The synthetic control method compared to difference in differences: discussion

Monica Costa Dias

April 2014

Outline of discussion

1. Very brief overview of synthetic control methods
2. Synthetic control methods for disaggregated data
3. Matching vs/and synthetic control methods
4. Indirect effects

Synthetic control methods: overview

- Goal: to evaluate the impact of a treatment implemented at the aggregate level in one (or very few) unit using a small number of controls to build the counterfactual

- Synthetic control methods
  - use (long) longitudinal data to build the weighted average of non-treated units that best reproduces characteristics of the treated unit over time, prior to treatment
  - this is the synthetic cohort
  - impact of treatment is quantified by a simple difference after treatment: treated vs synthetic cohort

- Goal: to evaluate the impact of a treatment implemented at the aggregate level in one (or very few) unit using a small number of controls to build the counterfactual

- Synthetic control methods
  - use (long) longitudinal data to build the weighted average of non-treated units that best reproduces characteristics of the treated unit over time, prior to treatment
  - this is the synthetic cohort
  - impact of treatment is quantified by a simple difference after treatment: treated vs synthetic cohort

Synthetic control methods: formalisation

- Units: $j = 0, 1, \ldots, J$ where $j = 0$ is the treated and $j = 1, \ldots, J$ are controls
- Time frame: $t = 1, \ldots, T_1$ split in two periods - before treatment $t = 1, \ldots, T_0$ and after treatment $t = T_0 + 1, \ldots, T_1$
- Potential and observed outcomes for the treated unit are $\left(Y_{0t}^0, Y_{0t}^1\right)$ and

$$Y_{0t} = \begin{cases} Y_{0t}^0 & \text{for } t = 1, \ldots, T_0 \\ Y_{0t}^1 & \text{for } t = T_0 + 1, \ldots, T_1 \end{cases}$$

- Aim is to estimate $\alpha_{0t} = Y_{0t}^1 - Y_{0t}^0$ for $t = T_0 + 1, \ldots, T_1$

Synthetic control methods: formalisation

Model of untreated outcomes for unit $j = 0, \ldots, J$ and time $t = 1, \ldots, T_1$

$$Y_{jt}^0 = \delta_t + \theta_t Z_j + \lambda_t \mu_j + \varepsilon_{jt}$$

- Z_j are the observed, pre-treatment covariates
- μ_j are permanent unobserved variables
- δ_t are common time effects
- ε_{jt} are unobserved transitory shocks at the unit level with zero mean

Synthetic control method: formalisation

Choose $W^* = \left(w_1^*, \ldots, w_J^*\right) \in [0,1]^J$, adding to 1, to minimise distance in pre-treatment characteristics between treated and weighted average of controls

- Treatment effect estimated by the simple difference

$$\hat{\alpha}_{0t} = Y_{0t} - \sum_{j=1}^{J} w_j^* Y_{jt} \quad \text{for } t = T_0 + 1, \ldots, T_1$$

- Ideally, one would want to select W^* such that

$$\sum_{j=1}^{J} w_j^* Z_j = Z_0 \quad \text{and} \quad \sum_{j=1}^{J} w_j^* \mu_j = \mu_0$$

  so $\hat{\alpha}_{0t}$ is unbiased (ε is mean-independent of (Z, μ) and independent across units and over time)

- Not feasible since μ is unobserved

Synthetic control methods: formalisation

- Solution: choose W^* satisfying

$$\sum_{j=1}^{J} w_j^* Z_j = Z_0, \quad \sum_{j=1}^{J} w_j^* Y_{j1} = Y_{01}, \quad \ldots \quad , \sum_{j=1}^{J} w_j^* Y_{jT_0} = Y_{0T_0}$$

- Bias can be bounded under mild conditions

$$|\mathrm{E}(\hat{\alpha}_{0t} - \alpha_{0t})| < \beta J^{\frac{1}{p}} \max \left\{ \left(\frac{m_p}{T_0^{p-1}} \right)^{\frac{1}{p}}, \frac{\bar{\sigma}}{\sqrt{T_0}} \right\}$$

