Y_{i,g+\ell}(0)|D_i = 1]$ (no pre-trend $\tau_\ell = 0$ for $\ell < -1$). Given the DGP, assuming $\phi(0) = \phi(1)$ is necessary for parallel trends to hold.
   3. Under covariate imbalance, the generalization to a matching estimator relies on the conditional parallel trends assumption and sufficient overlap. $E[\Delta Yi|D_i = 1, X_i = x] - E[\Delta Yi|D_i = 0, X_i = x]$ can then identify λ by averaging. For each treated i, choose an untreated unit $m(i)$, estimate $\hat{E}[\Delta Yi|D_i = 0, X_i]$, assign G_i to $m(i)$ and estimate matched event study as $Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \in L} 1[t = g + \ell]\delta_\ell + \sum_{\ell \in L} D_i P_\ell^t \tau_\ell + \epsilon_{it}$, where δ_ℓ is the avg. trend in treated / matched controls. Or choose balancing weights based on the propensity score (estimated) such that the distribution of X_i is the same in across the D_i groups after re-weighting. Combining both: double robustness.
   4. Pool of untreated observations increases: 1) closer matching, 2) weaker version of conditional parallel trends assumption: the matched never-treated counterfactual would evolve in parallel for all adoption groups (might not be parallel with the never-treated group). In addition to no anticipation and parallel trends, the dynamic DID specification now restricts homogeneity on treatment effects: The dynamic effect only depends on the relative time, but not on the treatment timing. The ATT could then by summarized by taking averages of the estimated τ_ℓ for $\ell \geq 0$. The introduction of an additional cohort with sufficient overlap also enables constructing a counterfactual cohort whose outcomes are within the support of the other cohorts: the synthetic control method for staggered rollout. With multiple units, N_0 never-treated and $N_1 = N - N_0$ treated, to avoid contamination from comparison with the already treated cohort, the heterogeneity robust estimation strategy targets the unit-specific dynamic treatment effect for each $i : G_i < \infty$ which could then be used to compute the dynamic ATT: $\tau_\ell = \frac{1}{N_1} \sum_{i:G_i<\infty} \tau_{i\ell}$ (no anticipation $\tau_{i\ell} = 0 \quad \ell < -1$).

