



Kontrafaktische Methoden – Ein Beitrag zu mehr Qualität?

„Qualität von Evaluierung – methodische und inhaltliche
Anforderungen, Perspektiven, Instrumente“

Frühjahrsworkshop des AK Strukturpolitik der DeGEval
21. Bis 22. April 2016, SMUL, Dresden

INHALT

1. Einführung in kontrafaktische Methoden
2. Wichtigste Methoden und deren Voraussetzungen
 - 2.1 Randomisierung (Experimente)
 - 2.2 Matching
 - 2.3 Differenz-von-Differenzen (DvD)
 - 2.4 Regressions-Diskontinuitätsansatz (RDD)
 - 2.5 Instrumentenvariablenansatz (IV)
3. Herausforderungen bei der Suche nach einer Vergleichsgruppe
4. Fallbeispiele
 - 4.1 Evaluation des Bundesprogramms Kommunal-Kombi
 - 4.2 Evaluation der Modellprojekte „Bürgerarbeit“

1. EINFÜHRUNG IN KONTRAFAKTISCHE METHODEN

Ausgangssituation

- Problemstellung
- Konzeption eines Programms/einer Maßnahme (Intervention):
- Ziele der Intervention:

Zentrale Fragestellung:

Was wäre mit den Teilnehmenden geschehen, wenn sie **nicht** an der Intervention teilgenommen hätten?

- ➔ Diese kontrafaktische Situation ist in der Realität nicht zu beobachten
- ➔ Nachbildung der kontrafaktischen Situation mit kausalen Wirkungsanalysen

Fallbeispiel:

Integration Langzeitarbeitsloser in den Arbeitsmarkt

Programme öffentlich geförderter Beschäftigung (Bundesprogramm Kommunal-Kombi, Modellprojekte Bürgerarbeit)

Integration der Teilnehmenden in den ersten Arbeitsmarkt

Abgang aus dem Leistungsbezug

Erhöhung der Beschäftigungsfähigkeit

1. EINFÜHRUNG IN KONTRAFAKTISCHE METHODEN

- Deskriptiver Vergleich der Zielwerte vor und nach der Intervention
 - Berücksichtigt keine Veränderung in den Zielwerten unabhängig von der Intervention
 - Kausalität der gemessenen Veränderung nicht nachweisbar
- Einfacher Vergleich der Zielwerte einer Intervention zwischen Teilnehmenden und Nichtteilnehmenden
 - Keine Berücksichtigung des Zugangs in die Intervention
 - Je nach Zusammensetzung der Teilnehmendengruppe wird der Teilnahmeeffekt unter- oder überschätzt
- Berücksichtigung der Zusammensetzung der Teilnehmendengruppe bei der Auswahl einer Vergleichsgruppe
 - Vergleichsgruppe ist der Teilnehmendengruppe strukturell ähnlich
 - Identifizierte Effekte können kausal auf die Intervention zurückgeführt werden
 - ➔ Kommt kontrafaktischer Situation sehr nahe, baut allerdings auf sehr starken Annahmen auf

