

**Was wäre wenn nicht –
und wenn doch, wie viel?
Neue Ansätze der Wirkungsforschung**

Jochen Kluge

Ringvorlesung Entwicklungspolitik

Technische Universität Berlin

10. Januar 2013

- Angenommen, Sie hätten ein neues Mittel gegen Haarausfall erfunden.
- Wie kann man wissen, ob es wirkt?
- ... und wie gut?

- In Deutschland derzeit: 3 Millionen Arbeitsuchende.
- Fortbildungsmaßnahmen: Computerkurs, Umschulung, etc.
- Wie kann man wissen, ob der Kurs den Menschen hilft?
- ... und wie viel größer ihre Chance ist, einen Job zu finden?

- Ein Mittel, eine Maßnahme
- Frage: Was bewirkt dieses Mittel, diese Maßnahme?
- Kernfrage der Wirkungsforschung

Überblick

1. Methodische Grundlagen

- Was wäre wenn?
- ... was wäre wenn nicht?
- ... und wie viel?

2. Drei Beispiele zur Illustration

- Ausbildung benachteiligter Jugendlicher in Rio de Janeiro
- Evaluation des Elterngeldes
- Berufliche Ausbildung in Vietnam

1. Methodische Grundlagen

Was wäre wenn?

Kern der Wirkungsforschung (“program evaluation / impact evaluation”) ist das Verhältnis zwischen

- i) einer Maßnahme / Intervention / Mittel
und ->
- ii) einem Ergebnis bzw. einer Ergebnisgröße

Also z.B.:

- dem Haarwuchsmittel und -> der Anzahl der Haare
- der Fortbildungsmaßnahme und -> der Wahrscheinlichkeit, eine Anstellung zu finden

Was wäre wenn?

Der faktische Zustand ist relativ leicht ermittelbar:

- Ein Proband nimmt das Haarwuchsmittel und -> man zählt hinterher die Anzahl der Haare
- 100 arbeitssuchende Personen nehmen an der Fortbildungsmaßnahme teil und -> man ermittelt hinterher den Anteil, der eine Anstellung gefunden hat

Aber: weiß man mit dieser Information bereits, ob und wie das Mittel bzw. die Maßnahme wirkt?

... und was wäre, wenn nicht?

Um die Wirkung zu ermitteln, müsste man auch wissen

- wieviel Haare der Person verblieben wären, wenn sie das Haarwuchsmittel nicht genommen hätte
- wieviele der 100 Personen eine Anstellung gefunden hätten, wenn sie nicht an der Fortbildungsmaßnahme teilgenommen hätten

Entscheidend dabei: es geht um

- den selben Probanden
- die selben 100 Personen

... und was wäre, wenn nicht?

Diese Aussage darüber, was mit derselben Person(en)gruppe) geschehen wäre, wenn sie nicht das Mittel eingenommen bzw. an der Maßnahme teilgenommen hätte, wird als die

kontrafaktische Situation bzw. **“counterfactual”**

bezeichnet.

... und wie viel?

Die **Differenz** aus faktischer und kontrafaktischer Situation misst die Wirkung / den **kausalen Effekt** der Maßnahme:

- Differenz aus Anzahl der Haare des Probanden *mit* Haarwuchsmittel und Anzahl der Haare des Probanden *ohne* Haarwuchsmittel
- Differenz aus Anteil der Personen *mit* Fortbildung, die einen Job finden, und Anteil der Personen *ohne* Fortbildung, die einen Job finden

Kern dieser Methodik:

- Zwei Zustände, die sich durch nichts unterscheiden, außer durch die Maßnahme und die damit verbundene Ergebnisgröße (*ceteris paribus*).

Probleme bzw. methodische Herausforderungen

- Die kontrafaktische Situation ist nicht beobachtbar und daher hypothetisch -> zwei “Zustände der Welt”
- Einzelperson nicht doppelt beobachtbar -> Gruppen
- Teilnahme- bzw. Kontrollgruppe -> müssen identisch bzw. sehr ähnlich bzw. vergleichbar sein
- Man misst empirisch den sog. durchschnittlichen Teilnahmeeffekt auf die Teilnehmer, “Average treatment effect on the treated” ATET

$$E(Y_1 | D = 1) - E(Y_0 | D = 1)$$

Lösungsmöglichkeiten

Idealfall: Kontrolliertes Zufallsexperiment

- 100 Männer, von denen per Zufall 50 einer Teilnahmegruppe zugeordnet werden, die das Haarwuchsmittel bekommt, und 50 einer Kontrollgruppe, die ein Placebo erhält
- Analog: Von 200 arbeitssuchenden Personen werden 100 zufällig der Fortbildungsgruppe zugeordnet und 100 einer Kontrollgruppe

Lösungsmöglichkeiten

Idealfall: Kontrolliertes Zufallsexperiment

- Durch die zufällige Zuordnung sind in beiden Experimenten die Teilnahme- und Kontrollgruppe im Durchschnitt in allen beobachtbaren (z.B. Alter; Ausbildung) und unbeobachtbaren (z.B. genetische Prädisposition; Motivation) gleich.
- > Das Ergebnis der Kontrollgruppe misst die kontrafaktische Frage, “welches Ergebnis hätte die Teilnahmegruppe gehabt, wenn sie nicht teilgenommen hätte?”

Lösungsmöglichkeiten

Was macht man, wenn man kein Experiment durchführen kann?

- Sog. “natürliche Experimente”
 - Statistisches Matching
 - Differenz-in-Differenzen
 - u.a.
-
- Bei Wahl des Studiendesigns häufig sinnvoll zu fragen, wie das passende Experiment aussehen würde

2. Drei Beispiele

**Beispiel 1:
Randomisiertes kontrolliertes Experiment – Soziale
Kompetenzen Rio de Janeiro**

**Beispiel 2:
Natürliches Experiment – Elterngeld**

**Beispiel 3:
Differenz-in-Differenzen – Berufliche Bildung
Vietnam**

Beispiel 1:

Randomisiertes kontrolliertes Experiment (RCT)

—

Berufliche und soziokulturelle Bildung für benachteiligte Jugendliche in Rio de Janeiro

Hintergrund

- Debatte in der Forschung zu beruflicher Bildung in Entwicklungs- und Schwellenländern
- -> Hypothese: neben der technischen Ausbildung ist die soziokulturelle Ausbildung mindestens genauso wichtig -> „Life skills“ = soziale Kompetenzen
- Grund: Zielgruppe benachteiligter Jugendlicher ohne abgeschlossene Schulbildung, häufig mit problematischem familiären und sozialen Hintergrund
- Fokus daher auf: Innovative Ansätze, diese „Life skills“ zu vermitteln