2. 1. Identification of the ATT: $\tau \equiv \mathbb{E}[Y_{it}(1) - Y_{it}(0)|D_{it} = 1]$ in $D = [(0,0,1)^T (0,1,1)^T]$ requires $\mathbb{E}[Y_{i2}(0) - Y_{i1}(0)|G_i = g] = \mathbb{E}[Y_{i2}(0) - Y_{i1}(0)|G_i = \infty] \quad g \in \{1,2\}, Y_{i1} = Y_{i1}(0) \, \forall i : G_i = 2$; $\mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|G_i = 1] = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0)|G_i = 1] = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0)|G_i = 2]$ equivalent to additive separability of the conditional expectation $Y_{it} = \alpha_i + \gamma_t + D_{it}\tau_{TWFE}$
   2. $\tau_S = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0)|G_i = 2]$ identified under parallel trends for $G_i \in \{2, \infty\}$ and no anticipation for $i : G_i = 2$ by the conditional expectation $Y_{it} = \mu + \alpha D_i + \gamma P_t + \tau_S D_i P_t$, estimated via OLS or by 2x2 DID comparison of means (treated post - treated pre) - (control post - control pre) $\tau = \tau_S$ if $\mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|G_i = 1] = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0)|G_i = 1] = \mathbb{E}[Y_{i2}(1) - Y_{i2}(0)|G_i = 2]$, not in general
   3. Smoothness at c_t : $\lim_{r \uparrow c_t} E[Y_{it}(d)|R_{it} = r] = \lim_{r \downarrow c_t} E[Y_{it}(d)|R_{it} = r] \quad (d,t) \in (\{0,1\}, \{1,2\})$ $\lim_{r \downarrow c_t} E[Y_{it}(1)|R_{it} = r] - \lim_{r \uparrow c_t} E[Y_{it}(0)|R_{it} = r] = E[Y_{it}(1) - Y_{it}(0)|R_{it} = c_t] \equiv \tau_t(c) \neq \tau$ if $\tau_1(c) \neq \tau_2(c)$ or $\tau_1(c) \neq \tau_2(\tilde{c})$, $\tilde{c} \neq c$; also $\tau_S \neq \tau_2(c)$: former ATT $G_i = 2$ latter LATE at c_2. Obtain $\hat{\beta}_t$ from $Y_{it} = \alpha_t + \beta_t \mathbf{1}[R_{it} > c_t] + \sum_{k=1}^K \gamma_{0kt}(R_{it} - c_t)^k + \sum_{k=1}^K \gamma_{1kt} \mathbf{1}[R_{it} > c_t](R_{it} - c_t)^k + \epsilon_{it}$ or use local nonparametrics to estimate $\tau_t(c)$, $t \in \{1,2\}$, where $G_i = 1$ is ommitted in $t = 2$
   4. If for $G_i \in \{2, \infty\}, \quad d \in \{0,1\} : \lim_{r \downarrow c_2} \mathbb{E}[Y_{i2}(d)|R_{i2} = r] = \lim_{r \uparrow c_2} \mathbb{E}[Y_{i2}(d)|R_{i2} = r]$, and $f_1(d)$ is continuous at c_2 under a more general $Y_{it} = f_{it}(R_{it} - c_2) + \gamma(R_{it} - c_2) + \tau_2(c_2)\mathbf{1}[R_{it} > c_2]\mathbf{1}[t = 2] + \epsilon_{it}$, where $\gamma(R_{it} - c_2)$ represents the time-invariant discontinuity which could be due to sorting and/or the effects of other policies that change at c_2 in $t = 1$, it is possible to remove $\gamma(R_{it} - c_2)$ and recover the ATE at c_2 for $G_i \in \{2, \infty\} : \tau_2(c_2)$ by estimating a regression discontinuity on the first differenced outcome $\Delta Y_i = (f_{i1}(R_{i1} - c_2) - f_{i2}(R_{i2} - c_2)) + \tau_2(c_2)\mathbf{1}[R_{it} > c_2] + (\epsilon_{i2} - \epsilon_{i1})$ where e.g. $f_{it}(R_{it} - c_2) = \alpha_t + \beta_t \mathbf{1}[R_{it} > c_2] + \sum_{k=1}^K \gamma_{0kt}(R_{it} - c_t)^k + \sum_{k=1}^K \gamma_{1kt} \mathbf{1}[R_{it} > c_2](R_{it} - c_2)^k, \quad t \in \{1,2\}$

[1] mach5689@student.su.se

3. 1. 9 CEF equations (under CIA): $E[Y_{11}] = \alpha_1+\gamma_1, E[Y_{12}] = \alpha_1+\gamma_2+\beta, E[Y_{13}] = \alpha_1+\gamma_3+\beta, E[Y_{14}] = \alpha_1+\gamma_4+\beta, E[Y_{22}] = \alpha_2+\gamma_2, E[Y_{23}] = \alpha_2+\gamma_3+\beta, E[Y_{24}] = \alpha_2+\gamma_4+\beta, E[Y_{33}] = \alpha_3+\gamma_3, E[Y_{34}] = \alpha_3+\gamma_4+\beta$ in 8 unknowns parameters: one needs to be normalized to solve for/identify the rest

