

Ökonometrische Methoden zur Evaluation in der Forschungs- und Technologiepolitik: Ihre Stärken und ihre Grenzen in der Praxis

Dirk Czarnitzki, KU Leuven & ZEW Mannheim
Dept. of Management, Strategy and Innovation

Einleitung

- ❑ Ökonometrische Methoden werden heutzutage häufig genutzt, um die Effekte von Innovations- und Technologiepolitik abzuschätzen.
- ❑ Sie können für verschiedenste Arten von Programmen verwendet werden
 - Direkte Förderung in Form von (FuE) Zuwendungen
 - Indirekte Förderung, z.B. steuerliche FuE-Förderung
 - Innovationswettbewerbe (Preise)
 - Rahmenbedingungen im Patentwesen
 - Öffentliche Beschaffung innovativer Produkte
 - Innovation Vouchers
 - Etc.
- ❑ Es gibt zahlreiche Methoden, die auf verschiedenen Annahmen beruhen.
- ❑ Nicht jede Methode eignet sich für jede beliebige Evaluationsfrage.

Inhalt des Vortrages

- ❑ Nicht-technische Beschreibung der zugrundeliegenden Intuition der meistgenutzten, mikro-ökonomischen Methoden
- ❑ Hervorhebung von „Fallstricken“,
 - d.h. ökonomische Annahmen, die oft nicht hinterfragt/getestet werden
 - und die möglichen Konsequenzen
- ❑ Veranschaulichung an einigen Beispielen
- ❑ Grenzen der Methoden in der Praxis im Feld der Technologiepolitik
- ❑ Diskrepanz zwischen akademischer Literatur und Interesse der Politik

Ökonometrische Methoden zur Evaluierung

- ❑ Matching (und Propensity-Score Weighting)
 - Vergleich Y subventionierter Unternehmen mit ähnlichen nicht-geförderten.
- ❑ Difference-in-Difference (DiD)
 - Vergleich ΔY subventionierter Unternehmen mit anderen.
- ❑ Conditional Differences-in-Difference (CDiD)
 - Vergleich ΔY subventionierter Unternehmen mit ähnlichen nicht-geförderten.
- ❑ Regression Discontinuity Designs (RDD)
 - Nutzung von zusätzlichen Bedingungen zur Programmteilnahme
- ❑ Natürliche Experimente
 - Subventionen werden nicht zufällig vergeben, aber man benutzt z.B. exogene Veränderungen in Teilnahmebedingungen, um Effekte zu identifizieren (keine Selbstselektion der Teilnehmer).
- ❑ Randomized Control Trials (RCT)
 - Subventionen werden (teilweise) zufällig vergeben
- ❑ Synthetic Control Methods
 - Methode für „Fallstudien“ (ähnlich wie CDiD)

Randomized control trials

❑ Oft als „Goldstandard gehandelt“

- Förderung wird (teilweise) zufällig vergeben.
- Man kann das Experiment unter verschiedenen Rahmenbedingungen wiederholt durchführen.

❑ Beispiel: Sollte man sein Frühstücksei „anpiksen“, wenn man es morgens kocht?

❑ Eignung für Technologiepolitik (TP) (meiner Meinung nach) begrenzt:

- Will man Zuwendungen oder Steuerermäßigungen zufällig vergeben?
- Experimentieren sehr kostenintensiv, z.B. direkte Projektförderung in Flandern durchschnittlich € 220.000 pro Projekt
- Bei Zuwendungen findet ein Peer-Review Prozess statt, um die „schlechten“ Projekte herauszufiltern. Warum sollte man einen positiven Effekt finden, wenn man „zufällig“ fördert?
- Bei TP-Programmen häufig zu kleine Fallzahlen, um teilweise zu randomisieren.

Randomized control trials: Beispiel

- ❑ Christensen/ Kuhn/ Schneider/ Sorensen (2021), working paper
- ❑ Dänisches “knowledge voucher support program”
 - Innovation vouchers
 - KMU Zugang zu angewandeter Forschung
- ❑ Projektträger wurde von der Vielzahl der Anträge überrascht.
 - Keine Kapazität für Selektion.
 - Lotterie!