Hintergrund

- Forschungsprojekt mit der „Inter-American Development Bank“ IDB
- 2011/2012: Besuch zweier Programme in Favelas in Rio de Janeiro
 - i) „A Ganar“ -> Sport als Fortbildungsinstrument
 - ii) „Galpao Aplauso“ -> Theater als Fortbildungsinstrument

i) “Life skills” durch Sport -> “A Ganar”

ii) “Life skills” durch Theater -> Galpao Aplauso

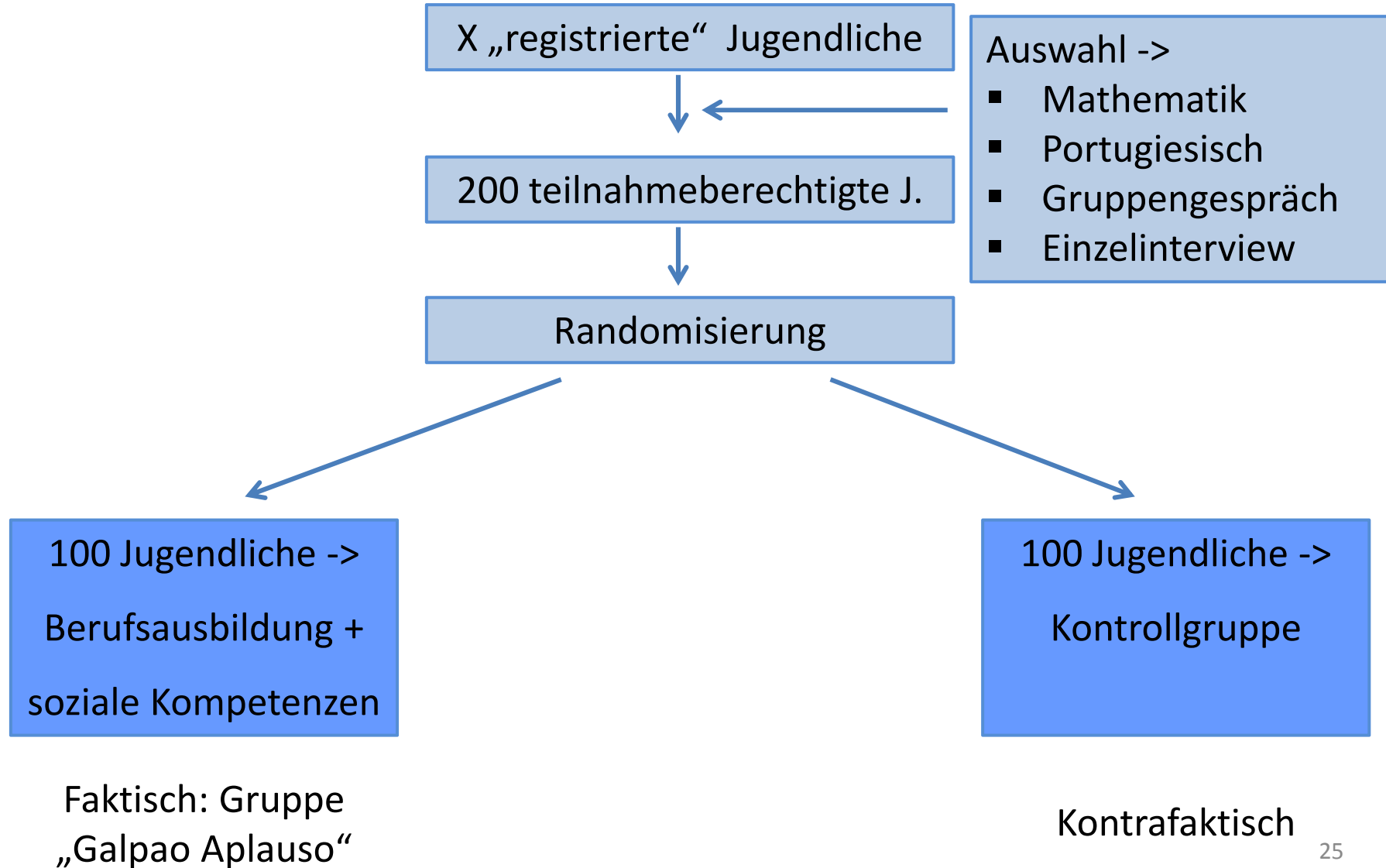
„Galpao Aplauso“

- Wie das Programm „A Ganar“ auch, zählt „Galpao Aplauso“ zu den innovativen Jugendprogrammen
- Kombination aus Berufsausbildung + Bildung sozialer Kompetenzen
- Die Jugendlichen verbringen pro Tag 3h mit technischer Fortbildung und 2h mit Bildung sozialer Kompetenzen
- Ansatz: Theater/Schauspiel
- Gesamtdauer 6 Monate

Studiendesign „Galpao Aplauso“

- Sozialwissenschaftliches Experiment → Möglich aufgrund der großen Nachfrage nach dem Programm
- (Kontrollierte) Zufallsauswahl aller teilnahmeberechtigten Jugendlichen für die Teilnahme- und Kontrollgruppe
- Keine Beeinflussung des Programms durch die Studie

RCT „Galpao Aplauso“



Befragungsdesign

- Drei Erhebungen: i) Baseline vor Programmbeginn sowie Follow-up ii) bei Programmende und iii) 6 Monate nach Programmende
- Insgesamt wurden 451 Jugendliche in Teilnahme- und Kontrollgruppe zugewiesen, davon wurden 381 in der Baseline befragt
- Persönliche Interviews (face-to-face)
- Umfangreicher Fragebogen: Soziodemographie der Person und des Haushalts, Arbeitsmarkterfahrung, Risikoverhalten, soziale Kompetenzen
- Erste Ergebnisse: Frühjahr 2013

RCTs und die Evaluation von EZ

- Interventionen der Entwicklungszusammenarbeit werden weltweit in zunehmendem Maße evaluiert, und immer häufiger auf Basis von RCTs
- J-PAL, Esther Duflo, www.povertyactionlab.org
- 3ie = International Initiative for Impact Evaluation, www.3ieimpact.org
- Deutsche EZ: s. Vortrag Herr Gajo; Evaluierungsinstitut

- Methodisch: Was kann man machen, wenn ein RCT nicht umsetzbar ist?

