Masterclass

Counterfactual methods in cohesion policy evaluation

Davide Ceccanti

REGIO B.1 - Policy Development & Evaluation
Directorate-General for Regional and Urban Policy

Evaluation Helpdesk 2025 Summer School

4 July 2025

European Commission

Intro: reading the room

- How many of you have some training on counterfactual methods?
- Did anybody here have to design or manage a counterfactual study?
- Who has direct experience in running a counterfactual themselves?

Intro: goals for today

Understanding counterfactuals -> 4 sections

1. What is a counterfactual?
2. What methods, how and when?
3. Expost evaluation ERDF-CF 2014-2020
4. Managing counterfactuals: pitfalls & ideas

NB -> Please interrupt for questions. Debate at the end.

I. What is a counterfactual?

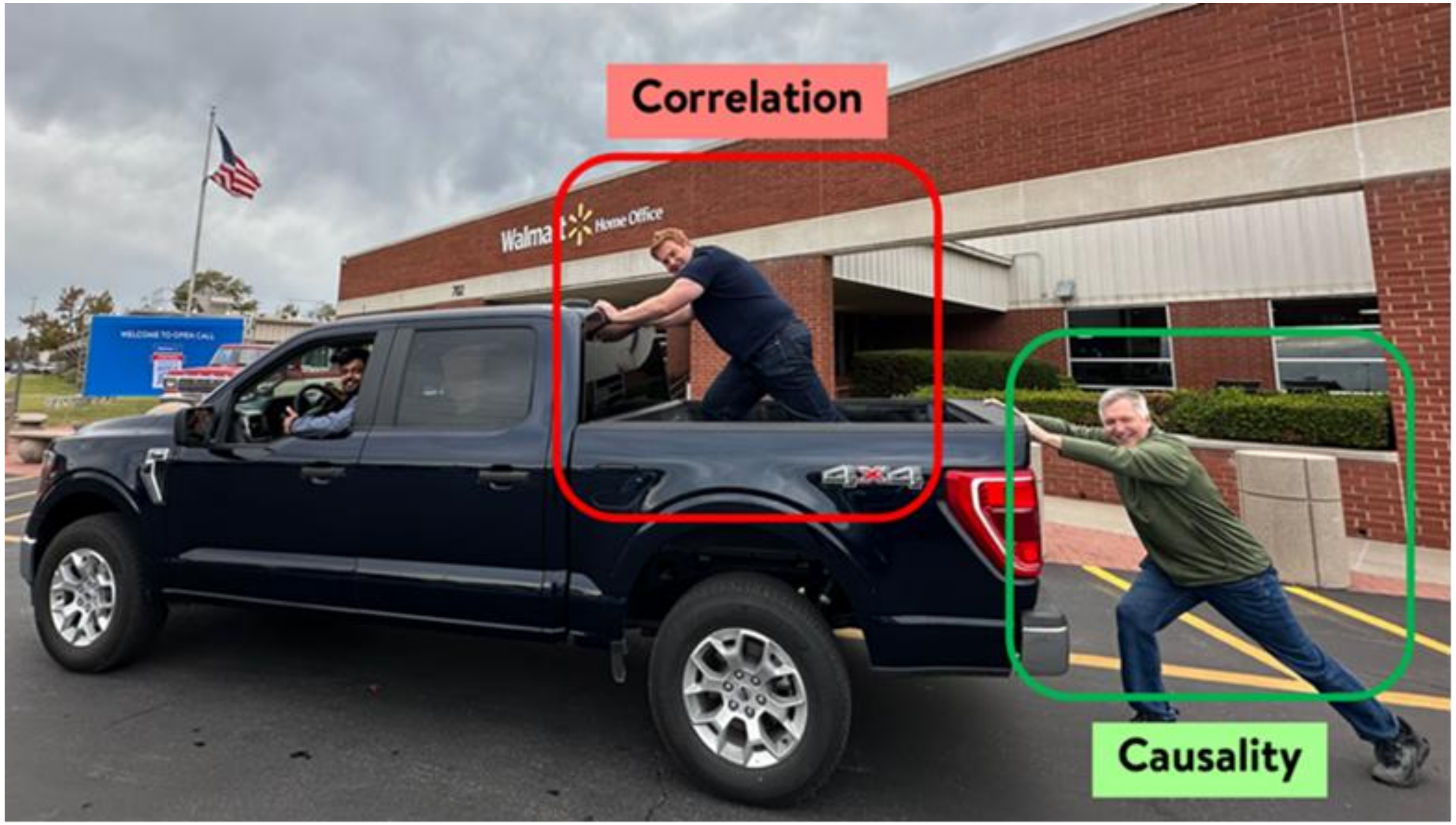
Causality

For policymaking, we are interested in the **IF-THEN** link in questions such as:

- If I build a new **hospital** in that neighborhood, will people be **healthier**?
- If I spend more money on **scholarships**, will students get **better jobs**?
- If I force firms to adhere to **green standards**, will they pollute less and not lose **competitiveness**?

-> Does a change in the **treatment** measure CAUSE a change in the **outcome** measure? How do we check this? Is two variables moving at the same time enough to claim causality?

