

Andreas Peichl*

Bessere Daten für die Wissenschaft!



Andreas Peichl

Durch den harten »Knextit« geht am ifo eine Ära zu Ende. Ich habe das Wirken von Meinhard Knoche zwar meistens nur aus der Ferne verfolgt und erst im letzten halben Jahr in den Verhandlungen über meinen Wechsel ans ifo direkt mit ihm zu tun bekommen. Gleichwohl habe ich schon sehr früh gemerkt, dass Meinhard Knoche immer alles gibt, um die Bedingungen für die Wissenschaftler am ifo (und auch an der LMU) zu verbessern. Eines seiner zentralen Themen hierbei ist der Zugang zu Daten, was man beispielsweise am LMU-ifo Economics & Business Data Center (EBDC) sieht. In diesem kurzen Beitrag argumentiere ich, dass es zwar in den letzten Jahren schon einige Fortschritte in der Datenverfügbarkeit in Deutschland gegeben hat, gleichzeitig aber noch weitere Anstrengungen erforderlich sind, um den Wissenschaftsstandort Deutschland international kompetitiv zu machen.¹

Wir brauchen eine bessere Datenlage für die Wissenschaft in Deutschland. Hierfür gibt es zwei Hauptgründe:

1. Der Wissenschaftsstandort Deutschland ist in Gefahr! In anderen Ländern tut sich einiges; nicht nur Skandinavien, wo man traditionell eine bessere Datenverfügbarkeit hat, sondern auch die USA und das UK ermöglichen deutlich mehr und besseren Zugriff auf administrative Daten (von Entwicklungsländern gar zu schweigen) – unter Berücksichtigung des Datenschutzes.
2. Bessere Politikberatung: Durch bessere Forschung mit besseren Daten können wir bessere evidenzbasierte Wirtschaftspolitik (-beratung) bieten. Und

das nützt allen Bürgerinnen und Bürgern und Unternehmen in Deutschland.

Die Datenlage in Deutschland ist besser als z.B. vor 20 Jahren. Aber »besser« ist nicht gleich »sehr gut«. Gerade im internationalen Vergleich mit anderen hochentwickelten Volkswirtschaften schneidet Deutschland schlecht ab. Deshalb muss noch mehr getan werden, insbesondere damit dem Wissenschaftsstandort Deutschland im internationalen Wissenschaftswettbewerb keine noch größeren Nachteile entstehen.

Begünstigt durch die technologische Entwicklung hat in der wirtschaftswissenschaftlichen Forschung seit Mitte der 1990er Jahre ein kontinuierlicher Schwenk weg von Befragungsdaten und hin zu administrativen Prozessdaten stattgefunden (vgl. Chetty 2012).

Der Zugang zu statistischen Daten hoher Qualität, insbesondere administrativen Daten der amtlichen Statistik, ist für eine effektive und internationalen Standards genügende wissenschaftliche Politikberatung entscheidend. Evidenzbasierte Politikberatung auf Basis von quasi-experimentellen Forschungsdesigns zur Etablierung kausaler Zusammenhänge erfordert dringend qualitativ hochwertige, große Mikrodatensätze. Erst der Zugriff auf große administrative Datensätze ermöglicht die Anwendung modernster empirischer Methoden, auf deren Basis Erkenntnisse für die evidenzbasierte wirtschaftspolitische Beratung abgeleitet werden können. Administrative Daten weisen drei entscheidende Stärken auf:

1. Sie bilden in der Regel die Grundgesamtheit der jeweiligen Wirtschaftssubjekte ab (beispielsweise alle der Umsatzsteuer unterliegenden Firmen), so dass allgemein gültige Aussagen gemacht werden und heterogene Effekte Berücksichtigung finden können.
2. Sie basieren nicht auf Selbstaussagen wie fragebogenbasierte Daten und sind somit genauer und weniger fehleranfällig.
3. Eine Verknüpfung über die Zeit über beispielsweise Sozialversicherungs- oder Steuernummern ist möglich, so dass lang- und kurzfristige Reaktionen auf Politikreformen gemessen werden können.

Solche administrative Daten sind in Deutschland jedoch für Forschungszwecke nur sehr eingeschränkt verfügbar. Positiv hervorzuheben ist das Datenangebot des Forschungsdatenzentrums der Bundesagentur für Arbeit, das Mikrodaten der Sozialversicherungs-

* Prof. Dr. Andreas Peichl ist Leiter des ifo Zentrums für Konjunkturforschung und Befragungen.

¹ Dieser Artikel basiert teilweise auf Drechsel-Grau, Peichl und Schmid (2015), vgl. auch Zentrum für Europäische Wirtschaftsforschung (2015) für Möglichkeiten zur Verbesserung der Datenlage in Deutschland.

meldungen der Wissenschaft zugänglich macht. Auch die Forschungsdatenzentren der Statistischen Ämter der Länder stellen eine Vielzahl von Daten zur Verfügung. Dazu gehören auch Daten im Bereich Steuern, aber gerade in diesem Bereich gibt es Potenzial für erhebliche Verbesserungen. Ein Hauptproblem des bestehenden Angebots ist aus wissenschaftlicher Sicht zum einen die fehlende horizontale Verknüpfung verschiedener Datensätze (beispielsweise Körperschafts-, Gewerbesteuer- und Umsatzsteuerstatistik sowie Unternehmensregisterdaten). Zum anderen fehlt in der Regel eine vertikale Verknüpfung über die Zeit zu Paneldaten. Wissenschaftlich sind jedoch gerade diese Informationen essentiell. Die zeitliche Verknüpfung ist entscheidend, um sowohl die kurzfristige als auch die mittel- und langfristige Reaktion von Wirtschaftssubjekten auf Politikreformen zu messen. Die horizontale Verknüpfung ist wichtig, damit eine Vielzahl von Informationen berücksichtigt werden können. Bestehende Datensätze aus einer Quelle, wie beispielsweise das Umsatzsteuerpanel, enthalten nur Informationen, die für das jeweilige Politikinstrument (hier die Umsatzsteuer) relevant sind. Wichtige andere Informationen (wie z.B. Gewinn oder Investitionen) sind jedoch nicht enthalten, und nur durch die Verknüpfung verschiedener Datensätze ist eine umfassende Analyse möglich. Eine solche Analyse ist erforderlich, um die Auswirkungen von Reformen zu analysieren, die für die Wirtschaftspolitik von besonderem Interesse sind, so beispielsweise die Auswirkungen von Steueränderungen auf Investitionen, Arbeitsplätze oder auch Steuereinnahmen.