2. WICHTIGSTE METHODEN UND DEREN VORAUSSETZUNGEN

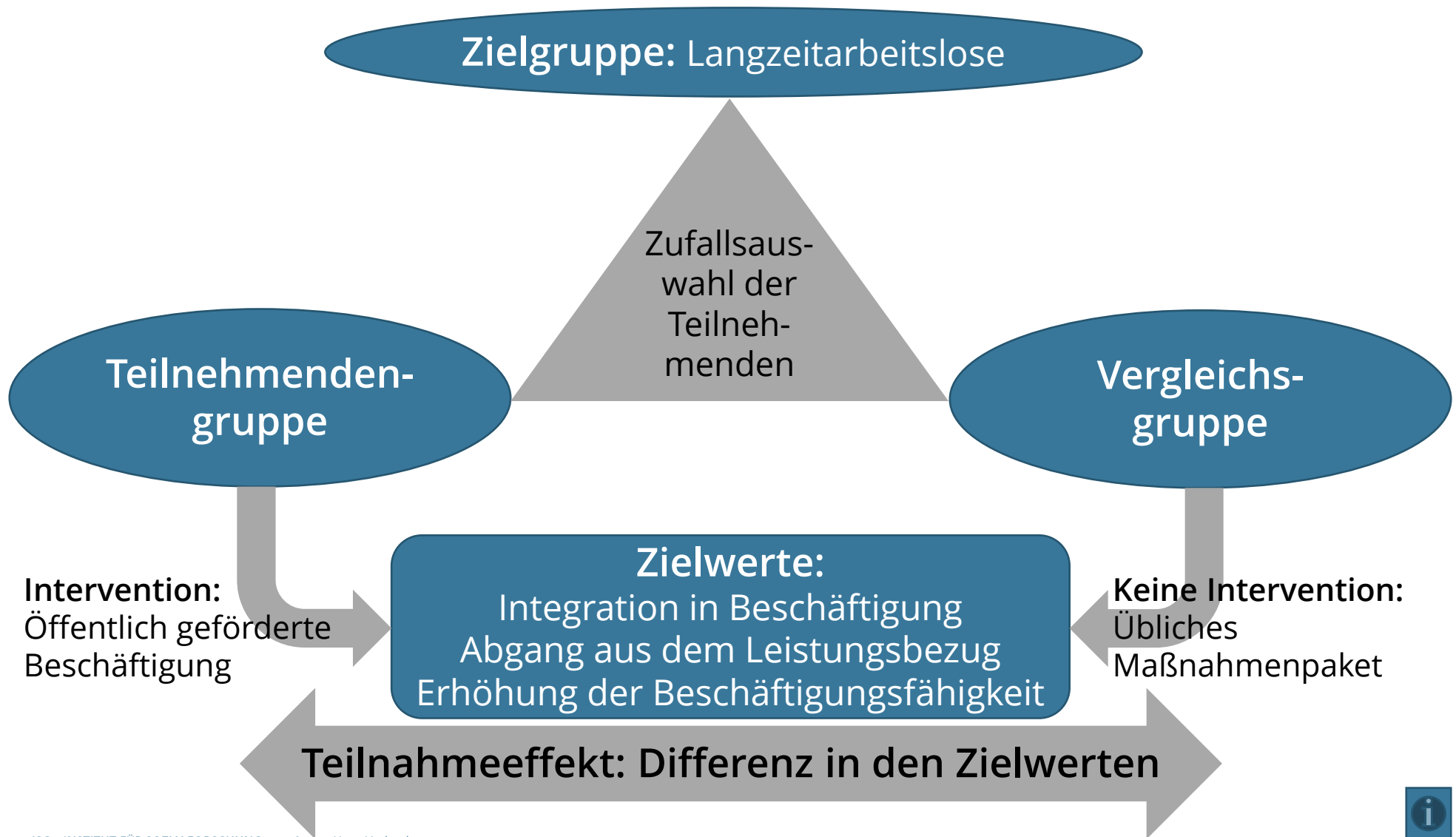
Experimentelle Methoden:

- Teilnehmendenauswahl Bestandteil des methodischen Ansatzes
- Die Evaluation...
 - ...beginnt zwingend vor dem Beginn der Intervention und
 - ...entscheidet über die Begünstigten der Intervention mit einem Zufallsverfahren.
- Aufgrund der Zufallsauswahl gleichen sich Teilnehmenden- und Vergleichsgruppe strukturell

Quasi-experimentelle Methoden:

- Teilnehmendenauswahl wird nachgebildet
- Die Methoden können auch noch nach dem Anlaufen der Intervention umgesetzt werden
- Auswahl der Vergleichsgruppe ist stärkeren Annahmen unterworfen, aber grundsätzlich praktikabel

2.1 RANDOMISIERUNG (EXPERIMENTE)



2.2 MATCHING

Zielgruppe: Langzeitarbeitslose

Teilnahmeauswahl kriteriengestützt:

- Bewerbungsverfahren
- Jobcenter

Identifikation statistischer Zwillinge:
Nachbildung der Auswahlkriterien der Teilnahmeauswahl