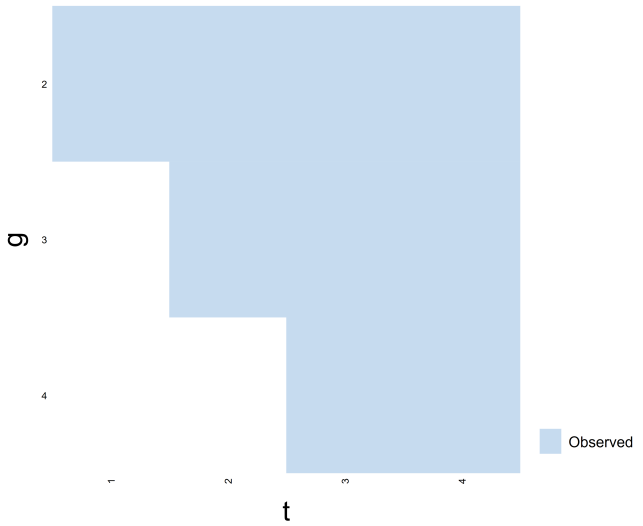
   2. Let $D_{it}^l = 1[t = G_i + l], L = \{-1, 0, 1, 2\}$ and set $D_{it}^{-1} = 0$ as a reference/to avoid multicollinearity $Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \in L} D_{it}^l \tau_\ell + \epsilon_{it}$, Define $\tau_{g+\ell}(g) = E\left[Y_{i,g+\ell}(g) - Y_{i,g+\ell}(\infty) \mid G_i = g\right]$ where ℓ is the number of periods since treatment. Identification of the dynamic ATT τ_ℓ requires parallel trends $E[Y_{it}(\infty)-Y_{i,t-1}(\infty) \mid G_i = 2] = E[Y_{it}(\infty)-Y_{i,t-1}(\infty) \mid G_i = 3] = E[Y_{it}(\infty)-Y_{i,t-1}(\infty) \mid G_i = 4]$, no anticipation $Y_{it}(g) = Y_{it}(g')$ for $g, g' > t$ and homogeneity across groups $\tau_\ell(g) = \tau_\ell$ (+ SUTVA). However, without never-treated units, path $\{\tau_\ell\}_{l \neq -1}$ not point identified in fully-dynamic OLS: adding a linear trend to this path, fits the data equally well: under-identification of the fully-dynamic specification requires an additional normalization $D_{it}^{l \neq -1} = 0$ to avoid multicollinearity.

   3. While untestable, not rejecting absence pre-trends can provide some evidence in favor of PTA, with at least one $G_i \in [3, \infty)$ run $Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \in L} \mathbf{1}[t = G_i + \ell]\beta_l + \tau D_{it} + \epsilon_{it}$, where $L = \{-g+1, \cdots -2\}$ is a set of lags, and test test $\beta_l = 0$ jointly or individually. The unbalanced treatment structure in this question implies the relative event time $\ell > -2$: pre-trends untestable.

   4. FWL: $\tilde{D}_{gt} = D_{it} - \bar{D}_g - \bar{D}_t + \bar{D} \implies \tilde{D}_{22} = 1-3/4-1/2+2/3 = 5/12, \tilde{D}_{23} = 1-3/4-2/3+2/3 = 3/12, \tilde{D}_{24} = 1-3/4-1+2/3 = -1/12, \tilde{D}_{33} = 1-2/3-2/3+2/3 = 4/12, \tilde{D}_{34} = 1-2/3-1+2/3 = 0, \tilde{D}_{44} = 1 - 1/2 - 1 + 2/3 = 1/6$ (note bottom-right of the treatment structure $(g, t) = (2, 4)$)
      Theorem dC & DH (2020): $w_{gt} = \frac{\tilde{D}_{gt}}{\sum_{(g,t):D_{it}=1} \tilde{D}_{gt}} \implies w_{24} = -1/12/[(5+3-1+4+2)/12] = -1/13$
      $E\left[\sum_{(g,t):D_{it}=1} \frac{N_{gt}}{N_1} w_{gt} \tau_{gt}\right] \implies \tau_{TWFE} = [1 - (-1/13)] \cdot 1 + (-1/13) E[\tau_{24}] \geq 0 \iff E[\tau_{24}] \leq 14$