Verknüpfungen über Datensätze hinweg erlauben darüber hinaus auch Qualitätskontrollen und Kostensenkungen für administrative Prozesse. Thomsen und Holmøy (1998) zeigen etwa für Norwegen, dass sich die Datenqualität durch die Verknüpfung verschiedener Datensätze deutlich erhöht hat.

Aufgrund des fehlenden Datenangebots fällt Deutschland im internationalen Vergleich insbesondere hinter skandinavische Länder, aber auch die Vereinigten Staaten oder das Vereinigte Königreich, zurück. Die entsprechende Ausweitung des Datenangebots würde einen substantiellen wissenschaftlichen Fortschritt bedeuten und die Attraktivität des Forschungsstandorts Deutschlands insgesamt sichtbar erhöhen.

Eine verstärkte Forschung mit deutschen administrativen Daten würde auch die Möglichkeiten der evidenzbasierten Politikberatung in Deutschland erheblich verbessern. Administrative Daten erlauben es, vielfältige wirtschaftspolitische Fragestellungen zu untersuchen, da Firmen und Individuen über den Zeitablauf verfolgt werden können. So können z.B. Einflussfaktoren für Unternehmensgründungen und -schließungen sowie die Standortwahl von Unternehmen untersucht werden: Welche Rolle spielen hierbei veränderte (steuerliche) Rahmenbedingungen? Wie reagieren Unternehmen auf die Bereitstellung von Infrastruktur oder lokalen öffentlichen Gütern? Welchen Einfluss haben

wirtschaftspolitische Maßnahmen auf Investitionsentscheidungen? Warum verlagern Unternehmen ihren Produktionsstandort nach Osteuropa oder Asien? Die durch die Forschung mit administrativen Daten gewonnenen Erkenntnisse können somit gezielt genutzt werden, um die Attraktivität des Standortes Deutschland zu verbessern.

Es ist jedoch klar, dass die Bereitstellung verknüpfter administrativer Paneldaten für die Wissenschaft mit Kosten verbunden ist. Auch sind die in Deutschland hohen Anforderungen an den Datenschutz und das Steuergeheimnis zu berücksichtigen. Skandinavische Verhältnisse wird es sicherlich nicht geben. Aber auch in Ländern mit Steuergeheimnis und hohen Datenschutzstandards sind teilweise mehr Daten zugänglich und insbesondere Verknüpfungen über Datensätze hinweg möglich. Vielfach geschah dies im Rahmen langjähriger Prozesse. Finnland etwa hat im Zuge der Umsetzung der EU-Verordnung über europäische Statistiken große Schritte hin zu weitgehenden Datenzugangsmöglichkeiten unternommen. In der Folge wurde eine Kommission gegründet, die insbesondere zum Ziel hatte, der Wissenschaft besseren Datenzugang zu bieten, um auf dieser Grundlage die wirtschaftspolitische Beratung zu verbessern. Großbritannien und die USA sind gute Beispiele für Länder mit hohen Geheimhaltungspflichten und dennoch umfangreichen Datenzugängen. Erreicht wird dies in beiden Ländern etwa durch hausintern von den Steuerbehörden selbst eingerichtete Datenzentren. Die Forschungsdatenzentren der Bundesagentur für Arbeit haben bereits langjährige Erfahrung mit entsprechenden Modellen. Auch in Deutschland wären entsprechende Modelle denkbar, in denen die bestehenden Datensätze nur in den Finanzverwaltungen selbst einsehbar sind. Es sei auch noch darauf hingewiesen, dass die wissenschaftliche Analyse nicht an einzelnen Unternehmen oder Individuen interessiert sind, sondern im Rahmen von ökonomischen Analysen an Mittelwerten entweder über die Gesamtpopulation oder über bestimmte Gruppen.

Die Erfahrungen anderer Länder könnten insofern auch als Pilotprojekt dienen, den Datenzugang zu amtlichen Daten in Deutschland weiter zu verbessern und damit die evidenzbasierte wirtschaftspolitische Beratung zu stärken. Dies wäre sicherlich im Sinne von Meinhard Knoche!

LITERATUR

- Chetty, R. (2012), »Time Trends in the Use of Administrative Data for Empirical Research«, NBER Summer Institute, Juli.)
- Drechsel-Grau, M., A. Peichl und K.D. Schmid (2015), »Bessere Daten für die Wirtschafts- und Sozialforschung – eine Erwiderung«, *Wirtschaftsdienst* 95(12), 864–867.
- Thomsen, I. und A.M.K. Holmøy (1998), »Combining Data from Surveys and Administrative Record Systems. The Norwegian Experience«, *International Statistical Review* 66(2), 201–221.
- Zentrum für Europäische Wirtschaftsforschung (ZEW) (2015), *Möglichkeiten zur Verbesserung der statistischen Datengrundlage zur Beschreibung höchster Einkommen und Vermögen*, Gutachten für das Bundesministerium für Arbeit und Soziales, ZEW, Mannheim.