“We find that

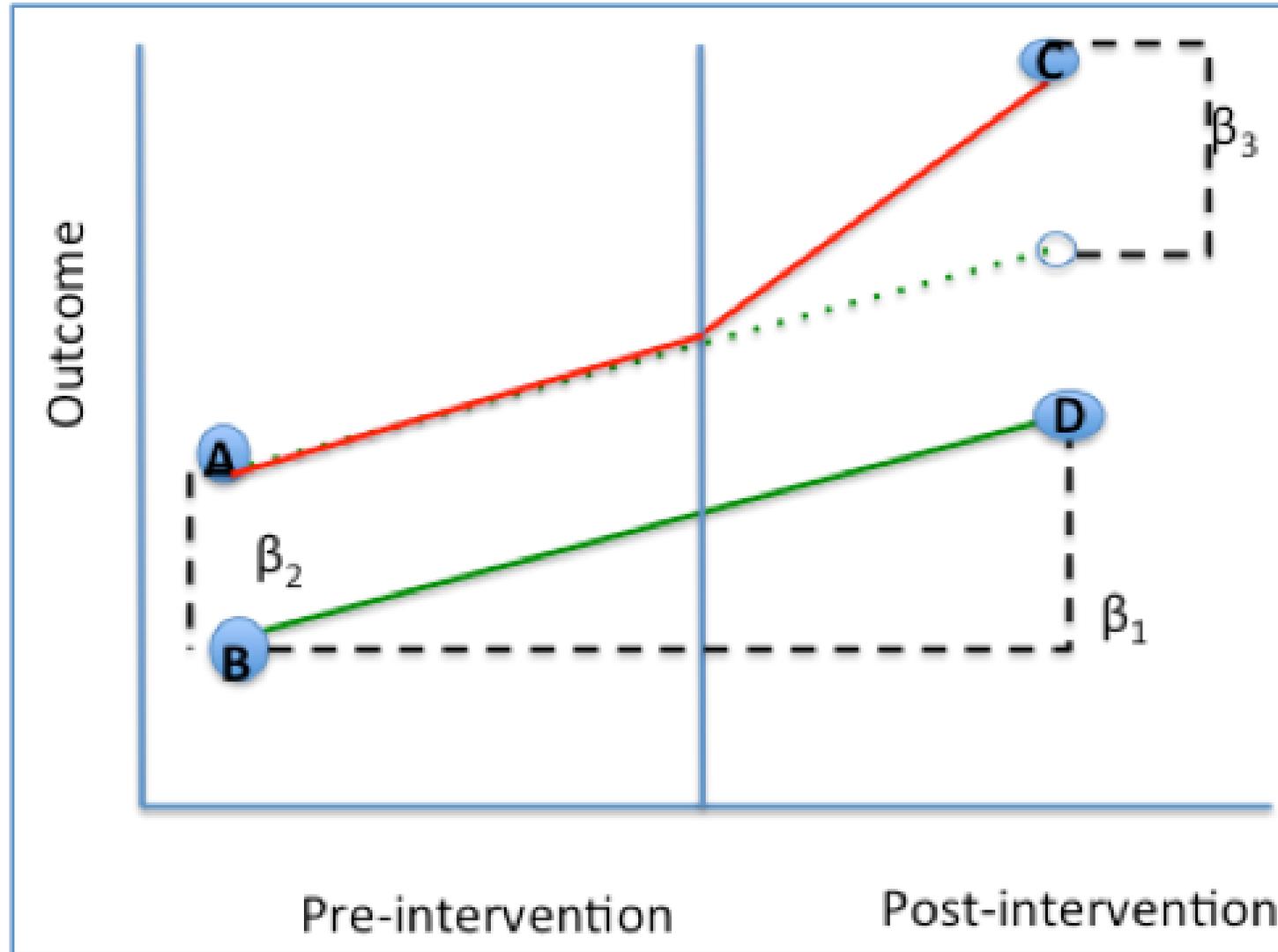
- I. the average effect of the consulting intervention on **profits and sales is not statistically** significant; we cannot reject crowding out,
- II. the program has **heterogeneous outcomes**, lottery winners in the upper end of the sales and profit distributions benefit substantially from the intervention,
- III. the treatment **did not affect employment and intermediate inputs.**”

Matching

- ❑ Eine der meist gebrauchten Methoden.
- ❑ Man vergleicht geförderte Unternehmen mit ähnlichen nicht-geförderten Unternehmen.
- ❑ Problem: der Vergleich kann nur auf Basis von beobachtbaren Variablen durchgeführt werden,
 - z.B. Unternehmensgröße, frühere Innovationserfolge, Sektor, Eigentümerstruktur, Liquidität, etc.
- ❑ Man kann stets argumentieren, dass unbeobachtete Determinanten den Antrag zur Förderung beeinflussen (Selbstselektion) bzw. die Förderung („picking the winner“)
 - Firma A hat kreativere Köpfe als Firma B.
 - Firma A wird mehr in FuE investieren als Firma B.

Difference-in-Difference (DiD)

- ❑ Man analysiert die Veränderung der Zielvariablen von geförderten und nicht-geförderten.
- ❑ Idee: Durch die Betrachtung der Veränderung der Zielvariablen beseitigt man Niveauunterschiede in der Zielvariablen
 - Firma A hat kreativere Köpfe als Firma B
-> investiert permanent mehr in FuE
 - Wenn wir Firma A an zwei Zeitpunkten beobachten - vor und während/nach der Förderung – und B bekommt nie Förderung, kann der Fördereffekt durch die Veränderung über die Zeit relativ zur Kontrollgruppe ermittelt werden.

