



AUS DER FORSCHUNG

Gestaffelte Difference-in-Differences braucht ein Kohorten-Zeit-Estimand

Wenn eine Behandlung zu verschiedenen Zeitpunkten beginnt, müssen Kohorten-Zeit-Effekt und Aggregationsziel vor dem Koeffizienten stehen.

Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand

VERÖFFENTLICHT

1. September 2026

LESEZEIT

8 Min.

WÖRTER

1.587

<https://isoglu.com/de/insights/journal/gestaffelte-difference-in-differences-braucht-ein-kohortENZEIT-estimand/>

Management Summary	2
Warum muss die Kausalanalyse das Kohorten-Zeit-Estimand vor der Schätzung benennen?	3
Warum sind Gruppen-Zeit-Effekte der unverzichtbare Startpunkt gestaffelter Rollouts?	5
Warum erzeugen klassische Two-Way-Fixed-Effects-Modelle negative Gewichte?	5
Warum sind parallele Vortrends ein wichtiges Indiz, aber kein Freibrief für Kausalität?	6
Wie beeinflussen nichtlineare Zielgrößen gestaffelte Difference-in-Differences-Schätzungen?	6
Welcher ökonometrische Prüfablauf sichert unverzernte Rollout-Analysen ab?	6
Wo liegen die methodischen Grenzen moderner Difference-in-Differences-Schätzer?	7
Quellen	8

MANAGEMENT SUMMARY

Gestaffelte Difference-in-Differences ist nicht einfach ein vertrauter Koeffizient mit neuer Bezeichnung. Wenn Kohorten die Behandlung zu verschiedenen Zeitpunkten erhalten, müssen Gruppe, Kalenderzeit, Ereigniszeit, Vergleichsgruppe, Ergebnis, Zeithorizont und Aggregationsziel benannt werden. Callaway und Sant'Anna definieren Gruppen-Zeit-Effekte und Aggregationsschemata. Goodman-Bacon zeigt, wie ein Two-Way-Fixed-Effects-Koeffizient gewichtete Zwei-mal-zwei-Vergleiche verbinden kann, die bei heterogenen Effekten schwer zu lesen sind. Roth und Kollegen trennen Estimand, Vortrendtests, Sensitivität und Inferenz. Wooldridge liefert eine begrenzte nichtlineare Erweiterung. Der Beitrag baut daraus eine Kohorten-Zeit-Freigabekarte. Er schätzt keinen aktuellen Effekt und bewertet keine Software.

Schlagwörter: Difference-in-Differences · Kausalinferenz · Heterogenität von Behandlungseffekten · Kohorten-Zeit-Estimand

Wenn eine Behandlung zu verschiedenen Zeitpunkten beginnt, ist die erste Aufgabe nicht die Auswahl einer vertrauten Regression. Die erste Aufgabe ist die Benennung des Effekts.

Die kurze Antwort lautet: Ein Design mit gestaffelten Behandlungszeitpunkten braucht ein Kohorten-Zeit- Estimand, eine explizite Vergleichsgruppe und eine erklärte Aggregationsregel, bevor ein Koeffizient eine kausale Bedeutung erhalten kann. Ein Two-Way-Fixed-Effects-Koeffizient kann in manchen Situationen brauchbar sein. Bei gestaffeltem Timing kann er jedoch mehrere Zwei-mal-zwei-Vergleiche mit Gewichten verbinden, die nicht zu dem Effekt passen, den eine Entscheidungsträgerin oder ein Entscheidungsträger zu lesen glaubt.

Callaway und Sant’Anna definieren Gruppen-Zeit-Durchschnittseffekte für Situationen mit mehreren Perioden und unterschiedlichen Behandlungszeitpunkten. Goodman-Bacon zerlegt den herkömmlichen Two-Way-Fixed-Effects- Schätzer in gewichtete Zwei-mal-zwei-Vergleiche. Roth, Sant’Anna, Bilinski und Poe ordnen die neuere Literatur um Heterogenität, Paralleltrends und Inferenz. Wooldridge entwickelt eine nichtlineare Erweiterung für einen engeren Paneldatenrahmen. Zusammen begründen die Quellen eine Entwurfsdisziplin, aber keinen universellen Schätzer.

Warum muss die Kausalanalyse das Kohorten-Zeit-Estimand vor der Schätzung benennen?

Nehmen wir an, zwei Kundenkohorten erhalten dieselbe kommerzielle Intervention. Kohorte A erhält sie im Januar. Kohorte B erhält sie im April. Eine dritte Kohorte hat sie bis zum Ende des Beobachtungsfensters noch nicht erhalten. Die Formulierung „der Behandlungseffekt“ ist unvollständig. Welche Kohorte? Welcher Monat? Verglichen mit welchen Einheiten? Wie lange nach dem Start? Mit welchen Gewichten?