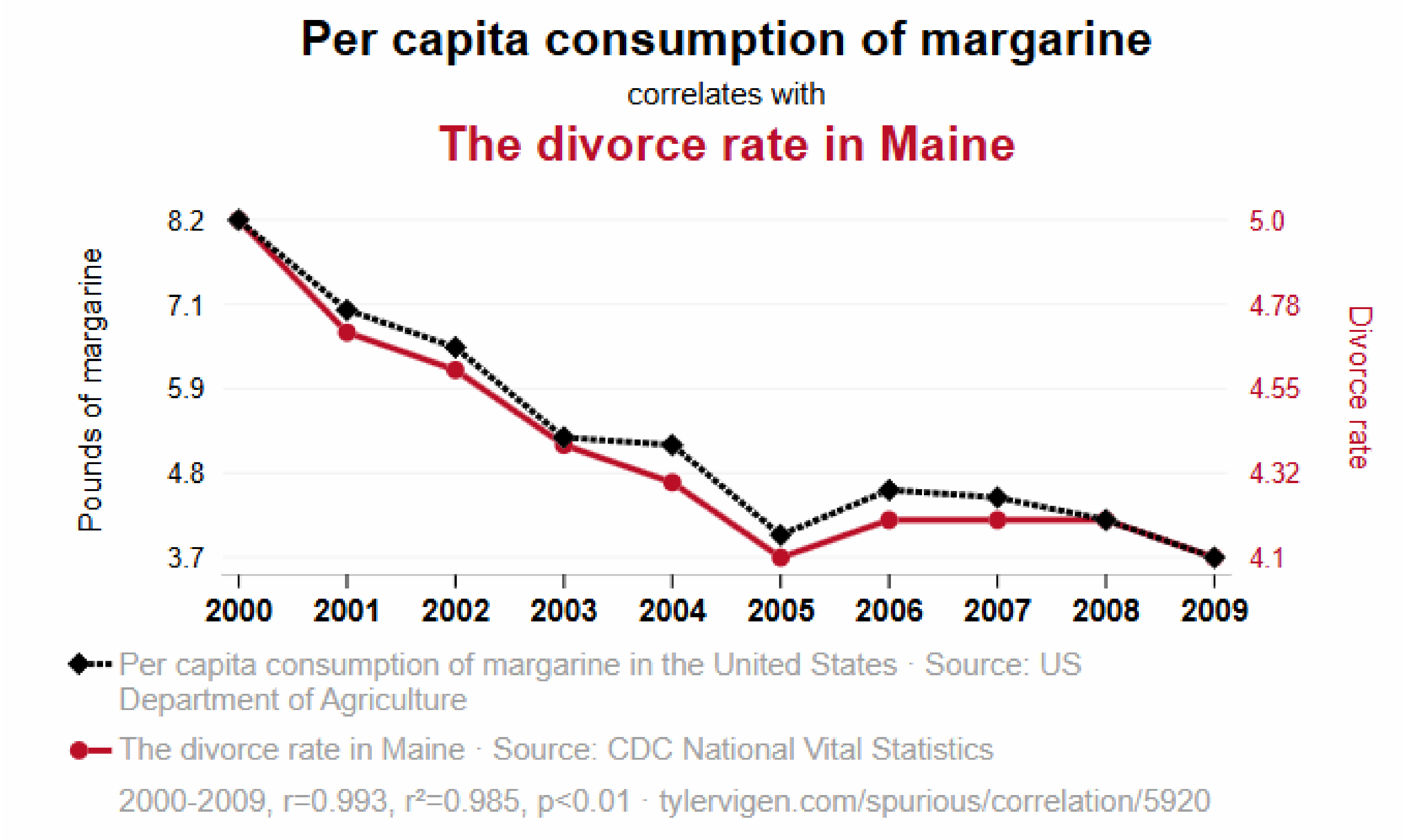
Causality vs correlation



Source: https://www.tylervigen.com/spurious-correlations

European Commission

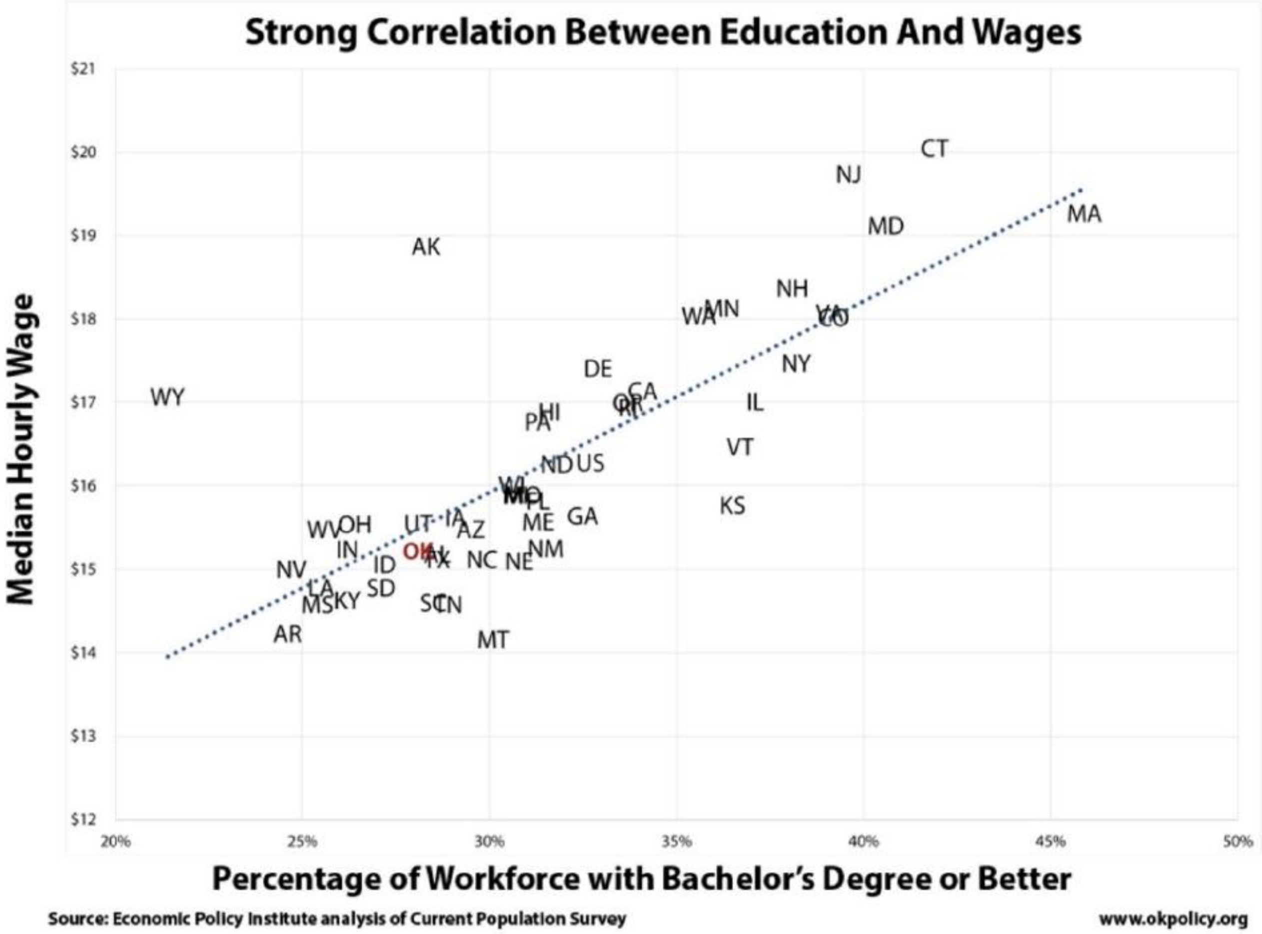
Causality vs correlation



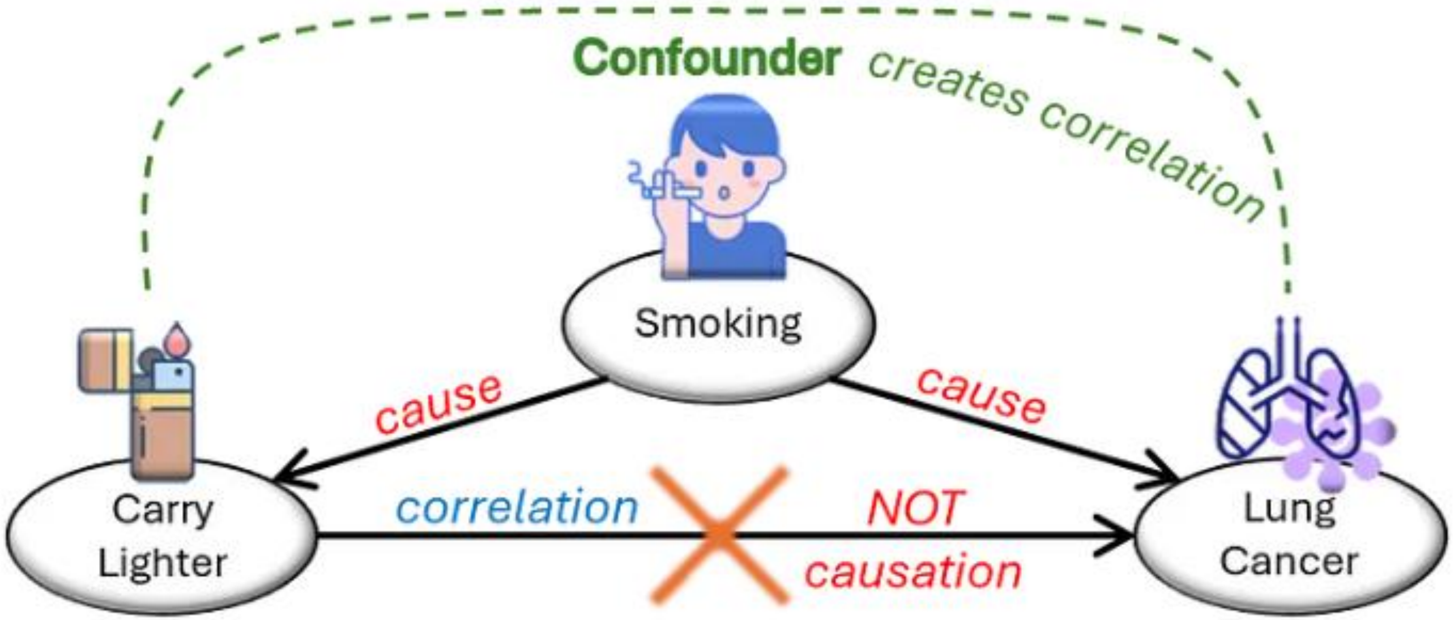
Source: https://www.tylervigen.com/spurious-correlations

European Commission

Causality vs correlation



Causality vs correlation



Source: https://medium.com/data-science-at-microsoft/causal-analysis-overview-causal-inference-versus-experimentation-versus-causal-discovery-d7c4ca99e3e4

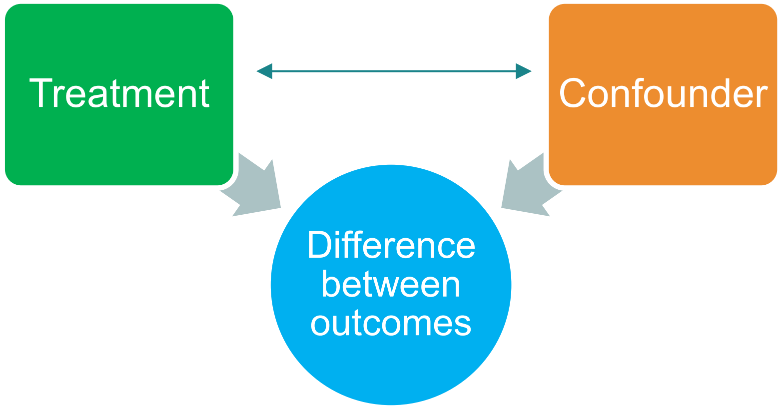
European Commission

How to establish causality?

- Observation -> covariance & correlation
- Credible theory -> Theory of Change
- Testing -> corroboration via qualitative and quantitative methods
- Quantitative -> experiments (ideally) & quasi-experiments (often the case)

If a **treatment** is administered to a subpopulation, you can compare the outcomes of ‘**takers’ vs ‘non-takers’**. The **difference** is attributed to the treatment… **BUT**

How to establish causality?



1. **Selection bias** -> if the selection is not random (often the case with evaluated policies), the difference may be due to some other factors.

2. **Confounders** -> Students going to BEST UNIVERSITY (treatment) get higher pay (outcome). BEST UNIVERSITY admits students with (confounder) most educated families / more money / smarter / more motivated. The higher pay may be caused by this characteristics rather than BEST UNIVERSITY.

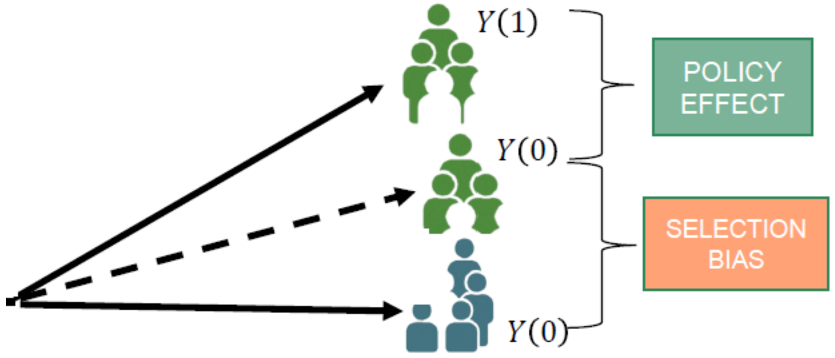
3. **Endogeneity** -> treatment is not independent of other factors affecting the outcome (indep. VAR correlated with error) because of omitted variables / simultaneity / measurement error / sample selection.

European Commission

How to establish causality?

- How do we deal with the selection bias? -> econometrics to the rescue: **COUNTERFACTUAL**

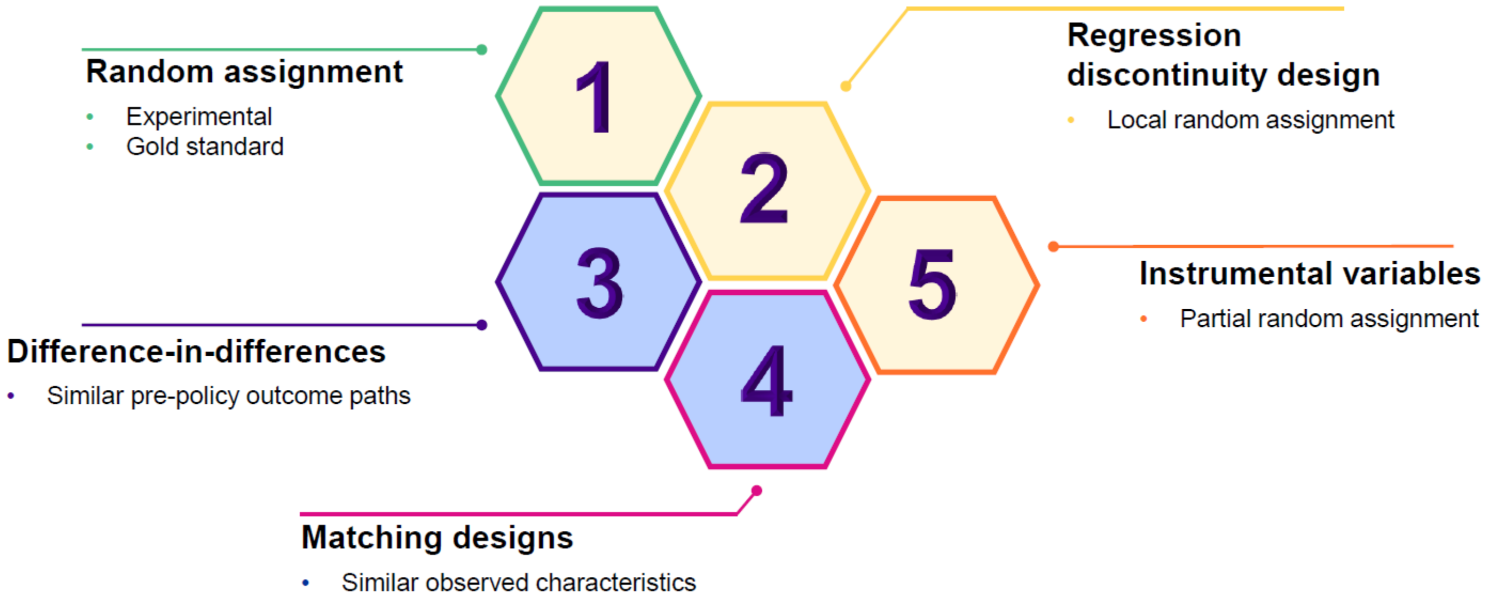
- The outcomes of ‘takers’ is compared to the hypothetical outcome of the same group in the absence of the treatment

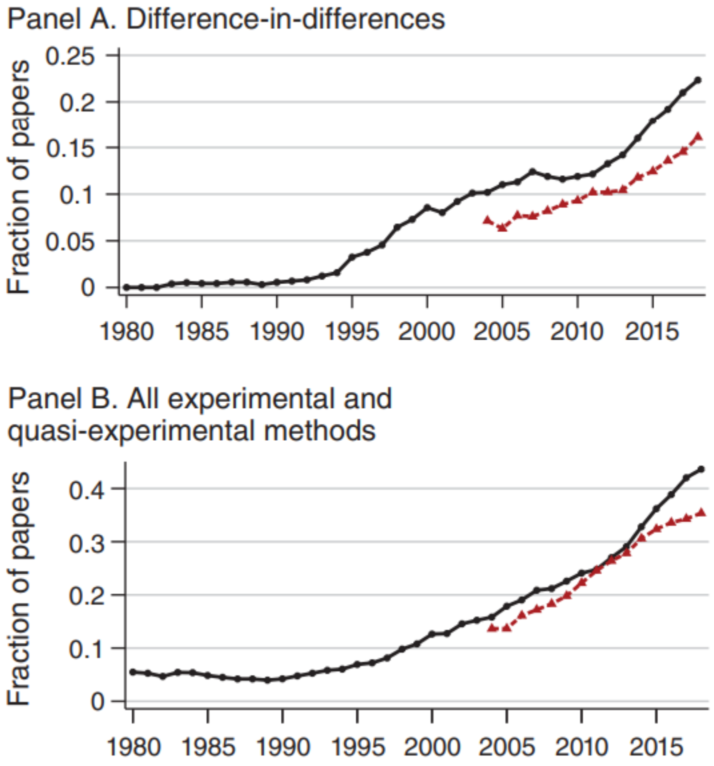


- The methodological challenge is then how to identify a credible counterfactual

II. What methods, how and when?

Different counterfactual methods



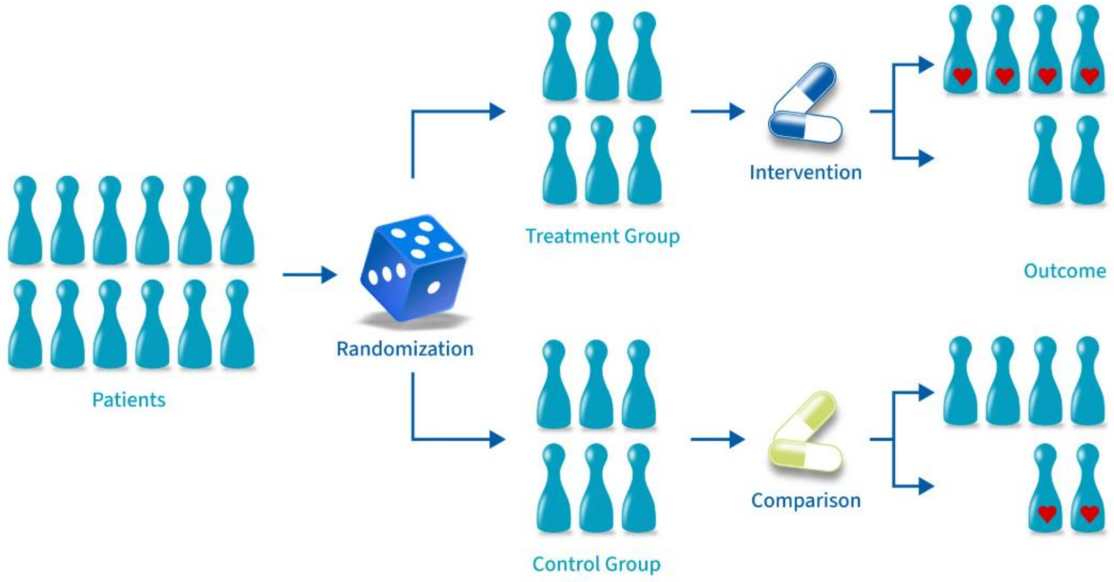


Source: JRC.

Currie, Kleven and Zwiers (2020)

European Commission

Different counterfactual methods



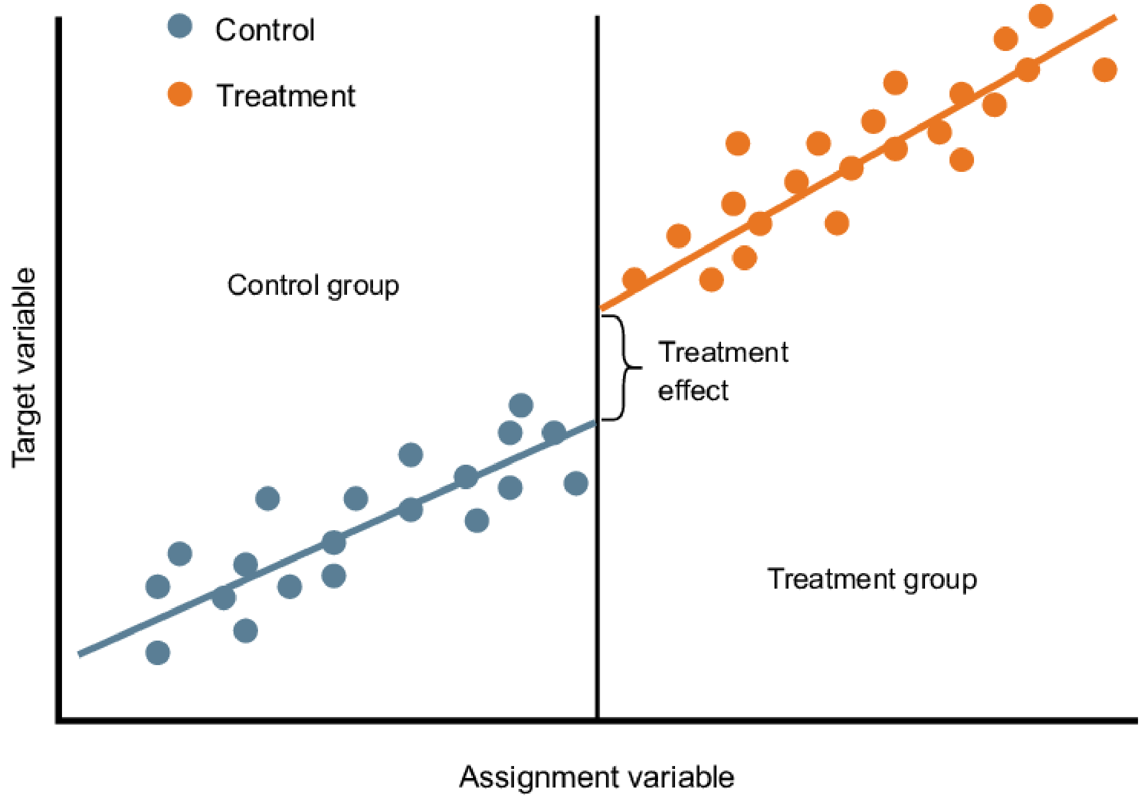
1. "RCT" Randomized Controlled Trials

- Random treatment assignment & placebo, pre-defined outcomes
- Gold standard causal inference (Nobel prize 2019 to Banerjee, Duflo, and Kremer)
- Often in medical research, limited but increasing application in policy evaluation: need for pre-intervention design, random assignment, ethical concerns, expensive etc.

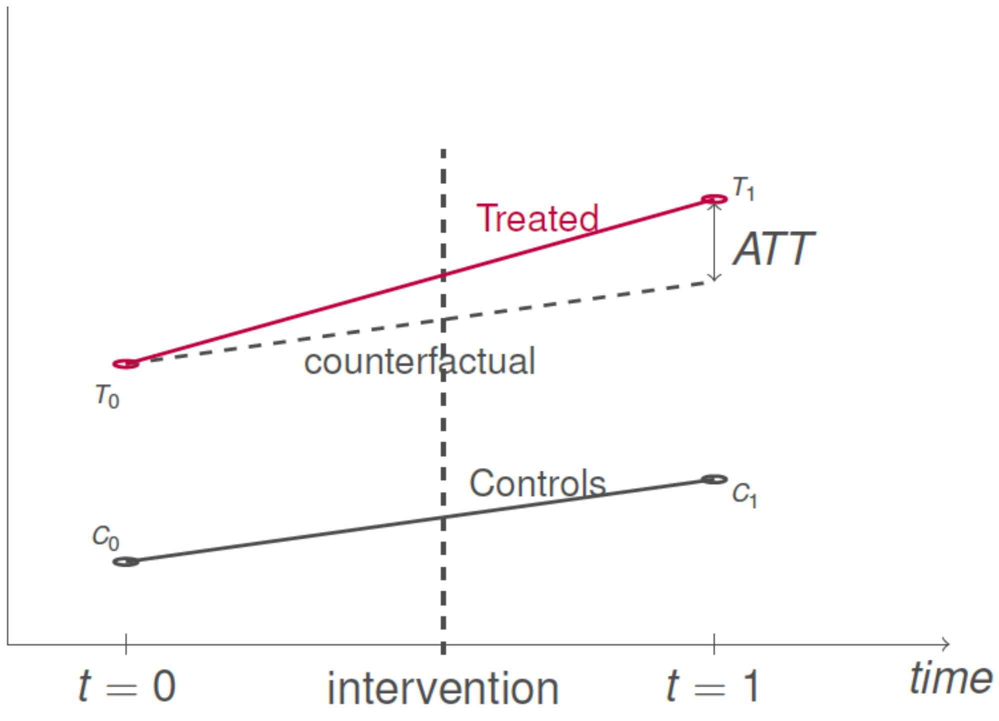
Different counterfactual methods

2. “RDD” Regression Discontinuity Design

- Similar units get treatment if their running variable (e.g., test score, income, age) exceeds a threshold
- Compare outcomes near cutoff (local treatment effect, quasi-random)
- Sharp (y/n based on cutoff) vs Fuzzy (probability changes sharply at cutoff)
- Requires a lot of data near cutoff; local validity; manipulation



Different counterfactual methods



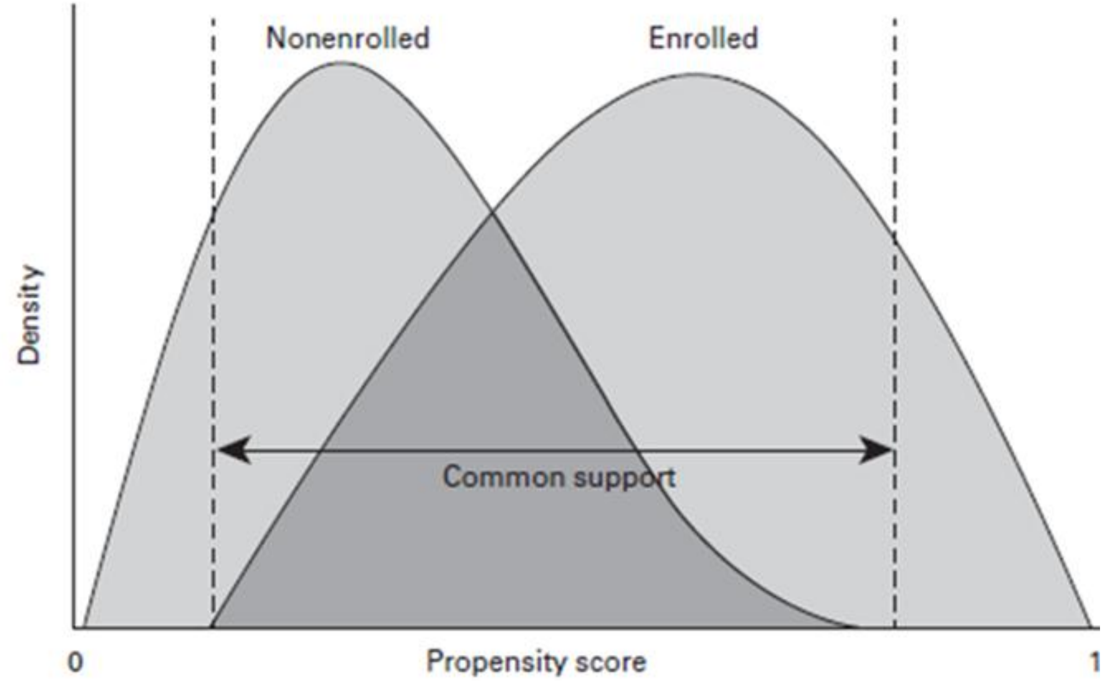
3. “DID” Difference-in-difference

- Difference between two groups before and after intervention: (YT,post−YT,pre)−(YC,post−YC,pre)
- Control for time-invariant differences between groups and common trends affecting both
- Convenient when no randomization but parallel trend assumption and lots of data

Different counterfactual methods

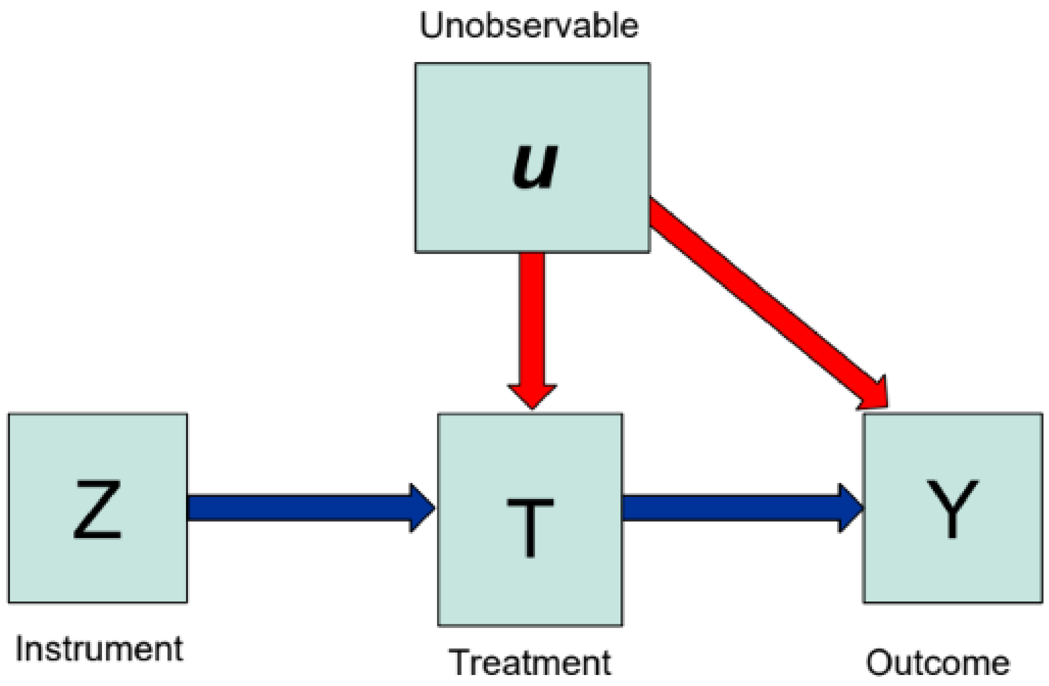
4. “PSM” Propensity Score Matching

- Techniques to match each treated unit with an untreated unit with similar covariates
- Propensity Score Matching -> estimate probability (0-1) via logistic regression (single measure), match similar P, difference mean
- Lots of individual data, lots of variables to consider



European Commission

Different counterfactual methods



5. “IV” Instrumental Variables

- When treatment is endogenous, randomization fails (no RCT and RDD) and not enough data or everybody treated (no DID and PSM)
- Use a third variable (instrument) to isolate (Wald estimator) exogenous variation in the treatment (get rid of evil endogenous) -> “pure” effect on outcome
- Must be relevant (strongly associated with treatment) and exogenous (not associated with confounder and outcome)

Member States CIEs on cohesion policy

	All Evaluations		Impact Evaluations	
Method	Total	Relative	Total	Relative
Qualitative	2,872	92%	1,100	90%
Quantitative	2,690	86%	1,097	89%
TBIE	444	14%	385	31%
CIE	**347**	**11%**	**345**	**28%**
MOD	88	3%	77	6%
CBA	87	3%	74	6%

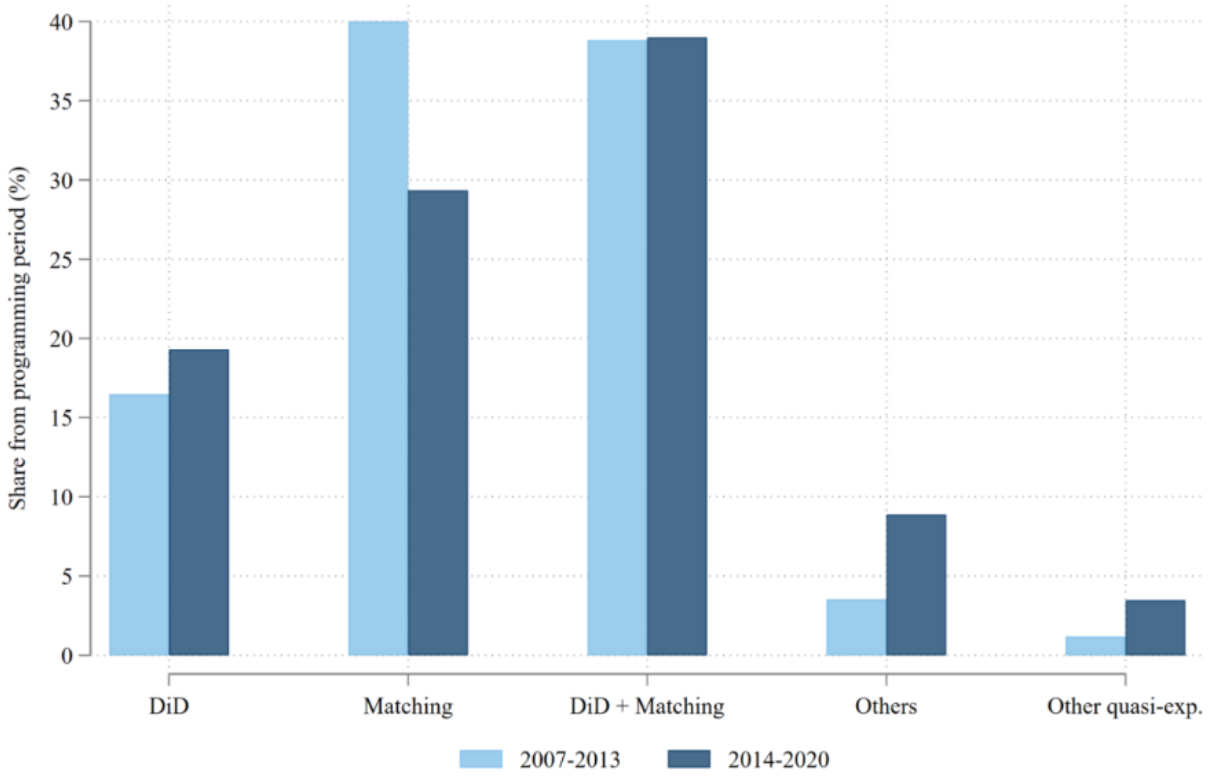
Notes: A single evaluation may fall under multiple methodological categories; thus, category totals can exceed the overall number of publications. Categories are designated according to the data dictionary.

Database: Evaluation Library (thanks to the Helpdesk)

- Info on 10% of evaluations, 30% of impact ones
- CIE share of evaluations increases over time
- Top 3 countries (PL, IT, CZ) contribute 50% of CIE
- Mainly DID, matching; no RCTs
- Themes: employment, social, education, SMEs, R&I

European Commission

Member States CIEs on cohesion policy



Database: [Evaluation Library](#) (thanks to the Helpdesk)

- Info on 10% of evaluations, 30% of impact ones
- CIE share of evaluations increases over time
- Top 3 countries (PL, IT, CZ) contribute 50% of CIE
- Mainly DID, matching; no RCTs
- Themes: employment, social, education, SMEs, R&I

Piloting experimentation

- **Goal** -> strengthen experimentation to obtain:
  - Better evaluation planning (embedded in the intervention)
  - Strong causal evidence (RCTs, where applicable)
  - Buy-in of policymakers (no isolated evaluators)
- **How** -> we work with World Bank on pilot projects (see Berlin event May 2025)
- **Status** -> currently at design stage; many strong projects; strong interest

European Commission

III. Ex post evaluation ERDF-CF 2014-2020

The case of cohesion policy

- Cohesion policy accounts for around 1/3 of the EU budget (i.e. the MFF)
- The main goal is to **reduce territorial & social disparities** but...
  also 15 thematic objectives 14-20
- One of the most evaluated policy in the world
- However, isolating its impact at the aggregate level is difficult:

➢ Short-run vs long run effects

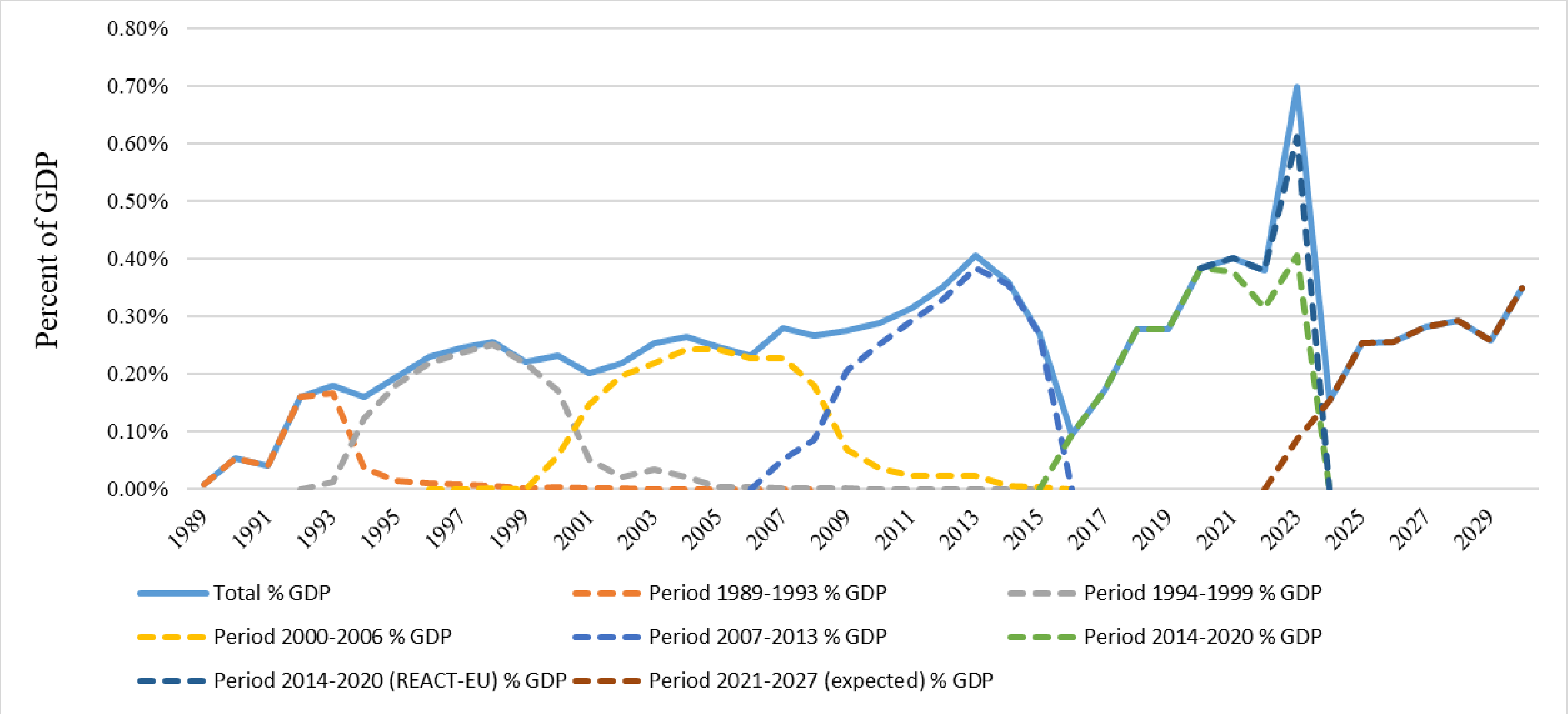
➢ Direct and indirect effects (data on programmes' outputs and results only provide information on direct effects);

➢ Spill-over and externalities

➢ Coincidence with other economic developments

The case of cohesion policy

Cohesion Policy funding 1989 to 2030 (% of GDP)



 Source: DG REGIO.

