

# Politikevaluation profitiert von Datenflut

Die zunehmende Verfügbarkeit von digitalen Daten bringt neue methodische Möglichkeiten wie das maschinelle Lernen. Dies hilft, die Wirksamkeit von wirtschaftspolitischen Massnahmen zu messen. *Martin Huber*

**Abstract** Die datenbasierte Politikevaluation zur Messung des kausalen Effekts einer Massnahme (zum Beispiel eines Qualifizierungsprogramms für Arbeitssuchende) auf ein interessierendes Ergebnis (zum Beispiel die Wiederbeschäftigung) hat in den letzten Jahrzehnten bedeutende Innovationen erfahren. Dies betrifft sowohl die Neu- und Weiterentwicklung von statistischen Verfahren zur Messung von Kausalität als auch die Verfügbarkeit und Verwendung von informativeren und umfangreicheren Daten. Der Beitrag gibt einen methodischen Überblick zu gängigen Ansätzen der Politikevaluation, illustriert anhand praktischer Beispiele auch aus der Schweiz. Ferner wird auf neue Methoden im Kontext der exponentiell wachsenden Datenverfügbarkeit eingegangen, die es erlauben, die Politikevaluation mit dem sogenannten maschinellen Lernen, einem Teilbereich der künstlichen Intelligenz, zu kombinieren.

**W**ie wirken sich Weiterbildungen auf die Wiederbeschäftigung von Arbeitslosen aus? Welchen Effekt haben Kinderbetreuungsangebote auf die Beschäftigung der Eltern? Wie beeinflusst eine Rentenreform das Pensionsalter? Mit solchen Fragen beschäftigt sich die Politikevaluation. Sie untersucht datenbasiert, welche Wirkung eine bestimmte Massnahme (zum Beispiel eine Weiterbildung) auf ein interessierendes Ergebnis (Wiederbeschäftigung) hat. In der Wirtschaftspolitik bilden Politikevaluationen oft die Grundlage von Kosten-Nutzen-Analysen staatlicher Interventionen.

Um die Wirkung einer Massnahme zu messen, müssen andere Merkmale, die das Ergebnis beeinflussen, für Gruppen mit und ohne Massnahme konstant gehalten werden. Zum Beispiel entspricht ein Vergleich der Durchschnittslöhne von Gruppen mit und ohne Weiterbildung nur dann dem Massnahmeneffekt, wenn beide Gruppen hinsichtlich arbeitsmarktrelevanter Merkmale (wie Alter, Bildung, Arbeitsmarkterfahrung) vergleichbar sind. Ansonsten werden Äpfel mit Birnen verglichen, sodass der Mass-

nahmeneffekt mit der Wirkung der Merkmale vermischt wird. In Experimenten wird die Vergleichbarkeit gewährleistet, indem eine Massnahme zufällig (also unabhängig von Merkmalen) zugewiesen wird. Im nicht experimentellen Kontext ist dies in der Regel nicht möglich. Beispielsweise unterscheiden sich Personen, die eine Weiterbildung besuchen, von jenen, die dies nicht tun, weil individuelle Merkmale die Teilnahmeentscheidung beeinflussen.

Die Politikevaluation hat diverse kausale Methoden entwickelt, unter anderem basierend auf sogenannten Quasi-Experimenten, die dem Kontext eines Experiments nahekommen. Anhand einer solchen Methode hat eine Lausanner Studie im Jahr 2008 beispielsweise untersucht, wie sich die verlängerte Arbeitslosenunterstützung von Personen ab 50 Jahren in Österreich auf die Dauer der Arbeitslosigkeit auswirkt.<sup>1</sup> Bei einem Vergleich der Ergebnisse (wie der Arbeitslosendauer) von Personen knapp über und unter der Altersgrenze, zum Beispiel von 50- und 49-Jährigen, erscheint es unter bestimmten Annahmen plausibel, dass beide Gruppen aufgrund des geringen Altersunterschieds vergleichbare Merkmale aufweisen. Somit lässt sich der Massnahmeneffekt für Personen nahe der Altersgrenze ähnlich einem Experiment evaluieren, was als Regressions-Diskontinuitäts-Analyse («regression discontinuity») bekannt ist.

## Per Los zur Weiterbildung

In einer weiteren Methode wird nicht die Massnahme selbst, sondern ein sogenanntes Instrument «quasi-zufällig» zugeteilt, welches das Ergebnis einzig über den Effekt auf die Massnahme beeinflusst. Ein Beispiel dafür ist die

<sup>1</sup> Lalive (2008).

<sup>2</sup> Schochet et al. (2008).

<sup>3</sup> Berger und Lanz (2019).

Analyse des amerikanischen «Job Corps»- Programms aus dem Jahr 2008: Anhand einer Auslosung wurden benachteiligte Jugendliche in den USA eingeladen, eine Weiterbildung zu besuchen.<sup>2</sup> Die Einladung (Instrument) hat einen Effekt auf die Massnahme (Weiterbildung), weil sie bestimmte Personen zur Teilnahme bewegt. Ferner beeinflusst sie das Ergebnis, beispielsweise den Lohn, annahmegemäss einzig über die Teilnahmeentscheidung. Der Effekt einer Massnahme auf den Lohn lässt sich hier ermitteln, indem man den Einfluss der Einladung auf den Lohn durch den Einfluss der Einladung auf die Massnahme dividiert.

Ein anderer Ansatz ist die sogenannte Differenz-in-Differenzen-Methode. Sie basiert auf der Annahme, dass sich die Durchschnittsergebnisse zweier Gruppen mit und ohne

Wie misst man die Wirksamkeit eines Sprachkurses für Stellensuchende?

Massnahme über die Zeit gleich verändert hätten, wenn keine Gruppe die Massnahme erhalten hätte. Als Illustration dient eine aktuelle Studie zur Einführung eines Mindestlohns in manchen Schweizer Regionen und dessen Effekt auf die Beschäftigung.<sup>3</sup> Hier entspricht ein naiver Vergleich der Beschäftigung zwischen Regionen mit und ohne Mindestlohn nicht dem Massnahmeneffekt, wenn Unterschiede in beschäftigungsrelevanten Merkmalen wie der Branchenstruktur bestehen. Auch ein Vergleich innerhalb der Regionen mit einem Mindestlohn über die Zeit, das heisst vor und nach der Einführung der Massnahme, scheitert, wenn die Konjunktur über die Zeit schwankt: Die Vorher-nachher-Differenz in der Beschäftigung vermischt den Massnahmeneffekt mit dem konjunkturellen Trend. Falls aber Re-



KEystone

---

## Die Universität Freiburg bietet neu einen Master in Data Analytics and Economics an.

gionen mit und ohne Mindestlohn im Durchschnitt dem gleichen Trend unterliegen, so kann dieser durch die Vorher-nachher-Differenz in Regionen ohne Mindestlohn ermittelt werden. Folglich kann der Massnahmeneffekt als Differenz der Vorher-nachher-Differenz der Regionen mit Mindestlohn (Massnahmeneffekt plus Trend) und der Vorher-nachher-Differenz der Regionen ohne Mindestlohn (Trend) berechnet werden. Nach diesem Prinzip funktioniert die Differenz-in-Differenzen-Methode.

Der Ansatz der «Selektion aufgrund beobachteter Merkmale» wiederum nimmt an, dass alle Merkmale, die sowohl die Massnahme als auch das Ergebnis beeinflussen, in den Daten beobachtet werden. In diesem Fall wird die Massnahme unter Personen mit vergleichbaren beobachteten Merkmalen «quasi-experimentell» zugeteilt und hängt nicht von unbeobachteten Merkmalen ab, die gleichzeitig das Ergebnis beeinflussen. Der Massnahmeneffekt lässt sich somit identifizieren, indem man Gruppen mit und ohne Massnahmen formt, die ansonsten vergleichbare beobachtete Merkmale aufweisen.

Die Politikevaluation hat sich in den letzten Jahrzehnten dynamisch weiterentwickelt – nicht nur hinsichtlich konzeptioneller Ansätze zur Messung von Kausalität, sondern auch hinsichtlich deren Umsetzung in statistischen Verfahren.<sup>4</sup> Traditionellere Verfahren unterstellen, dass der Zusammenhang zwischen Ergebnis und beobachtbaren Merkmalen und/oder der Massnahme einer bestimmten, zum Beispiel linearen Funktion entspricht, wodurch etwa ein zusätzliches Jahr an Arbeitsmarkterfahrung immer denselben Lohneffekt aufweisen muss, egal wie hoch die Arbeitsmarkterfahrung bereits ist: Derartige Annahmen sind in neueren Verfahren der Politikevaluation wie dem sogenannten Matching überflüssig.

Die Matching-Methode findet Paare von Personen mit und ohne Massnahme in den Daten, die sich hinsichtlich beobachtbarer Merkmale möglichst ähnlich sind. Somit können Gruppen mit und ohne Massnahme generiert werden, de-

ren Merkmale vergleichbar sind. Beispielsweise wurde die Matching-Methode bereits 2002 zur Evaluation von Arbeitsmarktprogrammen in der Schweiz eingesetzt.<sup>5</sup> Damals fanden sich positive Beschäftigungseffekte von temporären Lohnsubventionen.

Ein weiteres Verfahren gewichtet Personen basierend auf ihrer Wahrscheinlichkeit, die Massnahme zu erhalten, sodass die gewichteten Gruppen mit und ohne Massnahme wiederum vergleichbar in den beobachteten Merkmalen sind. Die Gewichtung kann zudem mit einer (dem Matching verwandten) Regression kombiniert werden. Da es ausreicht, wenn zumindest eines der beiden Verfahren korrekt funktioniert, wird diese Methode als doppelt robuste Schätzung bezeichnet. Eine Studie aus dem Jahr 2016 wies beispielsweise mit der doppelt robusten Schätzung in der Schweiz einen Zusammenhang zwischen den Direktverkäufen von Medikamenten durch Ärzte und dem Anstieg der Medikamentenpreise nach.<sup>6</sup>

## Maschinelles Lernen im Kommen

Vor dem Hintergrund der wachsenden Verfügbarkeit von digitalen Daten hat die Politikevaluation mit dem Kausalen Maschinellen Lernen (KML) eine weitere Innovation erfahren<sup>7</sup>: Maschinelle Lernalgorithmen lernen datengetrieben, eine Massnahme oder ein Ergebnis bestmöglich als Funktion beobachteter Merkmale vorherzusagen. Dies ist insbesondere dann hilfreich, wenn die Anzahl an Merkmalen so gross ist, dass sie ein Mensch kaum oder nicht mehr analysieren kann. Unter einer Selektion aufgrund beobachteter Merkmale wird KML deshalb zur datenbasierten Wahl der wichtigsten Merkmale benutzt, hinsichtlich derer Gruppen mit und ohne Massnahme vergleichbar zu machen sind. Ferner erlaubt die Methode, datengetrieben Subgruppen zu finden, für die der Massnahmeneffekt besonders gross oder klein relativ zu anderen Subgruppen ist. Dies ist wichtig, um politikrelevante Effektheterogenitäten zu finden und gegebenenfalls die Zielgruppe einer Massnahme anzupassen. Eine KML-Analyse von Stellensuchprogrammen in der Schweiz hat beispielsweise gezeigt, dass der Beschäftigungseffekt bei Arbeitslosen

<sup>4</sup> Siehe Imbens und Wooldridge (2009) für einen Überblick.

<sup>5</sup> Gerfin und Lechner (2002).

<sup>6</sup> Kaiser und Schmid (2016).

<sup>7</sup> Chernozhukov et al. (2018).

mit geringer Arbeitsmarkterfahrung und Bildung grösser ist als bei Personen mit besseren arbeitsmarktrelevanten Merkmalen.<sup>8</sup> Letzteren schadet die Massnahme generell sogar.

Darüber hinaus ermöglicht diese Methode das Lernen der optimalen Zuteilung einer (typischerweise begrenzt verfügbaren) Massnahme, damit sie ihre grösstmögliche Wirkung entfaltet.<sup>9</sup> Zum Beispiel maximiert ein optimal zuge teiltes Arbeitsmarktprogramm die Wiederbeschäftigung unter Arbeitssuchenden. Auch in der Privatwirtschaft lässt sich optimales Politiklernen («optimal policy learning») anwenden – etwa indem Detailhändler bestimmten Konsumenten gezielt Rabatte anbieten, um den Umsatz zu maximieren. Es ist deshalb nicht verwunderlich, dass nicht nur öffentliche und internationale Institutionen, sondern auch Technologieunternehmen verstärkt Ökonomen mit quantitativem Fokus für die Optimierung ihrer Prozesse einstellen. Diese Entwicklung

wird vermutlich weitere wirtschaftliche Bereiche erfassen. Auch Hochschulen reagieren zusehends auf das neue Anforderungsprofil einer ökonomischen Ausbildung gepaart mit modernen quantitativen Methoden. So hat die Universität St. Gallen jüngst die Masterausbildung «Quantitative Economics and Finance» neu strukturiert, und die Universität Freiburg bietet ab Herbst 2020 einen Master in «Data Analytics and Economics» an.

8 Knaus et al. (2018).  
9 Athey und Wager (2018).



**Martin Huber**  
Professor für Ökonometrie und Politikevaluation,  
Departement für Volkswirtschaftslehre, Universität  
Freiburg

**Literatur**

Athey S. and S. Wager (2018). Efficient Policy Learning, Diskussionspapier, Stanford University.

Berger M. und B. Lanz (2019). Adjusting to Minimum Wage Regulation: Evidence from a Direct-Democracy Experiment in Switzerland, Diskussionspapier, Universität Neuenburg.

Chernozhukov V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey und J. Robins (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters, *Econometrics Journal*, 21: C1-C68.

Gerfin M. und M. Lechner (2002). A Microeconomic Evaluation of the Active Labour Market Policy in Switzerland, in: *The Economic Journal*, 112: 854-893.

Imbens G. W. und J. M. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation, in: *Journal of Economic Literature*, 47: 5-86.

Kaiser B. und C. Schmid (2016). Does Physician Dispensing Increase Drug Expenditures? Empirical Evidence from Switzerland, in: *Health Economics*, 25: 71-90.

Knaus M., M. Lechner und A. Strittmatter (2018). Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach, Diskussionspapier, Universität St. Gallen.

Lalive R. (2008). How Do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach, in: *Journal of Econometrics*, 142: 785-806.

Schochet P. Z., J. Burghardt und S. McConnell (2008). Does Job Corps Work? Impact Findings from the National Job Corps Study, in: *American Economic Review*, 98: 1864-1886.