Canonical 2-by-2 Difference-in-Differences (DD)
Econ 672

1) The Takeaway
    a. A very common evaluation and identification strategy that can handle selection on observed and unobserved factors
    b. Unlike the Fixed Effects (within) estimator, which handles only time-invariant heterogeneity, the DD estimator can handle time-invariant heterogeneity and time-varying heterogeneity assuming the main assumption of DD holds
2) Pros
    a. A simple yet powerful identification strategy
    b. Can control for time-invariant and time-varying heterogeneity
3) Cons
    a. The DD estimator estimates the ATT not the ATE
    b. A dichotomous issue with the DD estimator
        i. When we only have 1 treatment group, we may have biased standard errors which may be downward biased
        ii. When we have more than 1 group adopting treatment at varying times, we may have a biased estimate of the ATT
        iii. (Note: Assuming we have many treatment groups simultaneously, then these two issues are not a problem)
4) Assumptions
    a. Parallel Trends Assumption
        i. The trends in the comparison group are the true counterfactual trends of the treatment group
    b. Stable Unit Treatment Value Assumption
        i. No spillovers; this is more problematic with DDD than DD
5) Testable Assumptions
    a. Parallel Trends is not directly testable, but we can indirectly test for it
    b. Pre-treatment Balance
        i. Plotting trends in outcomes or providing the data
    c. Pre-treatment Leads
        i. Coefficients on pre-treatment leads shows that outcomes are the same before the actual treatment is implemented
        ii. This needs a Two-Way Fixed Effects Difference-in-Differences (TWFEDD) to be implemented
    d. Placebo Falsification Tests
        i. Placebo tests by assigning treatment to comparison units that never receive treatment
        ii. Placebo timing by assigning treatment to treatment units earlier than the treatment units actually receive treatment
6) Estimation
    a. The basis of the DD is to compare outcomes before and after the treatment for treatment and control and then compare the 1st differences

b. In regression: $Y_{its} = \alpha + \gamma Treat_s + \lambda Post_t + \delta(Treat * Post)_{st} + \epsilon_{its}$
c. $\hat{\delta}_{kU}^{2x2} = E[Y_k^1|Post] - E[Y_k^0|Pre] + \left[E[Y_k^0|Post] - E[Y_k^0|Pre]\right] - \left[E[Y_U^0|Post] - E[Y_U^0|Pre]\right] = ATT + ParallelTrends$
d. Where
   i. $ATT = E[Y_k^1|Post] - E[Y_k^0|Post]$
   ii. $Parallel\ Trends = \left[E[Y_k^0|Post] - E[Y_k^0|Pre]\right] - \left[E[Y_U^0|Post] - E[Y_U^0|Pre]\right]$
   iii. If Parallel Trends assumption holds the second term is 0 and the DD estimator identifies the ATT

7) Difference-in-Differences-in-Differences (Triple DDD or DDD)
   a. DDD is a placebo test for a DD estimate
   b. We want to see if a group within the treatment unit is unaffected by the treatment when they do not receive the treatment
   c. An example is maternity benefits for men or women of non-child bearing age. A treatment of maternity benefits should not directly affect men or women of non-child bearing age
   d. In regression:
      i. $Y_{ijt} = \alpha + \psi X_{ijt} + \beta_1 Post_t + \beta_2 Group_j + \beta_3 Treat_i + \beta_4(Group * Post)_{jt} + \beta_5(Post * Treat)_{it} + \beta_6(Group * Treat)_{ji} + \beta_7(Treat * Group * Post)_{ijt\_}\epsilon_{ijt}$
      ii. β_7 is our triple difference interaction and our parameter of interest.