Die Mindestbeschreibung des Objekts lautet:

Tabelle 1 Warum muss die Kausalanalyse das Kohorten-Zeit-Estimand vor der Schätzung benennen?

Feld	Frage	Warum es zählt
Kohorte	Welche Einheiten erhalten in dieser Gruppe zuerst die Behandlung?	Der Behandlungszeitpunkt gehört zum Objekt.
Kalenderzeit	In welcher Periode wird das Ergebnis beobachtet?	Eine Kohorte kann sich über die Zeit unterschiedlich entwickeln.
Ereigniszeit	Wie weit liegt die Periode vor oder nach der Behandlung?	Dynamische Effekte sind nicht dasselbe wie ein einziger Nachbehandlungsdurchschnitt.
Vergleich	Welche unbehandelten oder noch nicht behandelten Einheiten liefern den Kontrast?	Bereits behandelte Einheiten sind nicht automatisch gültige Kontrollen.
Ergebnis	Was wird in welcher Einheit und über welchen Horizont gemessen?	Ein Koeffizient repariert kein unklar definiertes Ergebnis.
Aggregation	Welche Kohorten-Zeit-Effekte werden mit welchen Gewichten verbunden?	Die Gesamtaussage hängt vom Aggregationsziel ab.
Inferenz	Auf welcher Ebene wurde die Behandlung zugewiesen? Wo wird die Unsicherheit gebündelt?	Präzision ist Teil des Designs.

Tabelle aus diesem Essay. Quellen und Einordnung stehen im Beitrag.

Abbildung 1 Die Freigabekarte für gestaffelte Difference-in-Differences

Freigabefeld	Erforderliche Eingabe	Zulässiger Satz	Abbruchsignal
Kohorte und Zeit	Erstbehandlungsgruppe, Kalenderperiode, Ereigniszeit	„Dies ist der Effekt für Kohorte g zur Zeit t .“	Der Beitrag sagt nur „der Behandlungseffekt“.
Vergleich	Nie behandelte oder noch nicht behandelte Gruppe mit Bedingungen	„Dieser Vergleich liefert den benannten Kontrast.“	Bereits behandelte Einheiten dienen still als Kontrolle.
Ergebnis und Horizont	Ergebniseinheit, Messfenster, Nachbehandlungshorizont	„Das Ergebnis betrifft diese Einheit über diesen Horizont.“	Das Ergebnis wechselt zwischen Abschnitten.
Aggregation	Kohorten, Perioden und Gewichte des Gesamtwerts	„Dieses Ergebnis beantwortet diese gewichtete Frage.“	Die Zusammenfassung wird als natürlicher ATT gelesen.
Inferenz und Sensitivität	Zuweisungsebene, Unsicherheitsmethode, Trend- und Heterogenitätstests	„Unsicherheit und Sensitivität wurden an dieser Grenze geprüft.“	Ein Vortrendtest gilt als Beweis.

Freigaberahmen des Autors, begründet mit Callaway und Sant'Anna (2021), Goodman-Bacon (2021), Roth et al. (2023) und Wooldridge (2023). Die Fragen sind synthetisch und enthalten keine Effektschätzung.

Warum sind Gruppen-Zeit-Effekte der unverzichtbare Startpunkt gestaffelter Rollouts?

Callaway und Sant'Anna definieren Gruppen-Zeit-Durchschnittseffekte in einer Situation mit mehreren Perioden und unterschiedlichen Behandlungszeitpunkten. Die Gruppe zeigt, wann Einheiten erstmals behandelt werden. Die Zeit zeigt, in welcher Periode das Ergebnis geprüft wird. Der Effekt ist damit eine Menge von Objekten und nicht zwangsläufig eine natürlich gegebene einzelne Zahl.

Das macht die Entscheidungsfrage klarer. Ein Team kann den Effekt für frühe Anwenderinnen und Anwender im ersten Monat nach der Behandlung wissen wollen. Es kann den Effekt für späte Anwender nach einem vollen Quartal benötigen. Es kann einen Gesamtwert über Kohorten verlangen. Diese Fragen sind nicht austauschbar.

Auch die Identifikationsbedingungen gehören zum Objekt. Callaway und Sant'Anna erlauben in der relevanten Formulierung Paralleltrends nach Konditionierung auf beobachtete Kovariaten. Ihre Verfahren umfassen Outcome-Regression, Inverswahrscheinlichkeitsgewichtung und doppelt robuste Verfahren. Die Methode ersetzt nicht die Erklärung, was über unbehandelte potenzielle Ergebnisse angenommen wird.