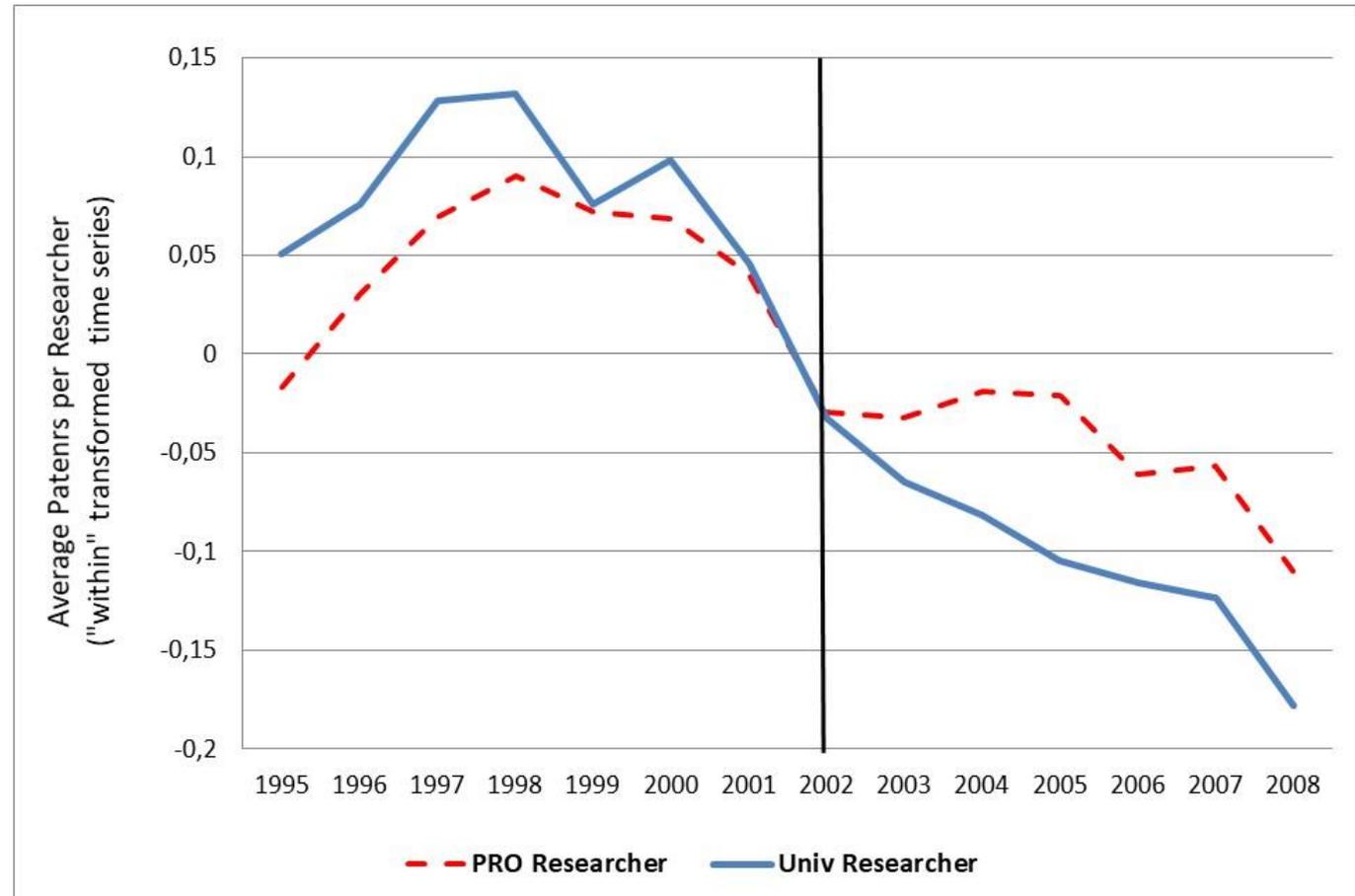


Problem

- ❑ Die Tatsache, dass A in der zweiten Periode einen Förderantrag stellt, kann daraus resultieren, dass die Forschergruppe eine sensationelle Idee hatte.
- ❑ Resultat: A könnte auch ohne Förderung in Investition in FuE erhöhen – ob Förderung erhalten wird oder nicht.
 - Sowohl Zielvariable als auch Förderstatus werden von der unbeobachtbaren „sensationellen Idee“ beeinflusst.
 - Fördereffekt wird überschätzt.
- ❑ DiD ist eigentlich gemacht für Situationen mit exogener Selektion
 - z.B. Änderung von Gesetzen
 - Dann ist die Anwendung ein „natürliches Experiment“ (keine Selbstselektion)

DiD als natürliches Experiment

- Czarnitzki/Doherr/
Hussinger/Schliessler/Toole
(2016, EER)
- Abschaffung des
Hochschullehrerprivilegs in DE
- Intellektuelles Eigentum wurde
von Profs auf Hochschule
übertragen. Sollte
Technologietransfer in der
Wirtschaft stärken.
- Gesetzesänderung: Professoren
konnten sich nicht „selbst-
selektieren“
- Vergleich: Patentanmeldungen
von Professoren vs.
Wissenschaftlern an ÖFE.
- Negativer Effekt der Politik!