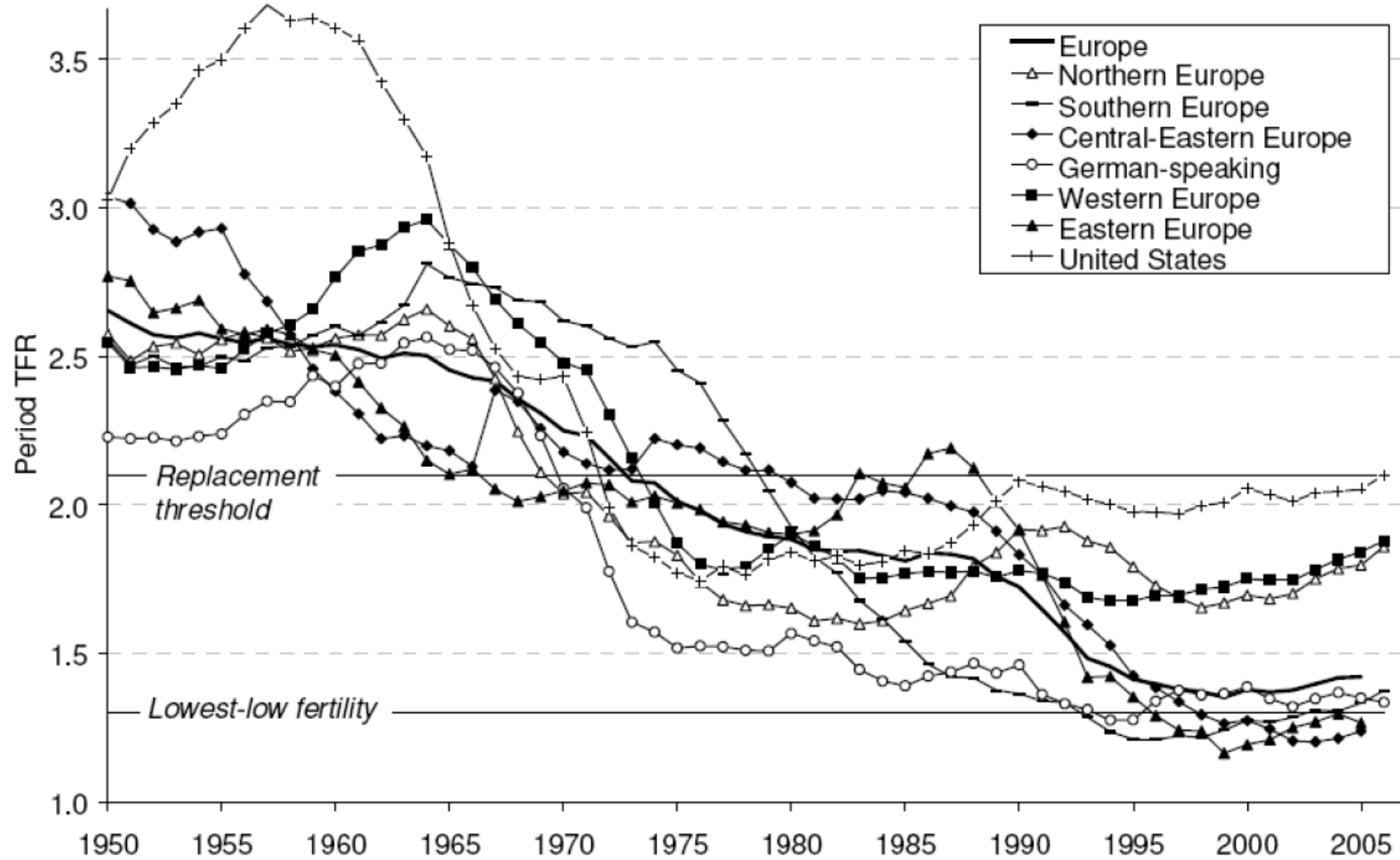
Beispiel 2:

Natürliches Experiment

—

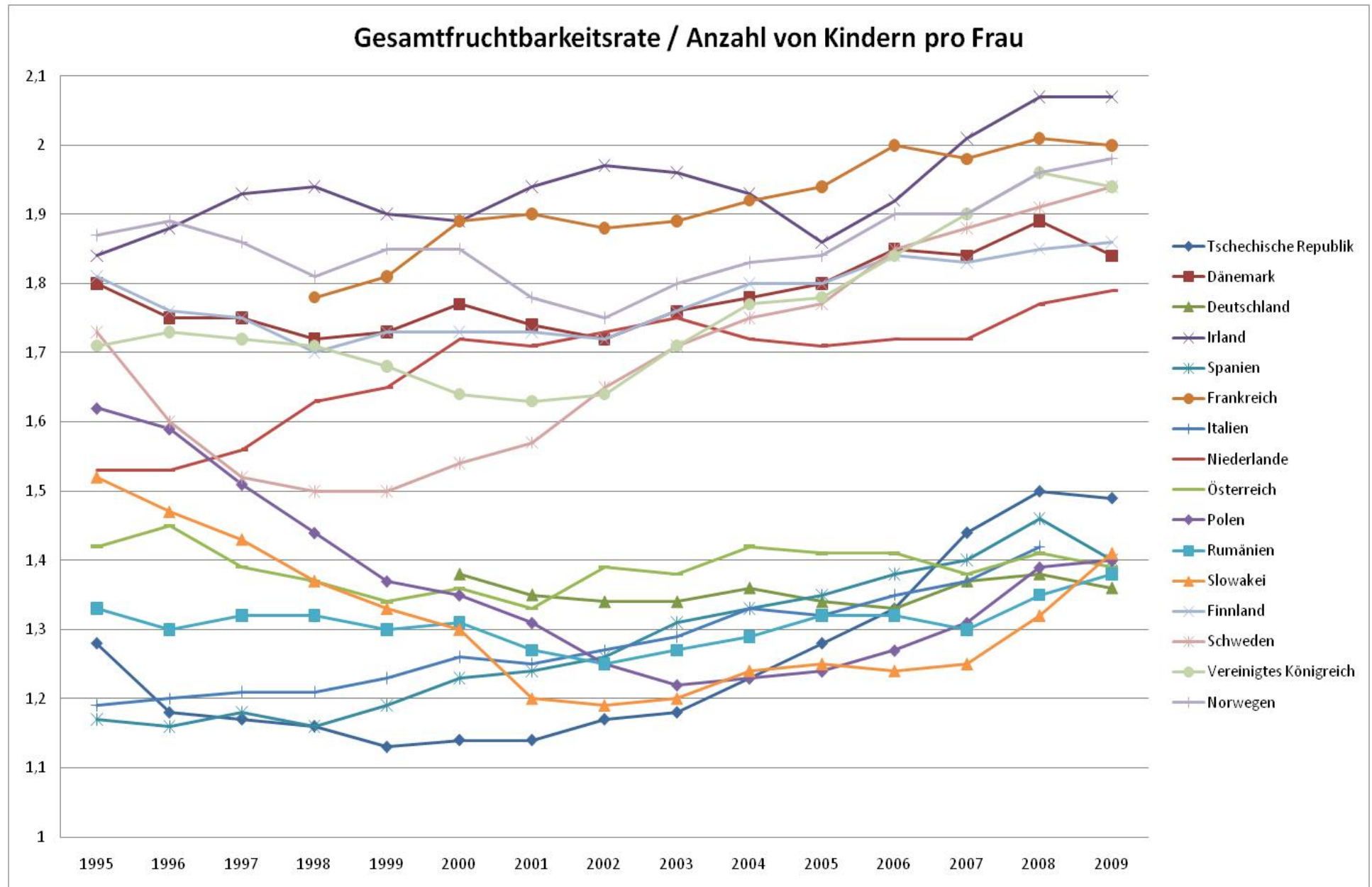
Evaluation des Elterngeldes

Kontext: Fertilitätsraten in Europa und USA, 1950–2006

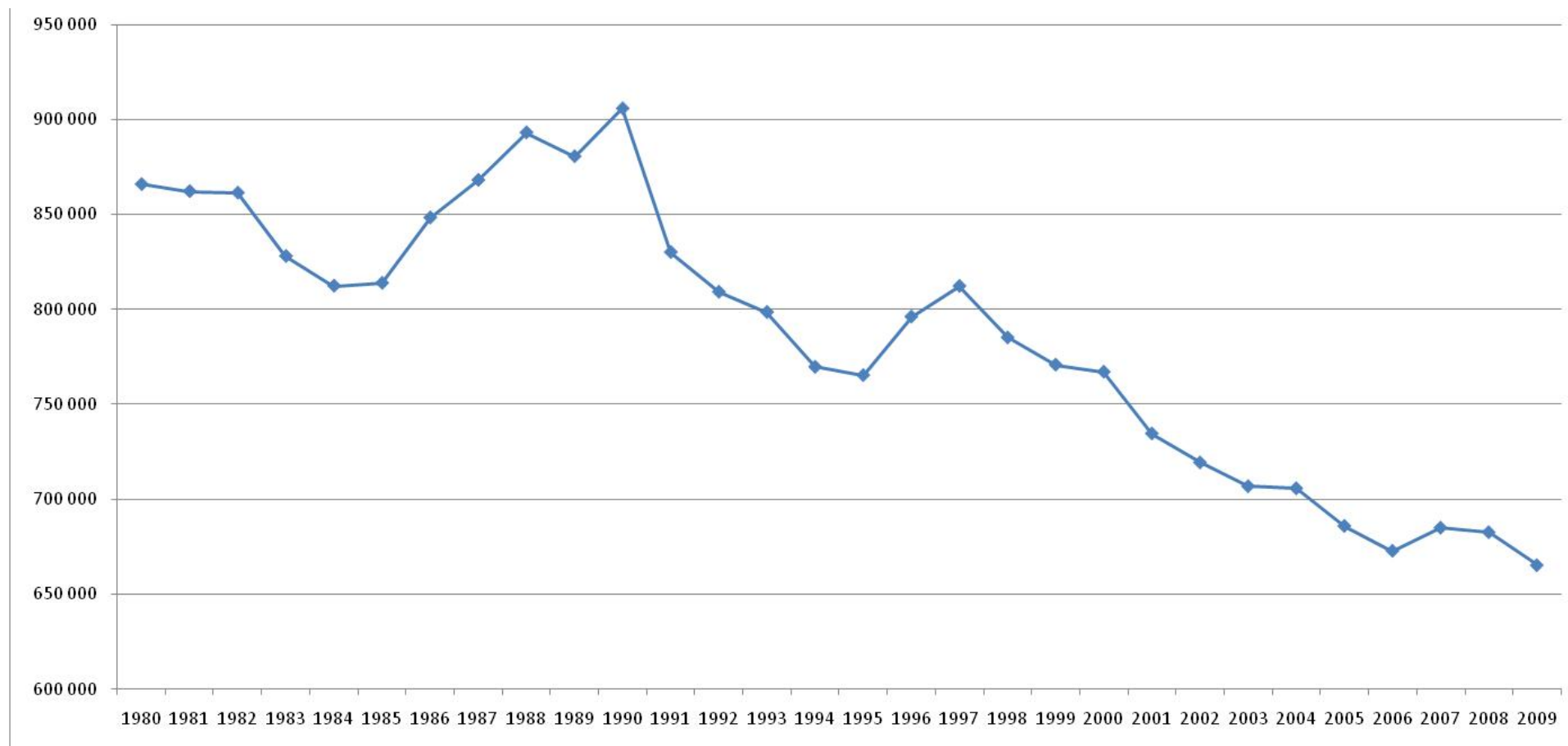


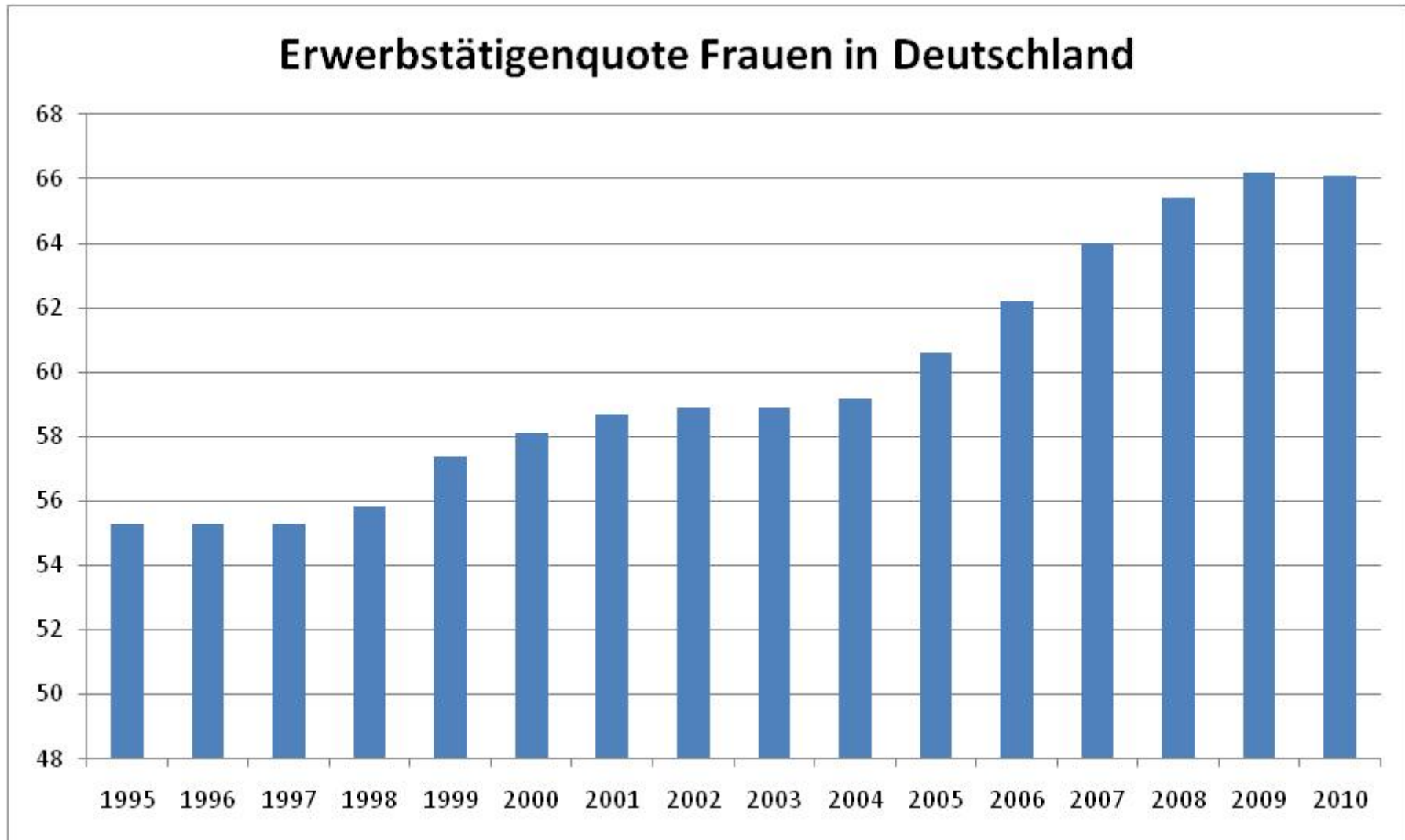
Source: Frejka and Sobotka (2008).

Kontext: Fertilitätsraten in Europa, 1995–2009



Kontext: Geburten in Deutschland, 1980–2009





Die Maßnahme: Elterngeld ab 01.01.2007

- Transitorischer Einkommenstransfer für Eltern neu geborener Kinder:
67% des vorangegangenen Arbeitseinkommens
- Maximal 1.800 Euro, mindestens 300 Euro pro Monat
- Partnermonate: 12 + 2
- Abgedeckter Personenkreis: faktisch 100% aller Familien

- ersetzt das bis 31.12.2006 geltende Erziehungsgeld
(i.d.R. 300 Euro für 24 Monate; ~75% der Familien)

Die Ziele -> Ergebnisgrößen

- **Einkommensrückgang** für Eltern nach der Geburt **abfedern** ⇒ ermöglicht eigene Betreuung des Kindes im ersten Jahr
- Schaffung von **Fertilitätsanreizen** insbesondere für erwerbstätige Frauen durch Berücksichtigung der Opportunitätskosten
- bei gleichzeitigem Versuch, Mütter nahe an ihrem Beruf zu halten und eine zeitnahe **Rückkehr in Erwerbstätigkeit** zu ermöglichen.
- Stärkere **Einbindung der Väter**

Evaluation Elterngeld: Gesetzauftrag

Gesetz zur Einführung des Elterngeldes vom 05.12.2006:

§ 25

Bericht

Die Bundesregierung legt dem Deutschen Bundestag bis zum 1. Oktober 2008 einen Bericht über die Auswirkungen dieses Gesetzes sowie über die gegebenenfalls notwendige Weiterentwicklung dieser Vorschriften vor. Er darf keine personenbezogenen Daten enthalten.

Methodische Herausforderung

Kontrafaktische Frage:

Wie hätten sich Eltern mit nach dem 01.01.2007 geborenem Kind hinsichtlich der Ergebnisgrößen in einer Situation ohne Elterngeldgesetz verhalten?

Methodische Herausforderung

Wie würde das Experiment aussehen?

- Potenzielle Eltern randomisieren in Teilnahmegruppe (Elterngeldberechtigt) und Kontrollgruppe
- > problematisch

Des Weiteren schwierig:

- Maßnahme betrifft die gesamte Population -> keine potenzielle Vergleichsgruppe derjenigen, die nicht teilnehmen
- Maßnahme intendiert Verhaltensänderungen -> Selektion

Lösungsvorschlag

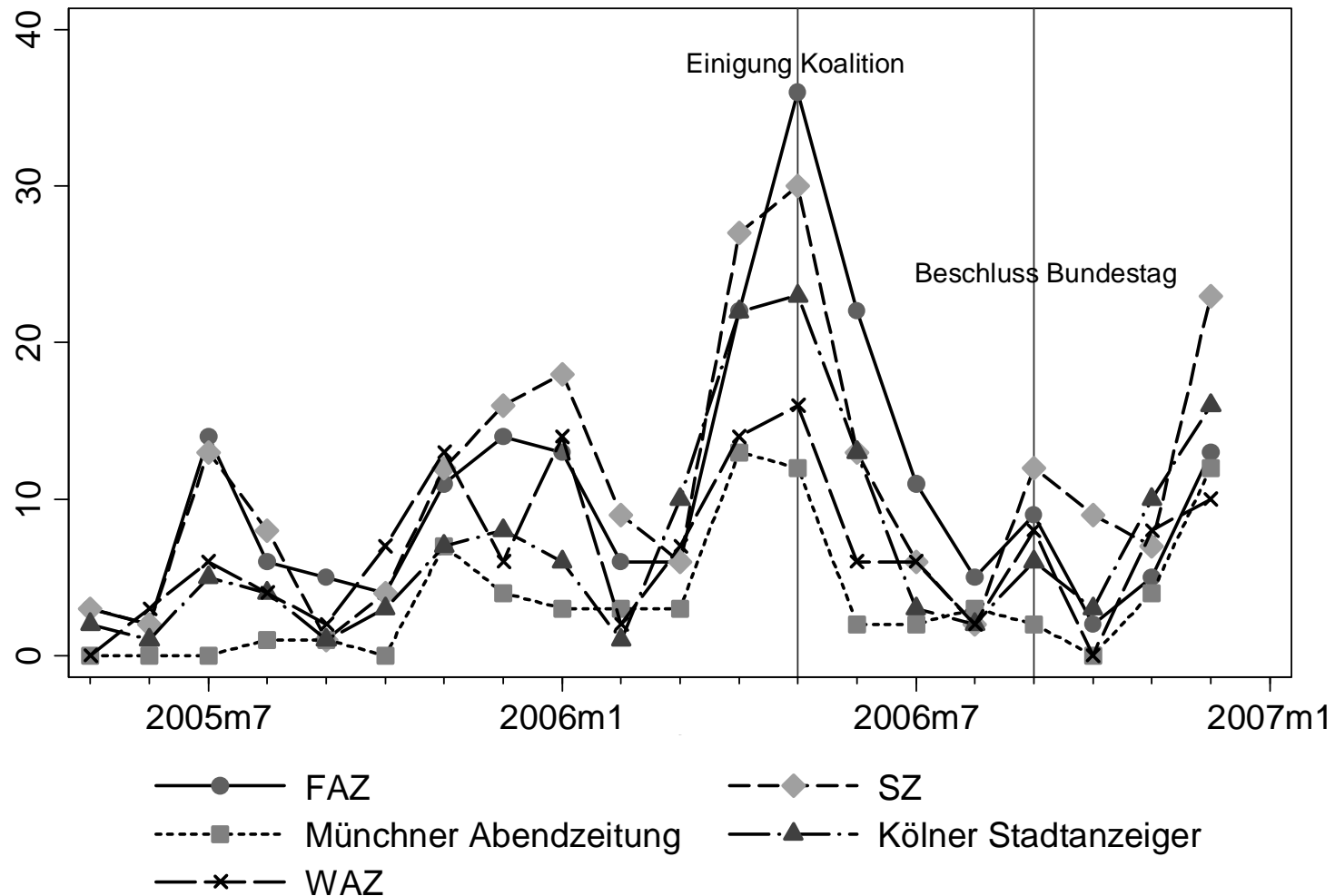
„Natürliches Experiment“

- \Rightarrow Gesetzgebungsprozess des Elterngeldgesetzes generiert zufällige Zuteilung in Elterngeld- und Vergleichsgruppe kurz nach / vor Inkrafttreten der Reform

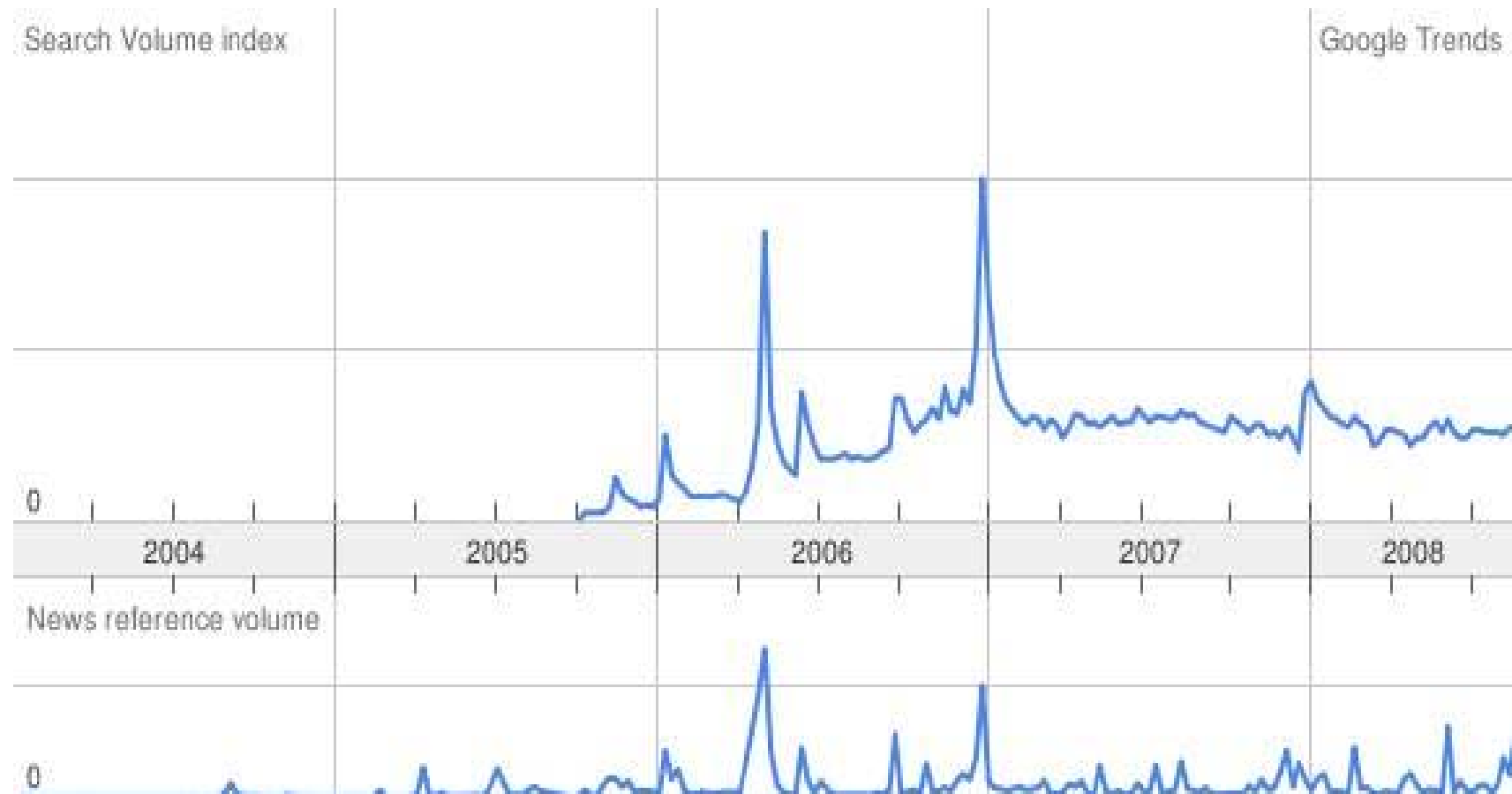
Warum?

- Zum Zeitpunkt der Zeugung jener Kinder, die kurz vor bzw. nach Inkrafttreten der Reform geboren wurden, konnten die Eltern nicht wissen, dass bei Geburt ihres Kindes die Neuregelung umgesetzt sein würde
- Vergleich Eltern Q1/07 mit Q4/06 kann Reformeffekte stichhaltig und unverzerrt messen

Methodik: Medienberichte zum Elterngeld in großen Tageszeitungen



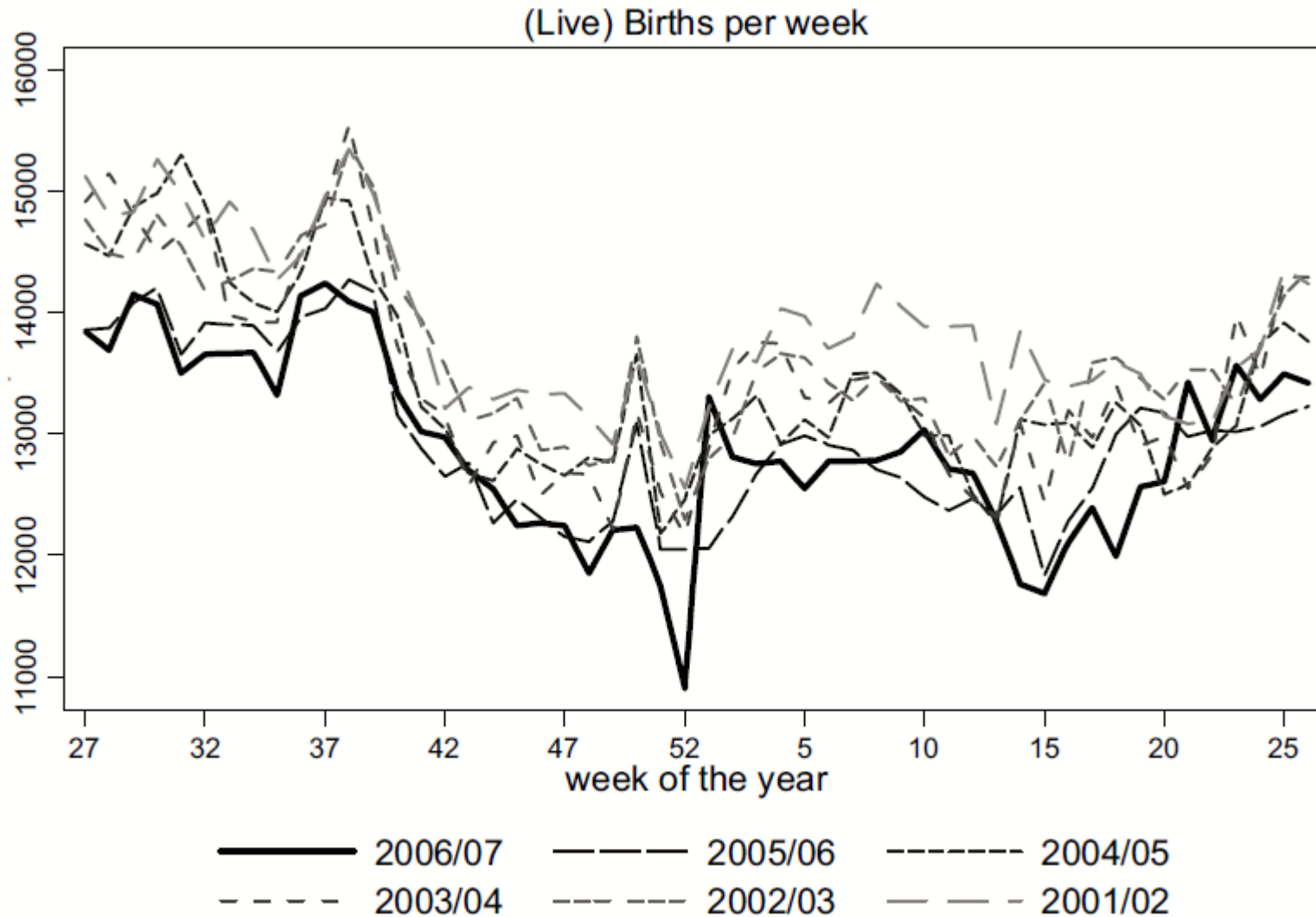
Methodik: Google Search Volume Index – Anzahl an “Elterngeld”- Suchen relativ zu allen Suchen (mit Ursprung Deutschland)



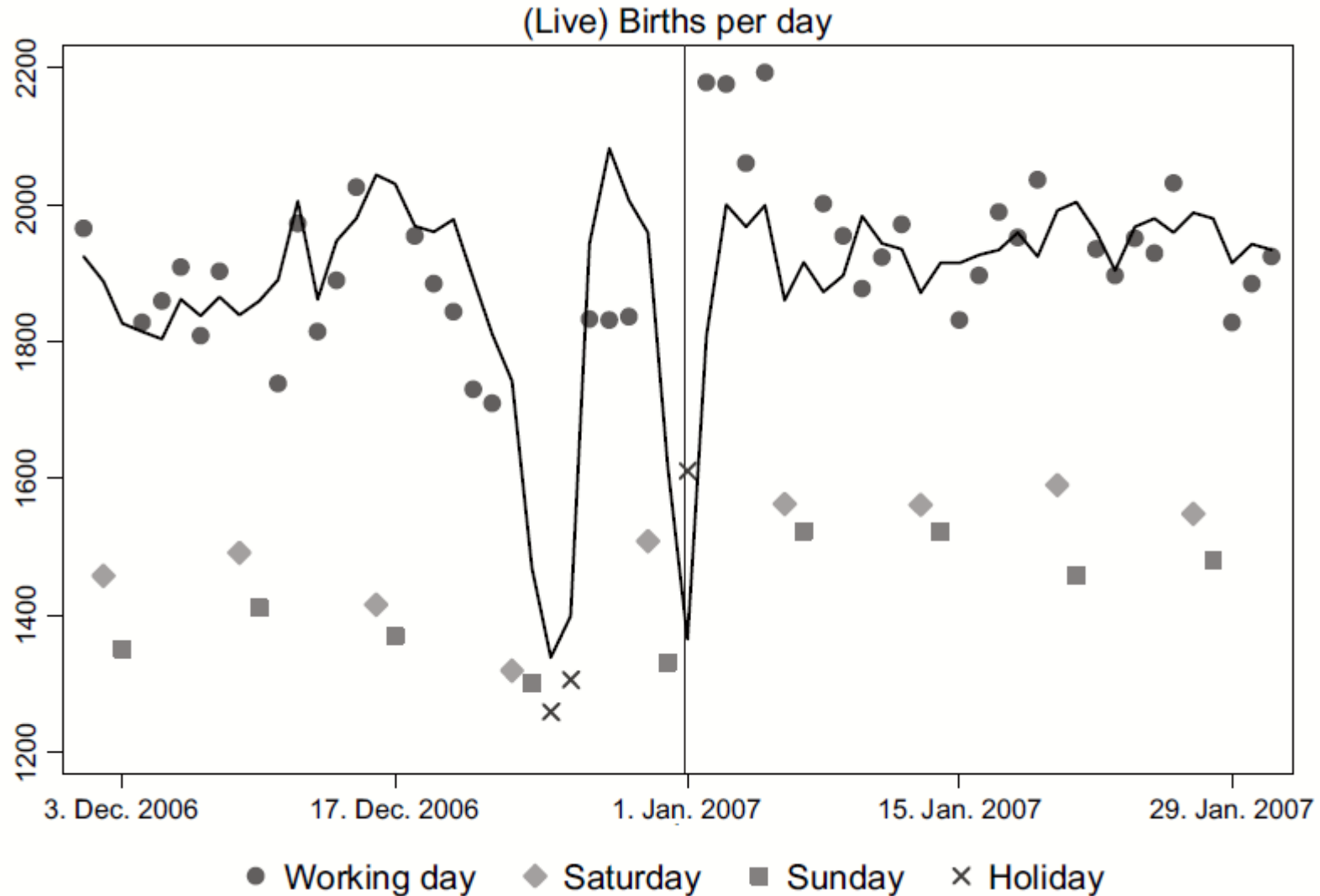
Die Crux steckt im Detail ...

- Eltern, die wussten, dass ihr Kind um den 1. Januar 2007 geboren werden soll, könnten auf den genauen Geburtstermin Einfluss nehmen
- Treten solche Verhaltensänderungen durch finanzielle Anreize tatsächlich auf?