Ebenso wichtig sind die Aggregationsschemata. Wenn Effekte nach Kohorte oder Ereigniszeit variieren, hängt eine Zusammenfassung davon ab, welche Gruppen und Perioden Gewicht erhalten. Ein Gesamtwert kann nützlich sein, wenn er eine erklärte politische oder geschäftliche Frage beantwortet. Er ist nicht automatisch der Durchschnitt, den Lesende bei einem einzelnen Behandlungskoeffizienten vermuten.

Warum erzeugen klassische Two-Way-Fixed-Effects-Modelle negative Gewichte?

Goodman-Bacon zeigt, dass der Two-Way-Fixed-Effects-Schätzer bei unterschiedlichen Behandlungszeitpunkten als gewichteter Durchschnitt von Zwei-mal-zwei-Vergleichen dargestellt werden kann. Dazu gehören Behandelte gegen nie behandelte Einheiten, früh behandelte gegen später behandelte Einheiten und später behandelte gegen früh behandelte Einheiten.

Das Problem ist nicht, dass jedes Two-Way-Fixed-Effects-Design mechanisch unbrauchbar wäre. Das Problem ist, dass Gewichte und Vergleichsobjekte des Koeffizienten vom Ziel-Effekt abweichen können. Eine früh behandelte Gruppe kann als Kontrolle für eine später behandelte Gruppe dienen. Sobald die frühe Gruppe selbst betroffen ist, ist dieser Vergleich kein einfacher unbehandelter Kontrast mehr.

Bei homogenen Effekten kann die Mischung dennoch eine klare Bedeutung haben. Bei heterogenen Effekten kann sie schwer interpretierbar werden und Gewichte erhalten, die negativ oder anderweitig problematisch sind. Die richtige Reaktion ist, Timing und Heterogenität sichtbar zu machen und dann ein Estimand und einen Schätzer für die eigentliche Frage zu wählen. „Two-Way Fixed Effects ist immer falsch“ ist ebenso ungenau wie „der Koeffizient ist der ATT“.

Deshalb sollte eine Regressionstabelle nicht das erste Exponat sein. Das erste Exponat sollte zeigen, wer behandelt wird, wann die Behandlung beginnt, wer als Vergleich zur Verfügung steht und welche Effekte aggregiert werden.

Warum sind parallele Vortrends ein wichtiges Indiz, aber kein Freibrief für Kausalität?

Roth und Kollegen ordnen Vortrendprüfungen in eine umfassendere Methodenagenda ein. Ihre Übersicht betont mehrere Perioden und gestaffeltes Timing, Verletzungen von Paralleltrends, Heterogenität und Inferenz. Sie warnt außerdem, dass Vortrendtests zu wenig Power haben können und dass ihre Verwendung als Vorprüfung Bias einführen kann.

Ein nicht bestandener Vortrendtest kann ein ernstes Problem anzeigen. Ein bestandener oder nicht auffälliger Test beweist Paralleltrends nicht. Ein Test kann eine relevante Abweichung nicht entdecken, oder die ausgewählte Vorperiode bildet den unbehandelten Verlauf nach der Behandlung nicht ab. Der Freigabesatz sollte daher Test, Zeitraum, Unsicherheit und Sensitivität gegenüber plausiblen Trendunterschieden nennen.

Dasselbe gilt für die Inferenz. Roth und Kollegen empfehlen explizite Entscheidungen zu Estimand und Schätzer, heterogenitätsrobuste Verfahren, Sensitivitätsanalyse und, wenn angemessen, eine Bündelung der Inferenz auf der Behandlungsebene. Clustering rettet keine Vergleichsgruppe, die bereits kontaminiert ist. Es beantwortet eine andere Unsicherheitsfrage.

Wie beeinflussen nichtlineare Zielgrößen gestaffelte Difference-in-Differences-Schätzungen?

Viele kommerzielle Ergebnisse sind nicht kontinuierlich. Es kann sich um Adoption, Bestellzahlen, begrenzte Raten oder positive, aber schiefe Größen handeln. Wooldridge entwickelt einfache nichtlineare Difference-in-Differences-Strategien für Paneldaten mit gestaffelten Interventionen und optionalen Kovariaten.