European Commission

The case of cohesion policy

Allocation of funds across 2014-2020 programmes

	Fund	EUR
Initial allocation	CF	61,455,291,486
	ERDF	199,733,761,176
	ESF	84,887,133,460
	YEI	8,950,645,385
	Sub-total	355,026,831,507
REACT-EU	ERDF	29,424,281,390
	ESF	19,361,713,312
	Sub-total	48,785,994,702
Total	Total	403,812,826,209

European Commission

Overview of the 2014-2020 evaluation

- Does cohesion policy “work”? What is the impact? What are the driving factors?
- One goal? -> improve policy and evidence-based policy-making, share lessons learnt, promote transparency and accountability, learn from beneficiaries…
- Complex level of analysis -> overall goal but thematic areas, EU aggreagte but also MS, MA, programmes projects, (potential) beneficiaries…
- Clear need for multiple methods
- Aim for robust evidence -> causality should be part of the picture

Overview of the 2014-2020 evaluation

- Approach
  - Theory-Based Impact Evaluation
  - 5 cross-cutting work packages (synthesis, data, **effects of funding**, territorial, crisis)
  - 8 thematic work packages (RTDI, ICT, SMEs, climate & environment, transport, employment & education & social, institutional capacity, territorial cooperation)
- Zoom on quantitative methods:
  - Descriptive / multivariate analysis
  - Macroeconomic modelling (RHOMOLO)
  - Counterfactual methods

European Commission

Methodology

- Cohesion policy in all EU regions
  - Funding function of GDP pc./EU GDP pc. (2014-2020: <75%; 75%-90%; <90%)
  - Also other factors, e.g. unemployment; also safety nets, capping, etc.
- Counterfactual? Macro
  - RDD at supposed-to-be cutoffs
  - “Generate” control group at macro level: machine learning and synthetic control
- Counterfactual? Micro
  - CIE case studies
  - RCTs?

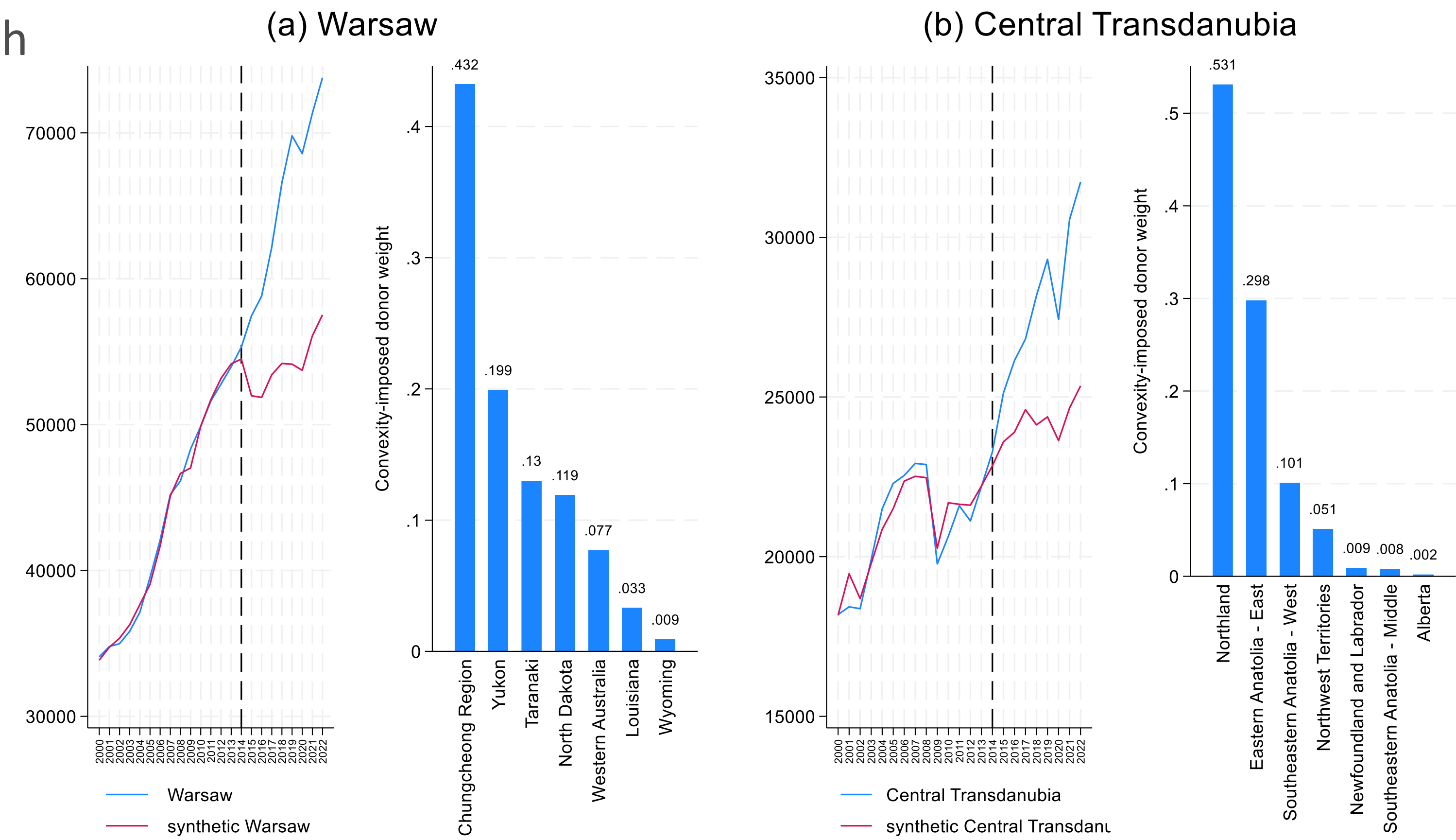
Main ideas of MLCM

- Cerqua, A., Letta, M., & Menchetti, F. (2024). Causal inference and policy evaluation without a control group. [SSRN Working Paper]
- Lots of data for all EU regions -> 1994-1999, 2000-2006, 2007-2013, 2014-2020
- Several different ML techniques are trained on 2/3 of pre-intervention data
- Forecasts compared with the remaining 1/3 real outcome data
- The most accurate ML wins the horse race (stochastic gradient boosting wins over random forest, bagging, LASSO and OLS)
- Average Treatment Effect -> ATE = E[Y(1) - Y(0)]

European Commission

Counterfactual impact evaluation methods

- Macroeconomic impact, e.g. GDP, employment
  - Regression Discontinuity Design (attempted)
  - Synthetic control: compare each beneficiary region against a counterfactual scenario based on a weighted combination of similar regions that did not receive funding



Counterfactual impact evaluation methods

- Macroeconomic impact, e.g. GDP, employment
  - Regression Discontinuity Design (attempted)
  - Synthetic control: compare each beneficiary region against a counterfactual scenario based on a weighted combination of similar regions that did not receive funding (diff-in-diff)
  - Machine Learning Control Method: historical data to forecast counterfactual scenarios
- Case studies, esp. concerning SME support
  - Matching, DID in FL, MT, PL, PT
  - MS evaluations (>300)