Lebendgeburten im Jahresverlauf: 2001 – 2007



Lebendgeburten pro Tag: Dezember 2006 – Januar 2007

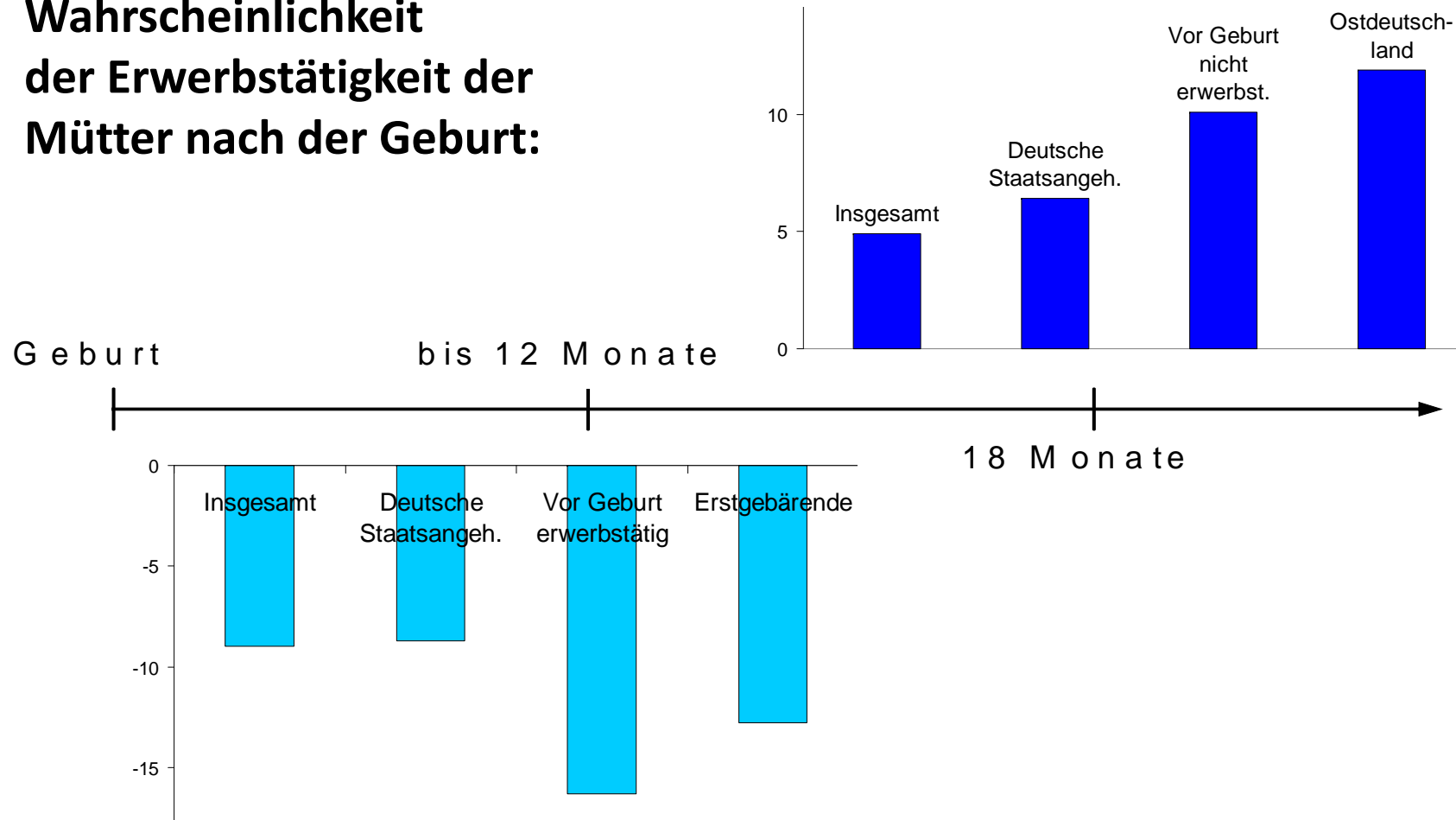


Die Crux steckt im Detail ...

- Diese „Verschiebungen“ kamen häufiger für erwerbstätige als für nicht erwerbstätige Mütter vor -> Selektion
- Empirische Lösung: Vergleich sowohl von Eltern Q1/07 mit Q4/06 als auch von Feb/März 2007 mit Okt/Nov 2006

Wie viel?

Wahrscheinlichkeit
der Erwerbstätigkeit der
Mütter nach der Geburt:



Weitere Ergebnisse

- Effekt auf Kinderbetreuung durch Väter nicht signifikant
- Wunsch der Eltern nach flexiblen Lösungen groß: Teilzeitbeschäftigung, Arbeitszeiten, KiTa-Betreuung

-> s. Kluge, J. und M. Tamm, „Parental leave regulations, mothers' labor force attachment, and fathers' childcare involvement: Evidence from a natural experiment“, *Journal of Population Economics*, im Erscheinen

Beispiel 3:

Differenz-in-Differenzen

—

Berufliche Bildung in Vietnam

Die Maßnahme

- „Promotion of TVET Vietnam“ -> bilaterale Kooperation zwischen Deutschland (GIZ) und Vietnam
- Maßnahme: Förderung von Berufsschulen
- Umfasst i) Lehrerfortbildung, ii) Lehrplanentwicklung und iii) technische Ausstattung
- Ist auf Schaffung nachhaltiger Strukturen angelegt

- Ziel: Passgenaue Ausbildung der Berufsschüler, d.h. Erhöhung der Beschäftigungswahrscheinlichkeit
- Geförderte Ausbildungsgänge: Metallarbeiten (Schweissen); Elektrotechnik / Elektronik; Textil / Kleidung

Methodischer Ansatz

- Evaluationspotenzial: Programm lief schon -> keine Baseline-Daten, keine anderen (administrativen) Daten; keine Vergleichsgruppe
- Ansatz zunächst: Vergleich des Arbeitsmarkterfolgs der geförderten und nicht-geförderten Ausbildungsgänge innerhalb der Schulen
- Vorteile: einzig mögliches Design; gleiche Schulen und lokaler Arbeitsmarkt
- Nachteile: unterschiedliche Arbeitsmarktchancen für unterschiedliche Ausbildungsgänge, Spill-over innerhalb der Schulen ?

Lösung: Eine besondere Zusammenarbeit

- Umsetzung einer Verbleibsuntersuchung („Tracer study“) in den Berufsschulen: Befragung der Absolventen zum Zeitpunkt des Abschlusses und 6 Monate danach
- Training der Schulen, diese Befragung selbst durchzuführen: Erhebung und Dateneingabe (+ einfache Analysen)
- Bereitstellung des Fragebogens
- Durch zahlreiche Workshops, Arbeitstreffen, etc.: Unterstützung durch das zuständige Ministerium und die Bildungsverwaltung
- Kostengünstig
- Schafft nachhaltige Strukturen, die nicht von der Förderung durch die GIZ abhängen

Erste Ergebnisse: Querschnitt 2009

Group	Supported Occupation	Unsupported Occupation	Difference
Treatment schools	0 .7817544	0 .6058182	0.1759362
Control schools	-	-	-

Aktuell: Differenz-in-Differenzen

- Daten: Absolventenjahrgänge 2010 und 2011, beide Befragungswellen
- 9 Schulen in der Teilnahmegruppe, 8 Schulen in der Kontrollgruppe
- $N = 8.999$ Absolventen
- Dieser Ansatz berücksichtigt Unterschiede zwischen Schulen und Ausbildungsgängen



Grün: Teilnahmeschulen,
Rot: Kontrollschulen

Differenz-in-Differenzen: “Roheffekt”

Group	Supported Occupation	Unsupported Occupation	DiD
Treatment schools	0.7714286	0.6009445	0.1704841
Control schools	0.7429399	0.7747385	-0.0317986
DiD	0.0284887	-0.173794	0.2022827*** (9.09)

Verteilung der Ausbildungsgänge

currVT_special_new	treatschool and treated occupation			
	0	1	0	1
Accounting	287		454	
Automobile technology	407		670	
Business Administration	41		80	
Computer repair and assembly	2		27	
Constructional Engineering	28		26	
Electrics		12		291
Industrial electrics		746		990
Industrial electronics		222		195
Informatics	118		369	
Metal cutting		226		372
Others	351	41	495	202
Refrigerating and air-conditioning	80		93	
Textile and garment		39		224
Welding		309		626
Total	1,314	1,595	2,214	2,900

Verteilung der Ausbildungsgänge, gewichtet

currVT_special_new	treatschool and treated occupation			
	0	1	0	1
Accounting	454		454	
Automobile technology	670		670	
Business Administration	80		80	
Computer repair and assembly	27		27	
Constructional Engineering	26		26	
Electrics		291		291
Industrial electrics		990		990
Industrial electronics		195		195
Informatics	369		369	
Metal cutting		372		372
Others	495	202	495	202
Refrigerating and air-conditioning	93		93	
Textile and garment		224		224
Welding		626		626
Total	2,214	2,900	2,214	2,900

Differenz-in-Differenz: gewichteter Effekt

Group	Supported Occupation	Unsupported Occupation	DiD
Treatment schools	0.7714286	0.6009445	0.1704841
Control schools	0.771192	0.764761	0.006431
DiD	0.0002366	-0.1638165	0.1640531*** (6.58)

Vielen Dank.

jochen.kluve@hu-berlin.de