Die Erweiterung ist nützlich, weil sie die Ergebnisskala sichtbar macht. Sie ist keine Erlaubnis, jede gestaffelte Analyse mit einem nichtlinearen Link zu versehen und denselben Satz zu verwenden. Estimand, Skala, Kovariaten, Timing und Inferenz bleiben modellspezifisch. Eine Wahrscheinlichkeitsdifferenz, ein Effekt auf einer Log-Skala und ein Verhältnis sind verschiedene Objekte, auch wenn sie dasselbe Panel nutzen.

Die praktische Regel lautet: Wenn sich die Ergebnisskala ändert, wird die Freigabekarte neu geschrieben. „Nichtlinear“ darf keine Fußnote sein, die den Kausalsatz unverändert lässt.

Welcher ökonometrische Prüfablauf sichert unverzernte Rollout-Analysen ab?

Nutzen Sie vor einer geschäftlichen Entscheidung die folgende Reihenfolge:

- 1 Schreiben Sie die Behandlungsdefinition und das Datum der ersten Behandlung für jede Kohorte auf.
- 2 Benennen Sie Ergebnis, Einheit, Beobachtungsfenster und Nachbehandlungshorizont.
- 3 Definieren Sie die Vergleichsgruppe jeder Kohorten-Zeit-Zelle und schließen Sie bereits behandelte Einheiten begründet aus.
- 4 Erklären Sie die Bedingung für den unbehandelten Trend, einschließlich der Kovariaten.
- 5 Wählen Sie die Schätzerfamilie nach dem Estimand, nicht davor.
- 6 Wählen Sie Aggregationsgewichte und erklären Sie, warum sie die Entscheidungsfrage beantworten.

- 7 Nennen Sie Inferenzcluster und prüfen Sie Vortrends, Heterogenität und Sensitivität getrennt.
- 8 Wenn das Ergebnis nichtlinear ist, erklären Sie Skala und Interpretation des Modells.
- 9 Schreiben Sie den zulässigen Satz und den stärkeren Satz, den die Evidenz nicht freigibt.

Diese Reihenfolge ist langsamer als das Kopieren einer Regressionsspezifikation. Sie ist schneller, als einen unerklärten Koeffizienten zu diskutieren, nachdem die Entscheidung bereits gefallen ist.

Wo liegen die methodischen Grenzen moderner Difference-in-Differences-Schätzer?

Die vier Quellen zeigen keinen aktuellen Behandlungseffekt für eine private Pipeline, eine Kundenkohorte oder einen Arbeitgeber. Sie bewerten nicht jede Software. Sie sagen nicht, dass ein Schätzer unter allen Timing-, Ergebnis- und Heterogenitätsbedingungen überlegen ist. Sie liefern die Bestandteile für einen Estimand-zuerst-Entwurf und eine Warnung vor automatischer Interpretation.

Die Abbruchregel ist daher konkret. Geben Sie ein Ergebnis aus gestaffelter Difference-in-Differences erst frei, wenn Ziel-Kohorte und -Zeit, Vergleichsgruppe, Aggregationsregel, Inferenzentscheidung und Sensitivitätsgrenze am selben Ort stehen. Fehlt eine davon, kann das Ergebnis berechenbar sein. Es ist noch keine belastbare Antwort auf eine Entscheidungsfrage.

Die Estimand-Grenze verbindet sich mit dem Attributionsmodell mit Kontrafaktum und der Uplift-Aussage, die eine Spezifikation braucht.

- Callaway, B., und P. H. C. Sant'Anna. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2), 200-230. DOI
- Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2), 254-277. DOI
- Roth, J., P. H. C. Sant'Anna, A. Bilinski und J. Poe. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2), 2218-2244. DOI
- Wooldridge, J. M. (2023). Simple Approaches to Nonlinear Difference-in-Differences with Panel Data. *The Econometrics Journal*, 26, C31-C66. DOI

ZITIERWEISE

Isoglu, S. (2026, 1. September). Gestaffelte Difference-in-Differences braucht ein Kohorten-Zeit-Estimand. Sinan Isoglu.
<https://isoglu.com/de/insights/journal/gestaffelte-difference-in-differences-braucht-ein-kohortenzeit-estimand/>

WEITERLESEN

Aus der Forschung

Was sind parallele Trends? Die Annahme hinter Difference-in-Differences

<https://isoglu.com/de/insights/journal/was-sind-parallele-trends/>

Aus der Forschung

Was ist Inkrementalität?

<https://isoglu.com/de/insights/journal/was-ist-inkrementalitaet/>

Aus der Forschung

Was ist Interferenz? Wenn die Behandlung einer Einheit ein anderes Ergebnis verändert

<https://isoglu.com/de/insights/journal/was-ist-interferenz/>



Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand