



# Counterfactual-Analyse

*Masterseminar WT 2020 an der Professur für  
Wandel und Nachhaltigkeit*



# Vorläufiger Zeitplan

Sitzung Nr.	Datum	Uhrzeit	Thema	Raum
1	14.01.2020	09:00-10:30	Theorie	36 / 1114
2	21.01.2020	09:00-10:30	Praxis	104 / 0117 PC-Pool 4
3	<b>28.01.2020</b>	<b>08:30-11:00</b>	<b>Praxis</b>	<b>104 / 0117</b> <b>PC-Pool 4</b>
4	03.03.2020	09:00-10:30	Zwischenstand Probleme, Einzeltermine?	36 / 1117
5	31.03.2020	-	Abgabe Seminararbeit	36 / 1117



## Synthetic Control Method - Wiederholung

- 1. Schritt: Konstruktion eines Gewichtungsvektors  $W^*$ , der die „Distanz“ der Variablen (zur Vorhersage des Outputs  $Y$ ) zwischen dem betroffenen Land ( $X_1$ ) und der selben Variablen der Kontrollgruppe ( $X_C$ ) minimiert (Pre-Treatment!)

$$\min(X_1 - X_C W) \mathcal{V}(X_1 - X_C W)$$

- 2. Schritt: Der Treatment-Effekt ist die Differenz der Outputs zwischen der betroffenen Einheit ( $Y_{1t}$ ) und dem gewichteten Durchschnitt der nicht betroffenen Einheiten

$$TE_{it} = Y_{1t} - \sum_{i=2}^{N+1} w_i^* Y_{it}, \forall t \geq T_0$$



## Synthetic Control Method - Stata-Intro

Synth: A Statistical Software Package for Synthetic  
Control Methods in Comparative Case Studies

Alberto Abadie, Harvard University

Alexis Diamond, IFC

Jens Hainmueller, Stanford University

▶ 0:02 / 10:43



<https://web.stanford.edu/~jhain/Video/SynthDemo.mp4>



# Synthetic Control Method

- Installation des Schätzers:

`ssc install synth, replace all`

- Hilfestellung für Schätzer

`help → Stata command → synth`

- Empfehlung - Arbeiten mit *Dofiles*



## Synthetic Control Method - Replikation Abadie, Diamond und Hainmüller (2010)

- `sysuse synth_smoking` (Laden des Datensatz)
- `tsset state year` (Definiton der Panelvariable und der Zeitvariable; „state“ muss numerisch definiert sein!!!)
- `synth cigsale beer(1984(1)1988) lnincome retprice age15to24 cigsale(1988) cigsale(1980) cigsale(1975),  
trunit(3) trperiod(1989)`
- ATE?



## Synthetic Control Method - Replikation Abadie, Diamond und Hainmüller (2010)

- synth cigsale beer(1984(1)1988) lnincome retprice age15to24 cigsale(1988) cigsale(1980) cigsale(1975), trunit(3) trperiod(1989) fig
- Zum Verständnis: Wie könnte man Graphik manuell erstellen bzw. wie können Datenpunkte der Kontrollgruppe per Hand berechnet werden?



## Synthetic Control Method - Replikation Abadie, Diamond und Hainmüller (2010)

### Weitere Befehle

- `synth cigsale retprice cigsale(1970) cigsale(1979) , trunit(33) counit(1(1)20) trperiod(1980) fig resultsperiod(1970(1)1990)`  
→ Unit: Texas, Treated Period: 1980, Graphik 1970-1990
- `synth cigsale beer lnincome retprice age15to24 cigsale(1988) cigsale(1980) cigsale(1975) , trunit(3) trperiod(1989) xperiod(1980(1)1988) nested`  
→ Durchschnitt der vorhersagenden Variablen von 1980-1988, Optimierung der Gewichtungsberechnung (nested, allopt), wirkt sich oftmals „positiv“ auf *root mean squared prediction error (MSPE)* aus

MSPE: Durchschnitt der quadrierten Distanzen zwischen Zigarettenverkäufe in Kalifornien und synth. Kalifornien während der Pretreatment-Periode (1970-1988)



Synthetic Control Method - Replikation Abadie, Diamond und Hainmüller (2010)

### Weitere praktische Befehle

- `cd "C:\Users\WOW\Desktop"` oder manuell über file → change working directory

→ Bestimmen des automatischen Speicherorts von Statadateien

- `synth cigsale beer(1984(1)1988) lincome retprice age15to24 cigsale(1988) cigsale(1980) cigsale(1975), trunit(3) trperiod(1989) keep (Tobacco)`

→ Speichern der neu geschätzten Daten (z.B. Kontrollgruppe)

## Synthetic Control Method - Replikation Abadie, Diamond und Hainmüller (2010)

- Robustheit der Ergebnisse oder „Zufallsergebnisse“ ?
- „Signifikanztest“ bzw. Robustheit der Ergebnisse (1)
  - Placeboläufe (betroffenes Land oder Zeitpunkt des Ereignisses ändern), weitere Kontrollvariablen einfügen, MSPE im Vergleich

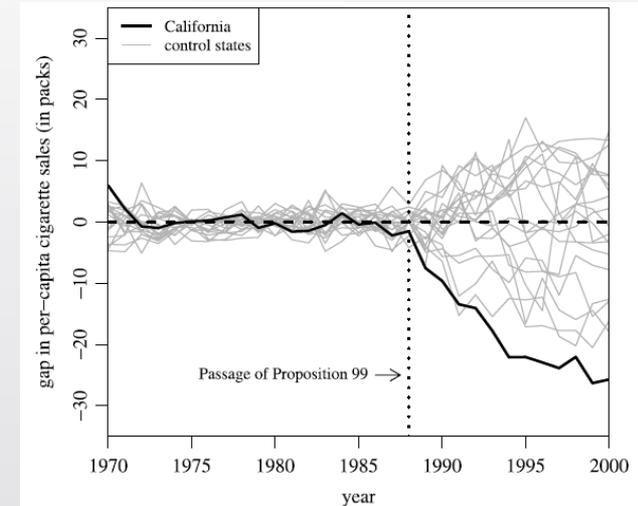


Figure 7. Per-capita cigarette sales gaps in California and placebo gaps in 19 control states (discards states with pre-Proposition 99 MSPE two times higher than California's).

Quelle: Abadie et al. (2010), S. 503



## Synthetic Control Method - Replikation Abadie, Diamond und Hainmüller (2010)

- „Signifikanztest“ bzw. Robustheit der Ergebnisse (2)
  - Teststatistik, ob Durchschnitt der Treatmentbeobachtungen signifikant vom Durchschnitt der Kontrollgruppe abweicht
    - Parametrischer t-Test (paired, da Annahme betroffenes Land = synthetisches Land)
      - nur anzuwenden bei Normalverteilung (Prüfen: z.B. `histogram _Y_synthetic, normal`)  
  
`ttest _Y_treated == _Y_synthetic if _time < 1989`  
`ttest _Y_treated == _Y_synthetic if _time > 1988`
    - Nicht-parametrischer Mann-Whitney-U-Test (zuvor Zusammenlegung der Variablen Treatment und Kontrollgruppe mithilfe einer Statusvariablen)  
  
`ranksum _Y_treated if _time < 1989, by(status)`  
`ranksum _Y_treated if _time > 1988, by(status)`



## Synthetic Control Method - Aufgabe

1) Download GDP per capita 1985-1995; worldbank.org (Excel)

2) Import Stata, in „brauchbare“ Form bringen, speichern

**(reshape long YR, i(CountryName IndicatorName) j(year))**

3) Download CO<sub>2</sub>-Emissionen (kt) 1985-1995; worldbank.org (Excel)

4) Import Stata, in „brauchbare“ Form bringen, speichern

**(reshape long YR, i(CountryName IndicatorName) j(year))**

5) Zusammenführung der beiden Datensätze (.dta) und erstellen einer ID für jedes Land

(Datenbrowser → data → combine datasets → merge two datasets)



## Synthetic Control Method - Aufgabe

6) Auswählen relevanter Länder für Überschaubarkeit

(z.B. `keep if country == "Algeria" | country == "Argentina" | country == "Australia" | country == "Austria"` (oder mithilfe ID!))

7) Synthetic Treatment für Deutschland; Event: 1990 (Wiedervereinigung); Auswirkung auf CO<sub>2</sub>-Emissionen; Kovariable: GDP per capita

8) Teststatistik?