Conditional Difference-in-Difference

- Wie DiD, aber man versucht die Annahmen besser zu approximieren, indem man als Kontrollgruppe nur vergleichbare, nicht-geförderte Unternehmen nutzt.
- Kombination aus DiD mit Matching.
- Praxis: man testet „common trends“
 - Bedeutet: man hofft dass z.B. die Forscher beider Firmen A und B zu jedem Zeitpunkt ähnlich gute Ideen haben.

Beispiel: Czarnitzki, 2021, Report Flämische FuE-Förderung (heute Nacht um ca. 1:30h fertig geworden...)

Innovation input: ln(R&D employees)

	No matching	NN1 PS matching	NN2 PS matching
	II	III	IV
	Coef.	Coef.	Coef.
	(Std. err.)	(Std. err.)	(Std. err.)
TREAT	0.448***	0.345***	0.385***
	(0.074)	(0.098)	(0.095)
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Number of obs.	8,976	1,924	2,782
Number of firms	3,276	629	858
adj. R-sq	0.752	0.729	0.739
Common pre-trend	Rejected at 5%	Not rejected	Not rejected

Notes: *** (**, *) indicate a significance level of 1% (5%, 10%).

Regression Discontinuity Design

- ❑ Man benutzt exogene Teilnehmerrestriktionen (oder ähnliches) um Effekte zu identifizieren.
- ❑ Beispiele:
 - 1) Programm gilt nur für kleine Unternehmen: Man vergleicht Zielvariable von kleinen Unt., die gerade noch teilnahmeberechtigt sind, mit etwas „zu großen Unt.
 - 2) Innerhalb der Teilnehmer: 10% Förderbonus für mittelgroße, 20% für kleine, im Vergleich zu großen Unternehmen.
 - 3) Qualitätsminium beim Peer-Review Score. Vergleich nur in der Nähe des Minimums.
- ❑ Probleme:
 - Definition des Bereichs um den Threshold Wert (trade-off: kleineres Sample vs. geringere Verlässlichkeit)
 - Siehe (3): warum würde man den treatment effekt im Bereich der „schlechteren“ bestimmen wollen?

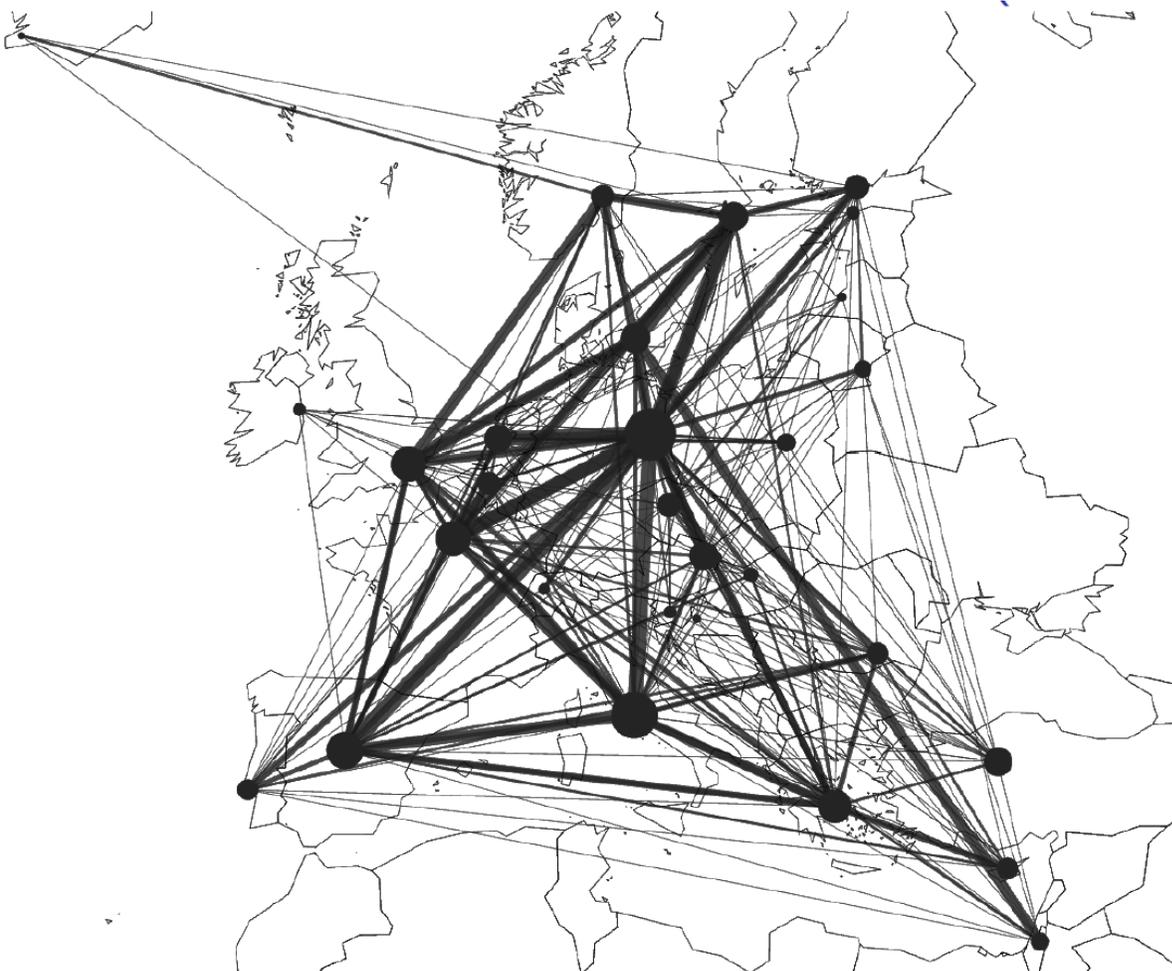
Weitere Ansätze:

- Kombination von DiD, Matching, RDD und Natural Experiment

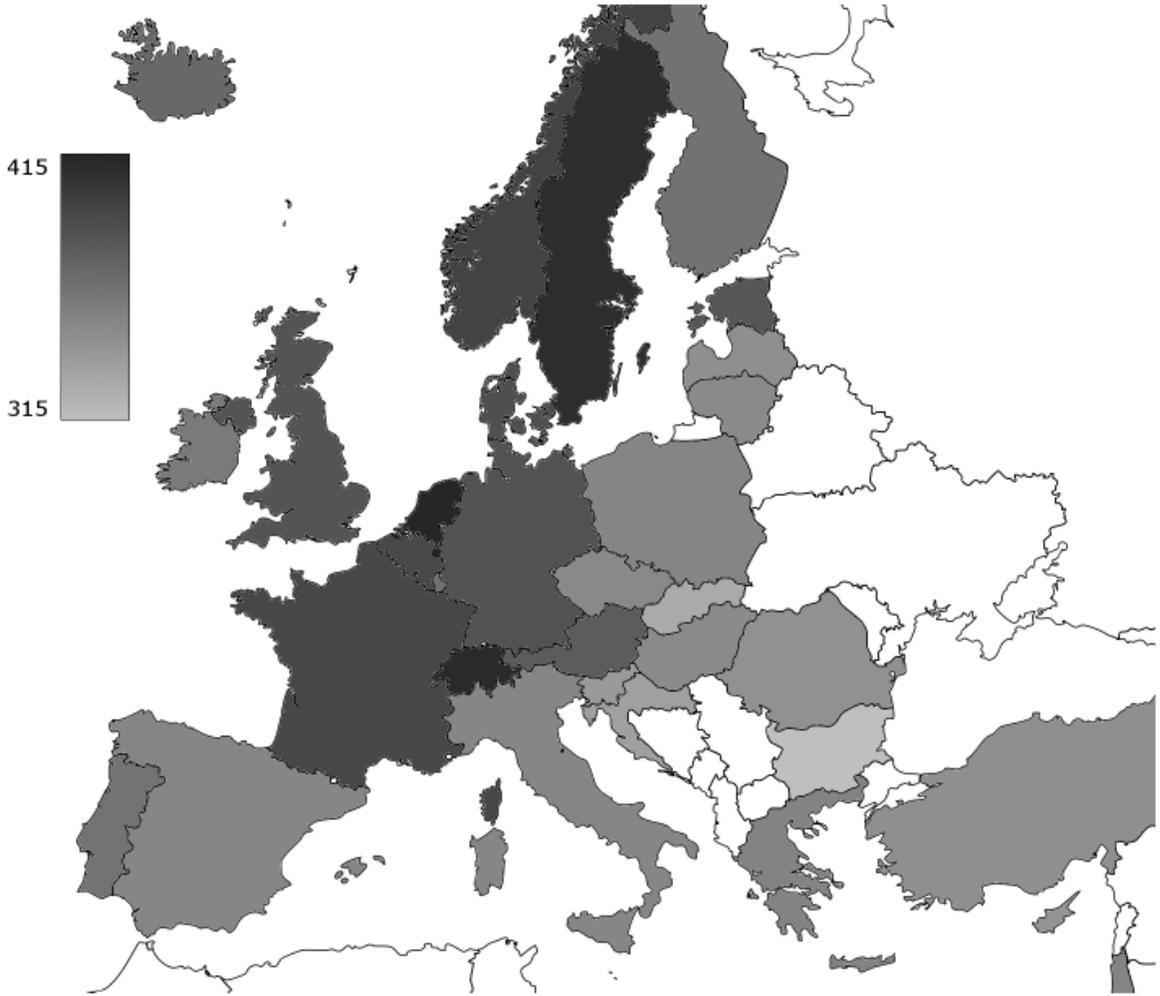
- Beispiel: Hünermund/Czarnitzki (2019, ResPol)

- Eurostars Programm: Innovative KMU
- Wir beobachten Antragsteller, sowohl gefördert als auch abgelehnt.
- Peer-Review Scores zwischen 0 und 600. Minimum threshold: 400.
- Unternehmen kooperieren international (gemeinsamer Antrag), werden aber national gefördert → „Virtual Common Pot“
- Nationale Budgets begrenzt.
- Es kommt zu „random drop-outs“ je nach Länderkombination der Antragsteller, obwohl das Forschungsvorhaben über dem Qualitätsminum liegt.
- Zielvariable in diesem Beispiel: Wachstum Beschäftigtenanzahl.

Network of international collaborations (3rd call)



Geographical distribution of project evaluation scores (range between 0 and 600)



Simulate: virtual vs. real common pot

Recall: Project evaluation scores from 0 to 600

- Minimum requirement: 400

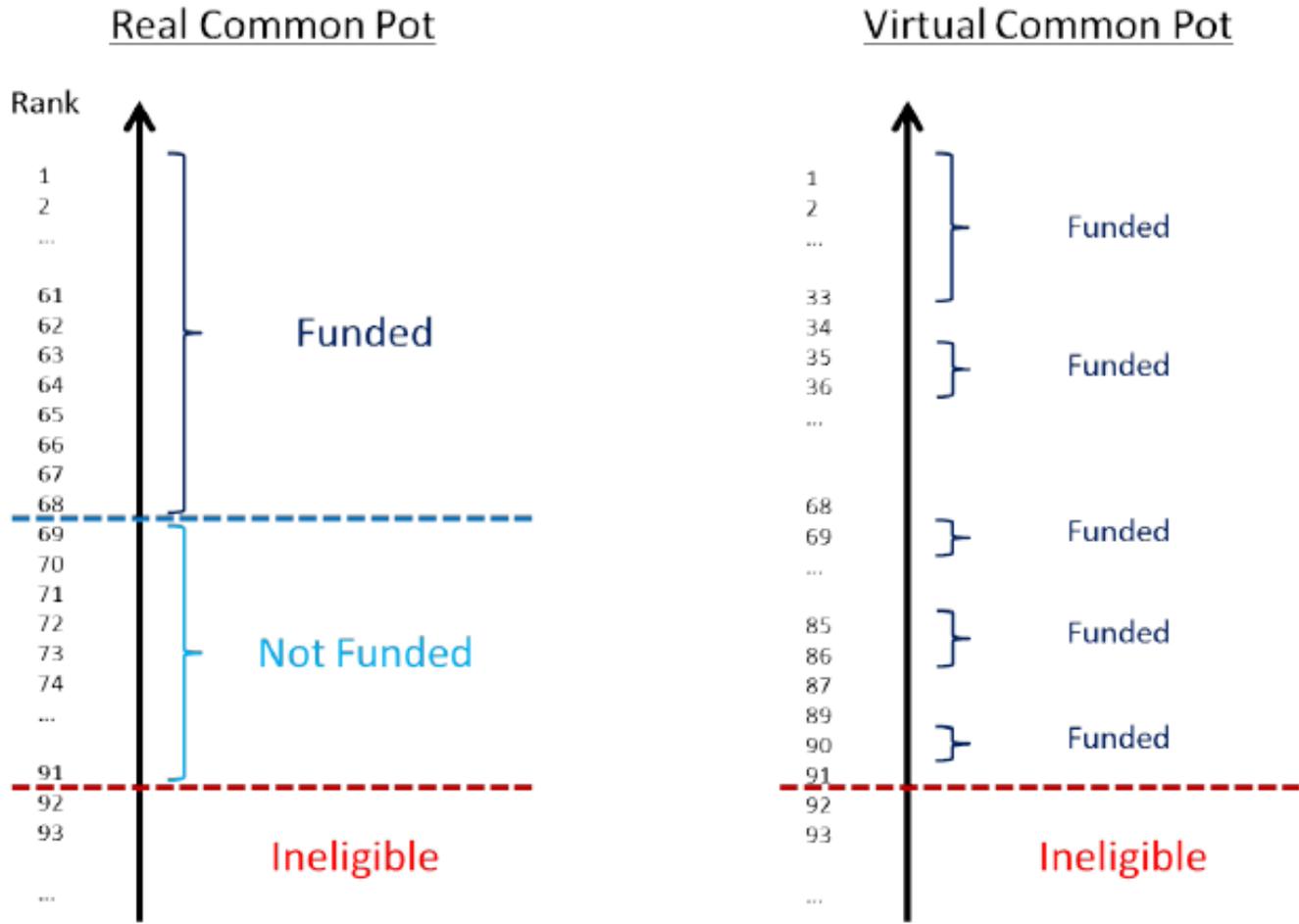
Hypothetical example

- Four countries: A, B, C, and D
- Each country has a budget to fund two firms

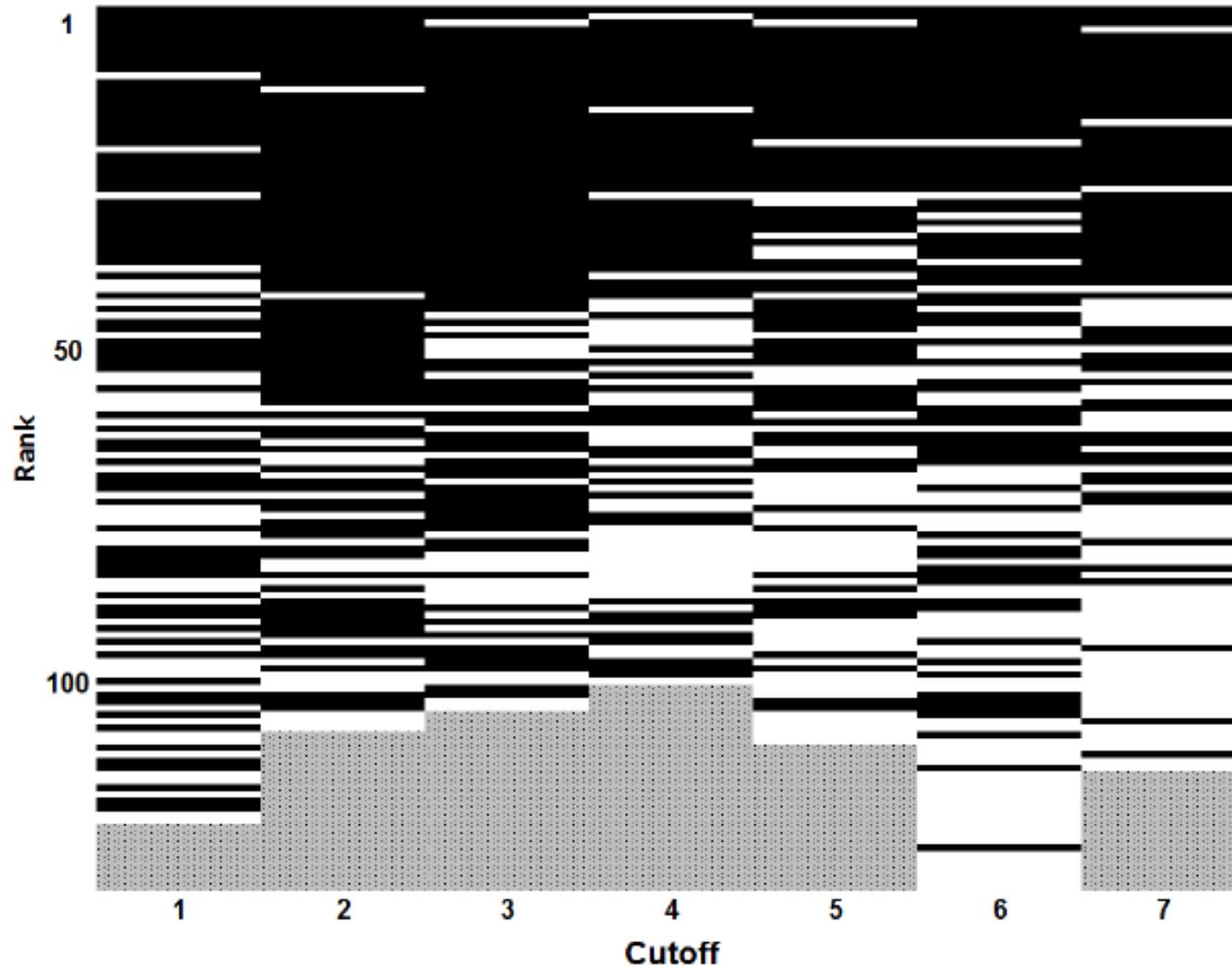
Who gets funding under VCP vs. RCP?

Rank	Consortium	VCP	RCP
1	A, B, B	✓	✓
2	B, B, C		✓
3	A, C	✓	✓
4	A, B, C, D		
5	C, D, D	✓	
6	A, C, D		