Synthetic control methods: bias

$$\left|\mathrm{E}\left(\hat{\alpha}_{0t}-\alpha_{0t}\right)\right| < \beta J^{\frac{1}{p}} \max\left\{\left(\frac{m_p}{T_0^{p-1}}\right)^{\frac{1}{p}}, \frac{\bar{\sigma}}{\sqrt{T_0}}\right\}$$

- Bias is small when T_0 is large relative to scale of ε
- *Intuition*: a synthetic cohort can fit $(Z_0, Y_{01}, \ldots, Y_{0T_0})$ for a large T_0 only if it fits (Z_0, μ_0)
- But a large J does not help reducing the bias once these conditions are met

$$\sum_{j=1}^{J} w_j^* Z_j = Z_0, \quad \sum_{j=1}^{J} w_j^* Y_{j1} = Y_{01}, \quad \ldots \quad , \sum_{j=1}^{J} w_j^* Y_{jT_0} = Y_{0T_0}$$

Some issues

1. "Many controls" not necessarily beneficial - although this depends on how small J is and whether the aggregate treated unit lies outside the domain of the controls
2. Scale of the transitory shock
   - can be larger at the disaggregated level if the aggregate outcome is the average of outcomes for smaller units
   - in which case need more time periods to keep bias down
3. Possibly more serious interpolation bias

Some issues

1. "Many controls" not necessarily beneficial - although this depends on how small J is and whether the aggregate treated unit lies outside the domain of the controls
2. Scale of the transitory shock
    - can be larger at the disaggregated level if the aggregate outcome is the average of outcomes for smaller units
    - in which case need more time periods to keep bias down
3. Possibly more serious interpolation bias

Some issues

1. "Many controls" not necessarily beneficial - although this depends on how small J is and whether the aggregate treated unit lies outside the domain of the controls
2. Scale of the transitory shock
   - can be larger at the disaggregated level if the aggregate outcome is the average of outcomes for smaller units
   - in which case need more time periods to keep bias down
3. Possibly more serious interpolation bias

Synthetic control methods for disaggregated data

Some issues

1. "Many controls" not necessarily beneficial - although this depends on how small J is and whether the aggregate treated unit lies outside the domain of the controls
2. Scale of the transitory shock
   - can be larger at the disaggregated level if the aggregate outcome is the average of outcomes for smaller units
   - in which case need more time periods to keep bias down
3. Possibly more serious interpolation bias

The synthetic control method relies on the linearity of the model of untreated outcomes

$$Y_{jt}^{0} = \delta_t + \theta_t Z_j + \lambda_t \mu_j + \varepsilon_{jt}$$

- Even if linearity is violated, the model can be a good local approximation
- But bias can be large if the characteristics of control units are far from those of the treated
- In aggregate studies: pick and choose the control units that more closely resemble the treated
- But this is difficult to implement with more and smaller units
- Or when treated and control units are of different nature

Matching and synthetic control methods

- DID has been used with matching to relax strict functional form assumptions
- Matching selects controls that have, each of them, characteristics $(Z_j, Y_{j1}, ..., Y_{jT_0})$ close to those of the treated
- Matching is a local estimator
  - hence it is less sensitive to interpolation bias
  - and allows for a more general specification of the model of untreated outcomes

Matching and synthetic control methods

A more general model, under the assumption that ε is balanced conditional on $\left(Z_j, Y_{j1}, ..., Y_{jT_0}\right)$:

$$Y_{jt}^0 = f_t\left(Z_j, \mu_j\right) + \varepsilon_{jt}$$

- Matching is unbiased if it removes unobserved permanent differences between treated and control units
- But it may be impossible to match closely on the many characteristics $\left(Z_j, Y_{j1}, ..., Y_{jT_0}\right)$
- Combine matching with synthetic cohorts for more disaggregated data?
  - if f is smooth, a local polynomial approximation of f is accurate
  - matching prior to applying a synthetic control method could help ensuring the comparability of admissable controls

Indirect effects

- At the aggregate level, it is quite plausible that what happens in one unit affects other units
- Abadie's and co-authors applications
  - California's tobacco control programme may have influenced behaviour and legislation in other states
  - The unification of Germany (and indeed aggregate shocks to its economy) may affect the economic outcomes of closely connected countries
- More similar control units may be more exposed to indirect effects
- Trade-off between interpolation bias and indirect effects