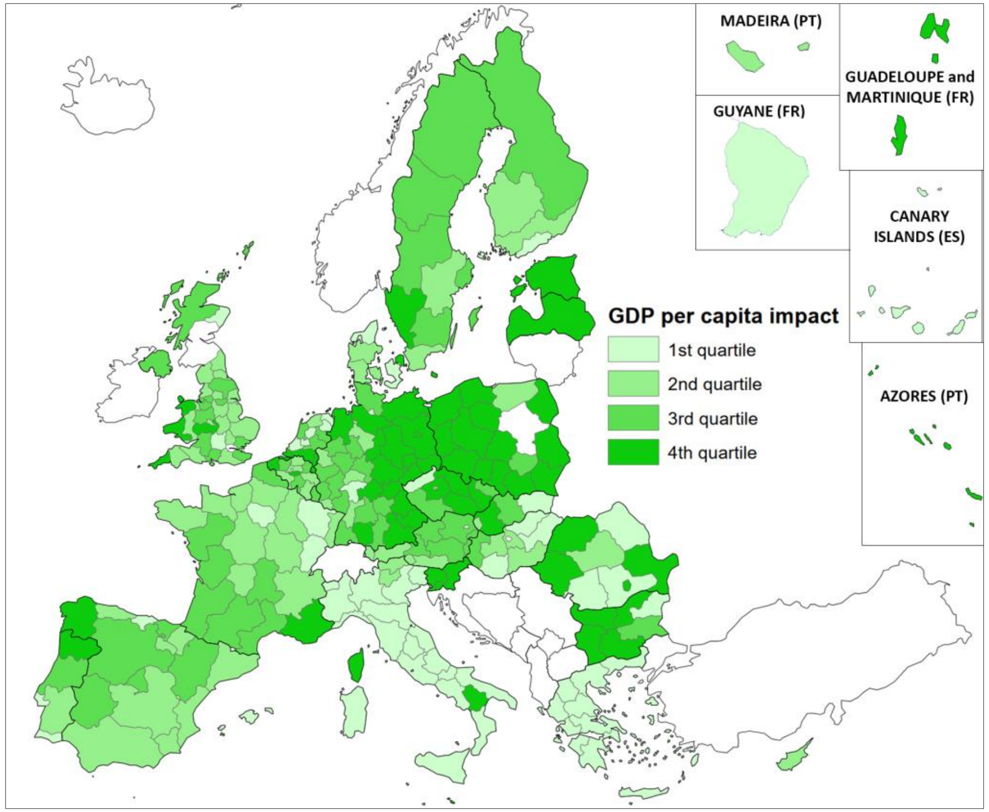
Macro – Main findings

Outcome	EU	Regions
GDP p.c.	+1.3% / +5.3%	LD: +4.5% / +6.4%
Employment	+3%	LD: +4.3%
… in advanced manufacturing	+2.1 pp.	LD: +1.5 pp.
… in high-skilled services	+0.9 pp.	LD: +0.8 pp.
Unemployment	-2.7 pp.	LD: -4.1 pp
Labour productivity	+5.3%	

- Overall positive outcomes, spatial heterogeneity
- Concentration of positive effects in less developed regions, hinting to convergence
- Interpretation needs caution awareness of assumptions

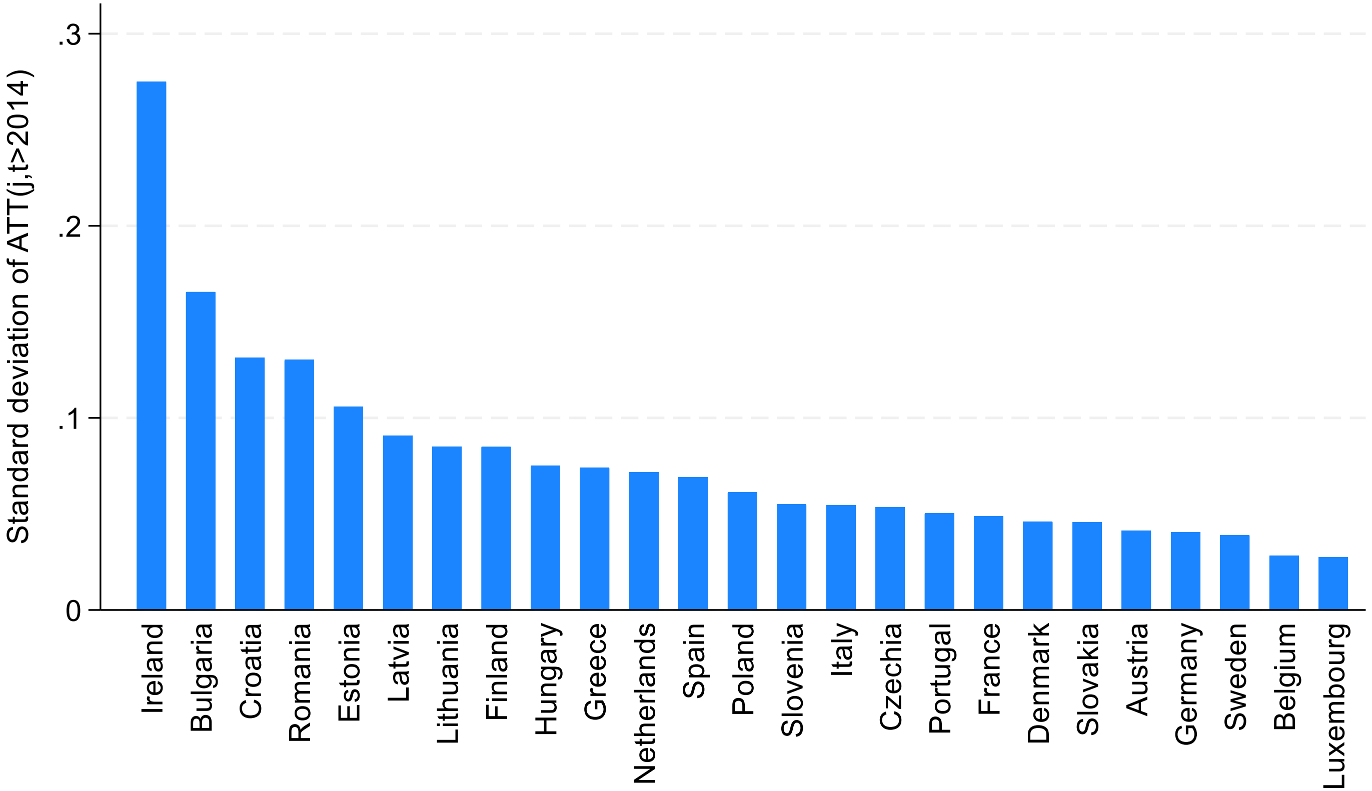
Macro – Spatial heterogeneity

Impact GDP per capita across EU regions



Cerqua and Letta (2025)

GDP per capita within-country effect dispersion



Spruk (2025)

European Commission

Data challenges and solutions: Macro

- Relevant: Models; macro-CIEs
- Challenges:
  - Rich in content: planned/selected/spent amounts, very granular thematic split, etc.
  - But: cohesion policy data from MS to EC via SFC: made for monitoring, not evaluation
  - **Geo** -> Not all data at NUTS 3/2 -> enrichment (e.g. Kohesion and surveys) & estimation
  - **Time** -> Only approximate expenditure time path (reporting delay, few times a year)
  - **Beneficiary** -> not always clear; cross-period identification; headquarter effect?
  - **Transmission** -> Slow, manual (i.e. scope for errors), hard to enrich

Outlook

- Consolidation of theory-based approach & focus on impacts across EU
- Stronger focus on causal inference and capacity support for MS / MAs
- Trial projects to encourage experimentation
- Automatic data transmissions + link with admin data + system-to-system integration
- More granular data (beneficiary, project, geo, time) = higher analytical capacity

European Commission

IV. Managing counterfactuals: pitfalls & ideas

European Commission

Typical goals

- Sound methodology to test ToC and claim causality
- Complement other qualitative and quantitative methods
- Influence policymaking to improve policies
- Communication / political push for numbers

European Commission

Typical constraints

- Lack of in-house skills
- Lack of in-house time capacity
- Small local evaluation market
- Information asymmetry with hierarchy/political level
- DATA DATA DATA

Follow the right sequence

- The ToC comes first -> expected causal pathway from intervention to outcome
- Clarify the needs (transparency, accountability, engagement, etc.) and goals (e.g. influence policymaking) of the evaluation -> ideal outcome
- Ex ante mapping of data needs and availability -> clarify constraints and pick the right question(s) and outcomes (right variables are key)
- Decision on methods -> evaluations respond to different needs, goals and constraints that normally require a combination of different methods
- Counterfactual can be part of the picture (which we encourage)

Pick the best counterfactual method

- RCTs work only in certain contexts (selection, ex ante, level of analysis, cost…)
- Not all types of interventions allow for all types of methods
- Question at hand + outcome selection + data availability drive the decision
- Allow time for discussion with relevant experts and broader community
- If contracting, the counterfactual method doesn’t need to be decided ex ante

Contracting

Main advantages: access academic skills + functional independence + capacity

Main risk: wrong/unclear design + information asymmetry + over-spending

- You should have some level of technical understating, the more the better
- If time constraints or incomplete skills, outsourcing is a good option
- If total lack of skills, it could be near impossible to ensure value for money
- Ideally internal capacity before contracting **-> The Commission can support**

Small/specific CIE contract -> you combine & own evaluation components

Next: Q&A, discussion, exchange of experience

Thank you!

Contact: davide.ceccanti@ec.europa.eu

© European Union 2025

Unless otherwise noted the reuse of this presentation is authorised under the CC BY 4.0 license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.