**Teilnehmenden-
gruppe**

**Vergleichs-
gruppe**

Intervention:
Öffentlich geförderte
Beschäftigung

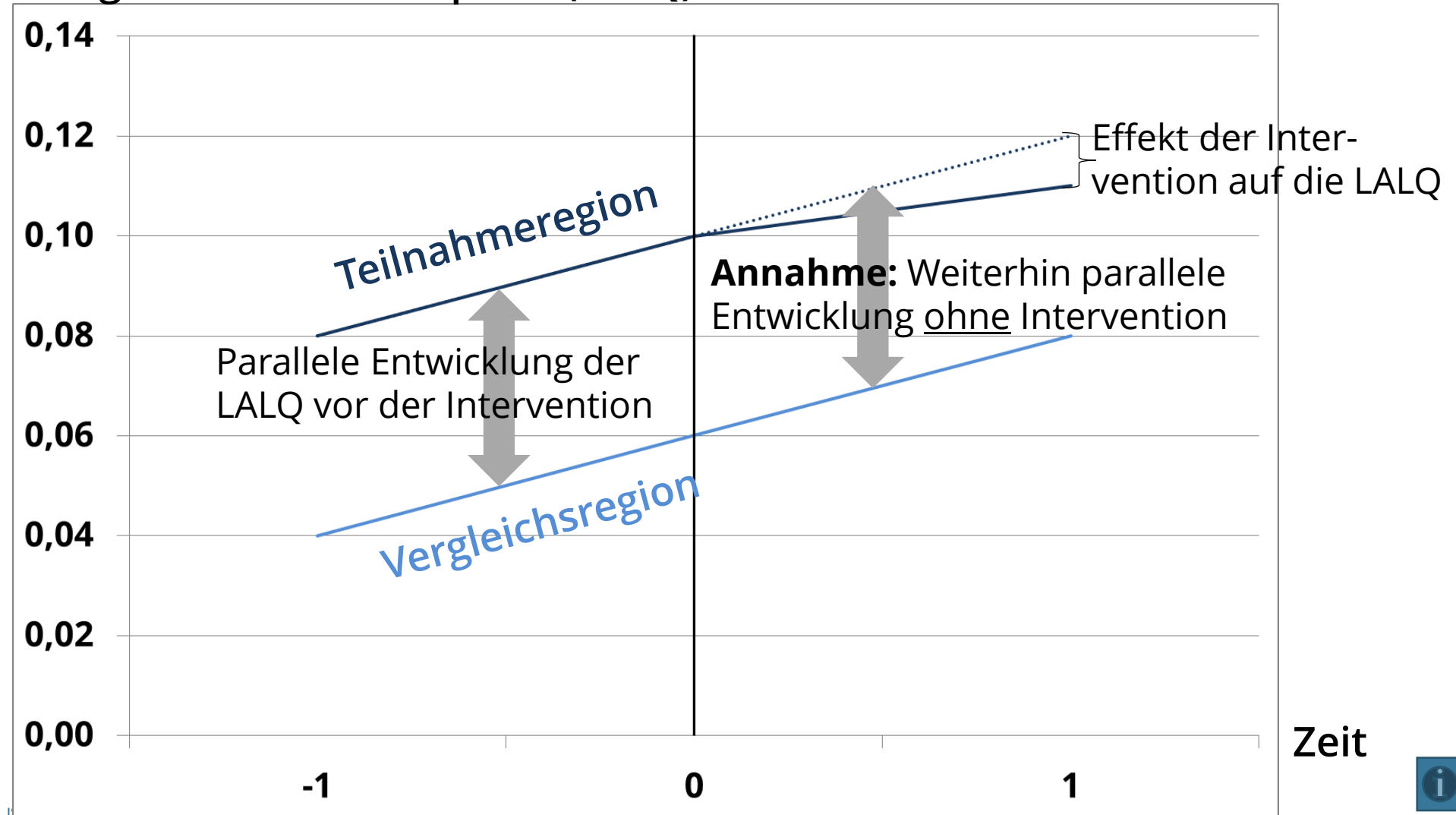
Zielwerte:
Integration in Beschäftigung
Abgang aus dem Leistungsbezug
Erhöhung der Beschäftigungsfähigkeit

Keine Intervention:
Übliches
Maßnahmenpaket

Teilnahmeeffekt: Differenz in den Zielwerten

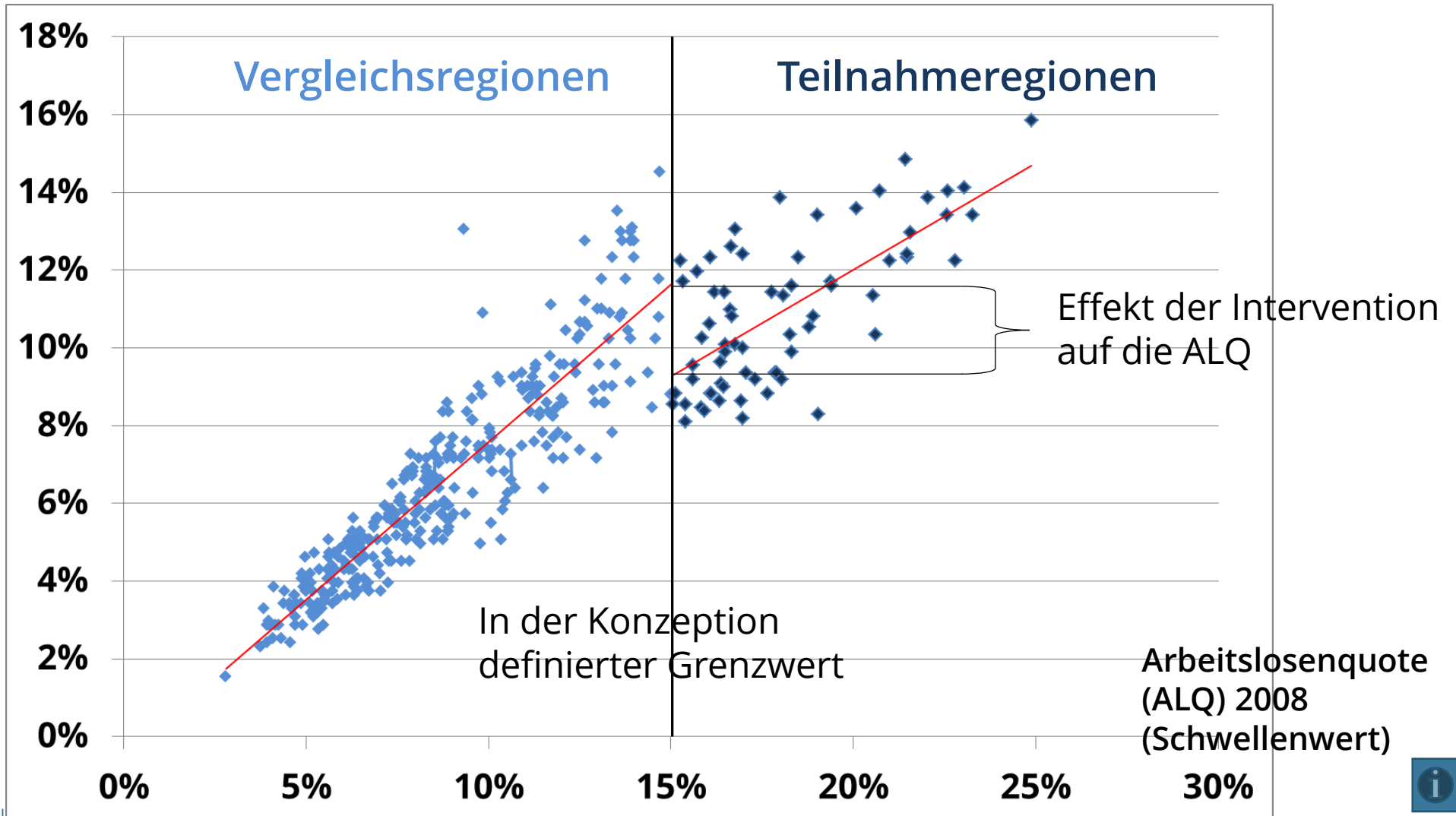
2.3 DIFFERENZ-VON-DIFFERENZEN (DVD)

Langzeitarbeitslosenquote (LALQ)

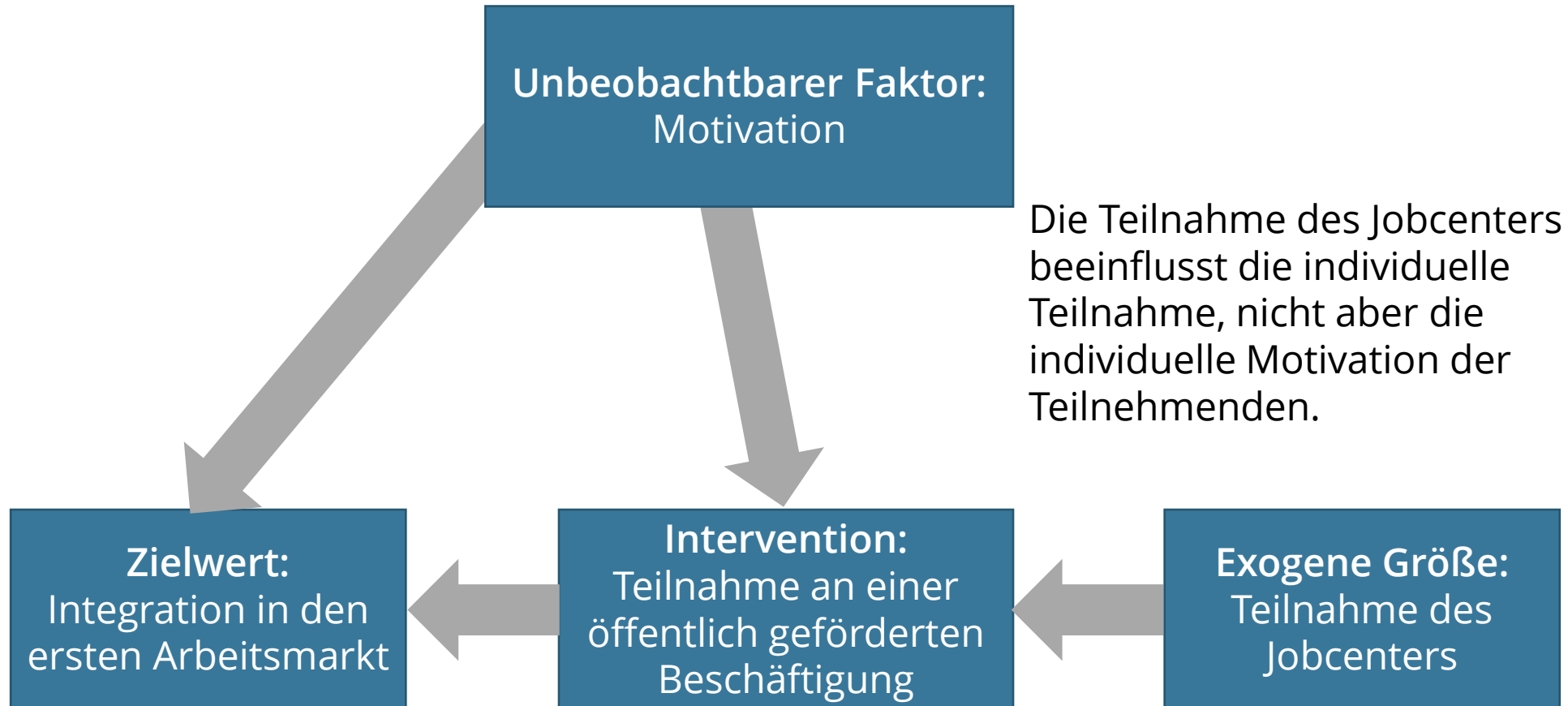


2.4 REGRESSIONS-DISKONTINUITÄTSANSATZ (RDD)

Arbeitslosenquote (ALQ) 2012 (Zielwert)



2.5 INSTRUMENTENVARIABLENANSATZ (IV)



3. HERAUSFORDERUNGEN BEI DER SUCHE NACH EINER VERGLEICHSGRUPPE

- 3 zentrale Fragen:
 - Ist die Teilnahme an der Intervention verpflichtend oder freiwillig?
 - Wurde die Intervention flächendeckend oder nur in bestimmten Regionen durchgeführt?
 - Wie ist der Zeitraum der Intervention definiert?

Nutzung der Variationen	Vergleichsgruppe	Nachteil
Regional	Zielgruppe aus nichtteilnehmenden Regionen	Unterschiedliche regionale Bedingungen
Zeitlich	Zielgruppe zu einem anderen Zeitpunkt vor oder nach der Intervention	Keine Berücksichtigung zyklischer Schwankungen
Teilnahmeberechtigung	Knapp nicht zur Teilnahme Berechtigte	Klare Regeln der Förderfähigkeit notwendig
Freiwilligkeit	Teilnahmeberechtigte Personen, die nicht teilgenommen haben	Abbildung des Entscheidungsprozesses notwendig

- Bei Zielgruppen mit vielen verschiedenen Interventionen kann die Maßnahmenteilnahme im Vergleich zu anderen Interventionen in der Zielgruppe gesetzt werden.

4. FALLBEISPIELE

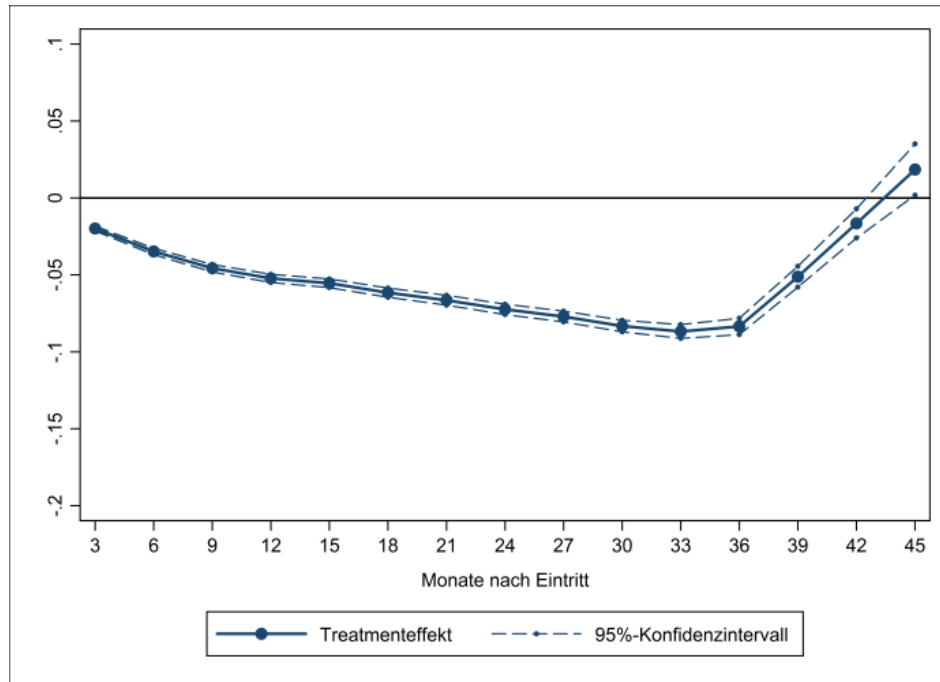
- Projekte des ISG mit kontrafaktischen Methoden im Rahmen einer ESF-Förderung:

Intervention	Förderzeitraum	Ergebnisse der Wirkungsanalyse
Bundesprogramm Kommunal-Kombi	2008 bis 2012	2013
Modellprojekte „Bürgerarbeit“	2010 bis 2014	2015
Evaluation des ESF-OP Bayern, Förderperiode 2014 bis 2020	2014 bis 2020	2024
Evaluation des ESF-OP Baden-Württemberg, Förderperiode 2014 bis 2020	2014 bis 2020	2025

4.1 EVALUATION DES BUNDESPROGRAMMS KOMMUNAL-KOMBI

Beschäftigungseffekte (Matching)

Übergänge aus dem ersten Arbeitsmarkt in Arbeitssuche (DvD)



Programm- gruppe	Vergleichs- gruppe	Zeitpunkte	Quartale
Frühe Förder- regionen insg. (n=63)	Nicht-Förder- regionen im Osten (n=14)	++ (3,1%)	++ (3,1%)
Frühe Förder- regionen im Osten (n=55)	Nicht-Förder- regionen im Osten (n=14)	+ (2,9%)	++ (3,1%)

+++ / ---: Zu 1% signifikant positiv/negativ

++ / --: zu 5% signifikant positiv/negativ

+ / -: Zu 10% signifikant positiv/negativ

4.1 EVALUATION DES BUNDESPROGRAMMS KOMMUNAL-KOMBI

Wirtschaftlichkeitsanalysen

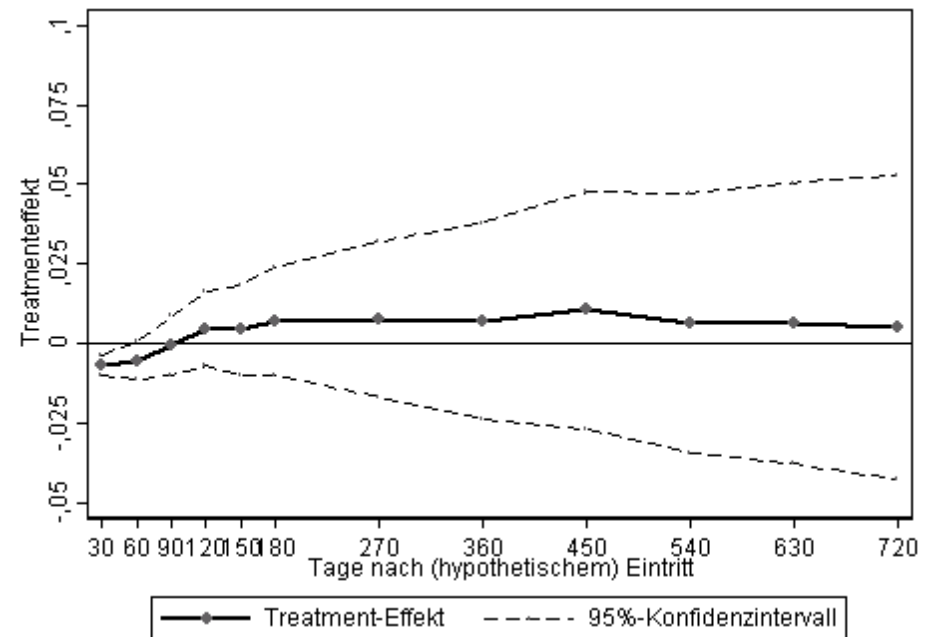
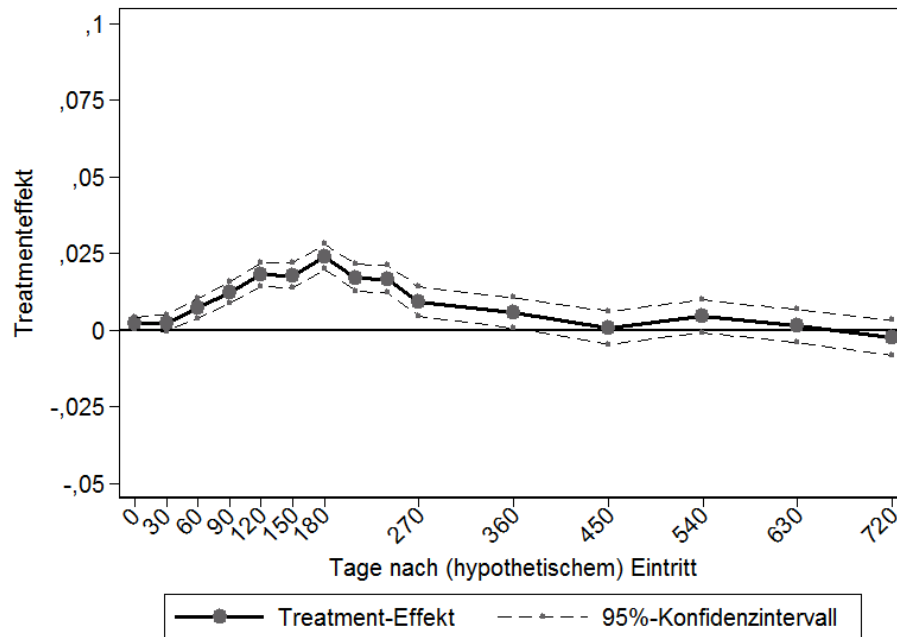
	Insg. (Mio. Euro)	Pro Stelle	Break-Even (in Jahren)
Direkte Kosten	-149,5	-9.445	
Nur Beschäftigungseffekte	-139,2	-8.796	42,2
Beschäftigungseffekte & Substitution und Verdrängung von 25 %	-119,2	-7.534	36,1
Beschäftigungseffekte & Substitution und Verdrängung von 50 %	-239,9	-15.158	72,7

Anmerkung: Der Break-Even berechnet sich auf Basis der Einnahmen durch einen konstanten Beschäftigungseffekt von zwei Prozentpunkten nach 45 Monaten.

4.2 EVALUATION DER MODELLPROJEKTE „BÜRGERARBEIT“

Beschäftigungseffekte der Aktivierungsphase (Matching)

Beschäftigungseffekte der Aktivierungsphase (IV)



4.2 EVALUATION DER MODELLPROJEKTE „BÜRGERARBEIT“

Wirtschaftlichkeitsanalysen Aktivierungsphase

	Insg. (Mio. Euro)	Bemerkungen
ESF-Mittel	0,0	Aktivierungsphase nicht ESF-gefördert
Bundesmittel	-127,3	Keine direkten Bundesmittel für Aktivierungsphase: Maßnahmekosten abzüglich der Nettoeinsparungen im Sozialleistungs- und Nettoeinnahmen im Steuersystem durch Integration in den Arbeitsmarkt
Sozialversicherungsträger	6,0	Nettoeinsparungen im Sozialleistungssystem durch Integration in den Arbeitsmarkt
Bundesländer	5,9	Nettoeinnahmen im Steuersystem durch Integration in den Arbeitsmarkt
Kommunen	2,6	Nettoeinsparungen im Sozialleistungs- und Nettoeinnahmen im Steuersystem durch Integration in den Arbeitsmarkt
Gesamt	-112,7	

LITERATUREMPFEHLUNG

EU Kommission (2012): Entwicklung und Beauftragung von kontrafaktischen Wirkungsanalysen – Ein praktischer Leitfaden für ESF-Verwaltungsbehörden (<http://t1p.de/EU-KOM2012>)

IAW und ISG (2013): Programmbegleitende und abschließende Evaluation des Kommunal-Kombi, Endbericht zum 30. Juni 2013 Tübingen/Köln/Berlin. (<http://t1p.de/KoKo>)

IAW und ISG (2015): Evaluation der Modellprojekte „Bürgerarbeit“, Endbericht zum 25. November 2015 Tübingen/Köln/Berlin. (<http://t1p.de/Buergerarbeit>)

Vielen Dank für die Aufmerksamkeit

Kontakt:

ISG Institut für Sozialforschung und Gesellschaftspolitik GmbH

Hans Verbeek

Gorgasring 2

13599 Berlin

Tel.: +49 (0)30 650 43 90

E-Mail: verbeek@isg-institut.de

Anhang

2.1 RANDOMISIERUNG (EXPERIMENTE)

- Teilnahmeauswahl aus der Zielgruppe erfolgt per Zufallsauswahl
- Zentrale Anforderung: Kontrolle der Auswahl für die Intervention per Zufallsverfahren

Vorteile	Nachteile
<ul style="list-style-type: none">• Unverzerrte Schätzungen der Effekte• Transparente und leicht verständliche Ergebnisse• Berücksichtigung unterschiedlicher politischer Kontexte und Interventionsumstände	<ul style="list-style-type: none">• Ausschluss der Kontrollgruppe von der Intervention• Einfluss der Randomisierung auf die Zusammensetzung der Untersuchungsgruppe• Zuweisungsstatus beeinflusst Verhalten der Teilnehmenden• Ethische Bedenken• Hohe Anforderungen an Planung und Design.• Kostenintensiv

2.2 MATCHING

- Bildung einer Vergleichsgruppe über die Zuordnung statistischer Zwillinge anhand beobachtbarer Merkmale
- Datenanforderungen:
 - Präzise Abgrenzung der Teilnehmendengruppe in den Daten erforderlich
 - Existenz einer Datenquelle mit Teilnehmenden- und Nichtteilnehmendeninformationen
- Detaillierte Kenntnis über Konzept der Teilnehmendenauswahl notwendig

Vorteile	Nachteile
<ul style="list-style-type: none">• Keine unmittelbare Kontrolle des Auswahlprozesses notwendig• Kann rückwirkend angewendet werden• Semi-parametrisches Schätzverfahren: erfordert weniger parametrische Annahmen• Verwendung bei mehrfachen Teilnahmen möglich	<ul style="list-style-type: none">• Datenhungriges Verfahren• Starke Abhängigkeit von der Qualität der Vergleichsgruppe sowie der gemeinsamen Stützmenge• Annahme: Teilnehmendenauswahl beruht ausschließlich auf beobachtbare Merkmalen

2.3 DIFFERENZ-VON-DIFFERENZEN (DVD)

- Vergleich der Entwicklungen der Zielwerte in der Teilnehmendengruppe mit der Vergleichsgruppe
- Annahme gemeinsamer Trends: Entwicklung der Teilnehmenden- und der Vergleichsgruppe würde ohne Intervention parallel verlaufen
- Datenanforderungen:
 - Mehrere Datenpunkte vor der Intervention notwendig um eine Plausibilisierung der Gültigkeit der Annahme gemeinsamer Trends durchzuführen

Vorteile	Nachteile
<ul style="list-style-type: none">• Kontrolle für unbeobachtbare zeitkonstante Merkmale• Kombination mit Matching möglich• Verfügbare Methoden für wiederholte Querschnitts- oder Paneldaten	<ul style="list-style-type: none">• Annahme gemeinsamer Trends kann nicht überprüft werden• Komplexe Analysen mit der Gefahr von Fehlinterpretationen• Nicht geeignet zur Schätzung der Wirkung mehrfacher Teilnahmen

2.4 REGRESSIONS-DISKONTINUITÄTSANSATZ (RDD)

- Teilnahmevoraussetzung abhängig von der Über- oder Unterschreitung eines Grenz-/Schwellenwertes
- Vergleich der Beobachtungen knapp oberhalb und knapp unterhalb des Grenz-/Schwellenwertes
- Zentrale Anforderung: Grenz-/Schwellenwert unabhängig von der Einstufung der Teilnehmenden
- Datenbedarf:
 - Ergebnisse der Einstufung der Zielgruppe
 - Grenz-/Schwellenwert

Vorteile	Nachteile
<ul style="list-style-type: none"> • Sowohl scharfe als auch unscharfe RDD-Ansätze verfügbar • Liefert unter bestimmten Bedingungen unverzerrte Wirkungen einer Intervention 	<ul style="list-style-type: none"> • Kontinuierliche Messung oder Einstufung, die über die Behandlung entscheidet, notwendig • Die Analysen können komplex und unsicher werden: <ul style="list-style-type: none"> • Probleme mit der funktionalen Form der Wirkungsregression • Nichtübereinstimmung • Zu kleine Stichprobengröße • Interpretation und Verallgemeinerung der Ergebnisse schwierig

2.5 INSTRUMENTENVARIABLENANSATZ (IV)

- Isolation exogener Veränderungen des Ablaufs einer Intervention
- Zentrale Anforderungen:
 - Einfluss des Instruments auf den Ablauf der Intervention, aber kein direkter Einfluss auf die Zielwerte
 - Keine Korrelation des Instruments mit anderen Faktoren, die Einfluss auf die Zielwerte haben
- Datenanforderungen:
 - Ausgangsdaten
 - Daten zum Ablauf der Intervention

Vorteile	Nachteile
<ul style="list-style-type: none"> • Qualitativ hochwertige Schätzungen • Löst das Problem der Verzerrung durch nicht berücksichtigte Variablen (bzw. der Auswahlverzerrung) • Kann rückwirkend angewendet werden • Erlaubt die Schätzung bestimmter Wirkungstypen 	<ul style="list-style-type: none"> • Identifikation eines plausiblen Instruments schwer • Schwer vermittelbar • Komplizierte Interpretation der Ergebnisse • Zugrundeliegende Annahmen nur begrenzt überprüfbar