Application for use of ML in IV Estimation

- Children and their Parents Labour Supply (Angrist and Evans, 1998)
- Is there a causal link between female labour supply and fertility (number of children)?
  - Both are endogenous and jointly determined i.e. may be the case that men/women who desire a bigger family size also have a weaker preference for labour
  - Paper uses a unique IV to estimate causal effect of fertility decision on labour force participation and labour supply decisions: gender mix of children as IV for fertility decision for families with two or more children
    - Parental preference for a sibling gender mix; parents of same-gender siblings more likely to have an additional child
    - Exogenous source of variation in family size

- ML application: GENERALIZED RANDOM FORESTS (Susan Athey, Julie Tibshirani and Stefan Wager, 2018)

- US data
  - 1980 Census and 1990 census: Public Use Micro Samples (PUMS)
  - Sample restricted to households with mothers between 21-35, with two or more children and eldest child being less than 18
  - Gender of the eldest sibling defines same-sex sibling pairs
  - Analysis for sample of (i) all women, (ii) married women (iii) husbands of married women
    - Separately for 1980 and 1990 data

- SAS dataset available at https://economics.mit.edu/faculty/angrist/data1/data

Application for use of ML in IV Estimation

- Analysis corresponds to Table 6, Table 7 (Columns 1,2, 4, 5, 7 and 8)and Table 8 (Columns 1,2, 4, 5, 7 and 8) of Angrist and Evans (1998)

- Variables
  - Labour supply measures (Y_i) - refers to work in year prior to Census
    - For labour force participation: "If worked for pay" as LHS variable (i) all women, (ii) married women (iii) husbands of married women
    - For all women: If worked for pay, then use (i)"weeks worked", (ii)"hours worked", (iii)"annual labour income" (iv) log of family income as labour supply LHS variables
    - For married women: If worked for pay, then use (i)"weeks worked", (ii)"hours worked", (iii)"annual labour income" (iv) "log of family income" (v) "log of non-wife income" as labour supply LHS variables
    - For husbands of married women : If worked for pay, then use (i) "weeks worked", (ii)"hours worked", (iii)"annual labour income" as labour supply LHS variables
  - Fertility measure
    - RHS endogenous variable: Dummy variable = 1 if mother has more than two children
    - IV is dummy variable =1 if same-sex siblings
    - Other controls mentioned in Notes to Table 7 and 8