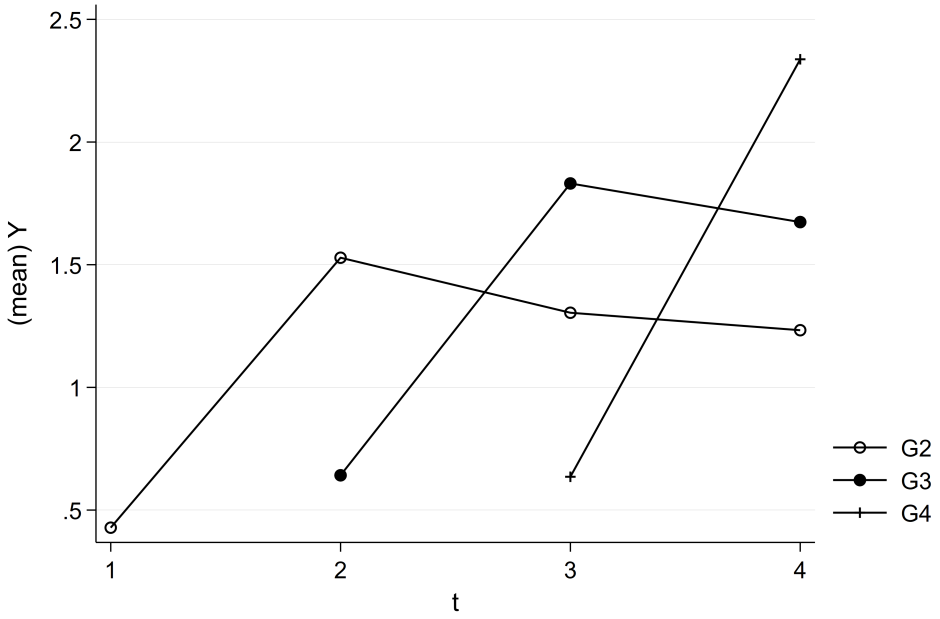
   5. ```
      set seed 42 // set obs 300 // gen i = _n // gen g = cond(i <= 100, 2, cond(i <= 200, 3, 4)) // expand 4 // bysort i: gen t = _n // bysort i: gen alpha = uniform() // replace alpha = alpha[_n-1] if i == i[_n-1]
      gen tau = 0// gen D = 0 foreach t of numlist 1/4 { replace D = (t >= g) if t == 't' // replace tau = cond(t < g, 0, cond(t == g, g/2 + rnormal(0,sqrt(0.1)), tau[_n-1] - (t-g)/5 + rnormal(0, 1))) if t == 't'}
      gen Y = alpha + D * tau + rnormal(0, 1) // replace Y = . if t < g - 1
      ```

   6. ```
      collapse Y, by(g t) // panelview Y, i(g) t(t) type(missing)
      ```



The dCdH switching estimator would not work here: we can not use the parallel trends assumption to measure the "joiner-effect": [0, 1] against [0, 0] since we do not observe the latter (which would allow for imputing the counterfactual trend of the treated) and there are no "leaver" [1, 0] cohorts.

7. `collapse (mean) Y, by(g t)`




```
sum tau // scalar att = r(mean) *.58302057
reghdfe Y D, absorb(i t) *1.593232
```

TWFE regression estimate 1.593 is significantly above the true ATT 0.583.

8. `sum tau if g==t // scalar attg= r(mean) *1.4690858`
9. `gen D0 = cond(g == t, 1, 0) // gen D1= cond(g==t[_n-1], 1, 0)`
`reghdfe Y D0 D1, absorb(i t) *1.062155`

Dynamic DID (separate estimates for the period of treatment and the period right after treatment) estimate for the period of treatment 1.062 is signigicantly below the ATT of the first period 1.469.

	TWFE	DynamicDID
D	1.593***	
	(0.147)	
D0		1.062***
		(0.0979)
D1		0.488***
		(0.130)
N	900	900
True ATT	0.583	1.469

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Estimating causal effects in staggered rollouts necessitates avoiding 'forbidden comparisons,' where already treated groups are compared to those that are becoming treated. Many robust estimators rely either on a never-treated group (Callaway and Sant'Anna) or on the not-yet treated groups (dC & DH, Borusyak et al) to impute the counterfactual trend of the treated groups. In this setting, due to the absence of a never-treated group and missing data, it is difficult to leverage these methods.