Real vs. Virtual Common Pot



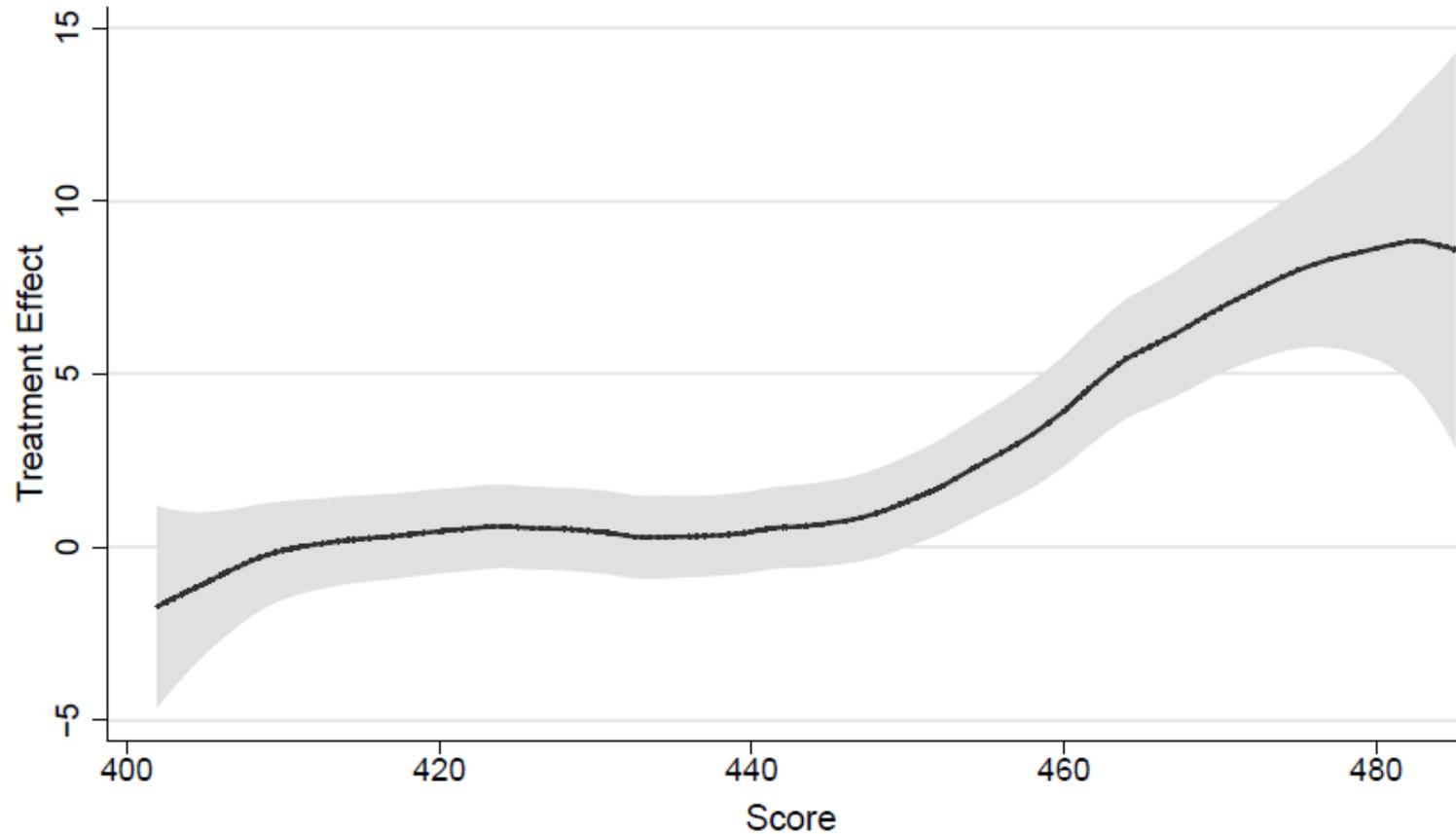
Actual allocation



Black: funded
White: not funded
Grey: below quality
threshold

Estimating LATE: Virtual vs. Real Common Pot

Figure 2: Treatment effects depending on project scores



Notes: Local polynomial regression of treatment effects on project scores. Degree = 1, kernel = epanechnikov, bandwidth = 7.56. Grey area depicts 95% confidence intervals.

Diskrepanz zwischen akademischer Forschung und Politik in der Praxis

- ❑ Akademische Literatur fokussiert die Bemühungen auf die Kausalität des geschätzten, durchschnittlichen Effektes eines Programms.
- ❑ Man kann den durchschnittlichen, kausalen Effekt in Beziehung setzen zu den Investitionen (extrapolation, back of the envelope calculation)
- ❑ Beispiel von meiner Eurostars Berechnung:
 - Extrapolation von sample auf alle Geförderten: 8679 Arbeitsplätze geschaffen.
 - Programmbudget dazu: € 314.27 Millionen für insgesamt 1556 Projekte.
 - 1 job kostete den “Eureka” Steuerzahler ca. 36,000 EUR.

Diskrepanz zwischen akademischer Forschung und Politik in der Praxis

□ Der Projektträger würde vermutlich lieber wissen:

- Für welche Teilnehmer funktioniert das Program besonders gut oder schlecht (Mitnahmeeffekte).
- Sind die Teilnahmebedingungen gut designed? Wie kann man diese verbessern?
- Sollen wir die Förderbeträge/-quote erhöhen oder verringern?
- Sollte man einer Firma nur maximal ein Projekt fördern, weil es sonst zu Mitnahmeeffekten kommt?
- → Heterogenität der Effekte. Problem: im TP-Bereich oft zu kleine Anzahl Firmen, um viele verschiedene Effekte zu berechnen.

Vielen Dank für Ihre Aufmerksamkeit!

- Fragen? → dirk.czarnitzki@kuleuven.be
- Übrigens: Frühstücksei anstechen hilft nicht!
- Bei einem Versuch des WDR vor einigen Jahren wurden 3.000 Eier gekocht, die Hälfte davon angepikst, die andere Hälfte ohne Loch.
- Ergebnis: Egal ob angestoichen oder nicht, im Test platzte etwa jedes zehnte Ei auf.