Modernisierung der Wirtschaftsstatistik

Potenziale der Digitalisierung nutzen und Rolle der Politikevaluation stärken

Die amtliche Statistik bildet eine wichtige Grundlage für eine evidenzbasierte Wirtschaftspolitik – eine Wirtschaftspolitik, die systematische und datenbasierte Analysen der Wirksamkeit von Maßnahmen in die Entscheidungsprozesse einbezieht. Im Zuge der Digitalisierung bieten sich vielfältige Möglichkeiten der Modernisierung der Wirtschaftsstatistik und Potenzial für effizientere und belastungsärmere Prozesse. Dabei sind stets die Anforderungen der statistischen Geheimhaltung zu beachten.

Informierte politische Entscheidungen sollten sich stets an der Lebenswirklichkeit der Menschen orientieren. Um diese erfassen und beschreiben zu können, kommt der amtlichen Statistik eine zentrale Bedeutung zu. Statistische Daten sind zudem wesentliche Grundlage einer evidenzbasierten, also an Daten und Fakten orientierten, Wirtschaftspolitik. Eine moderne Statistik steht dabei im Spannungsfeld zwischen der Erhebung qualitativ hochwertiger Daten auf der einen Seite und der Vermeidung unnötiger Kosten und Belastungen bei Bürgern, Unternehmen und Verwaltung auf der anderen Seite. Um beiden Zielsetzungen gerecht zu werden, müssen künftig in der Statistik vermehrt die Möglichkeiten der fortschreitenden Digitalisierung genutzt werden, ohne dabei die Anforderungen an die statistische Geheimhaltung aus dem Blick zu verlieren.

Durch die Digitalisierung eröffnet sich der Statistik vor allem ein großes Modernisierungspotenzial, das zu aussagekräftigeren, qualitativ besseren Daten führen kann, zugleich aber Erhebungs- und Verarbeitungsprozesse effizienter, kostengünstiger und belastungsärmer werden lässt. Jenseits der reinen Tatsachenbeschreibung von wirtschaftlichen oder sozialen Zusammenhängen sollten zudem neue, innovative Nutzungsmöglichkeiten der Statistik stärker ins Auge gefasst werden: Dazu gehört vor allem, statistische Daten stärker zur Evaluation politischer Maßnahmen oder zum Zwecke der wissenschaftlichen Forschung zu nutzen. Hier müssen im verfassungsrechtlichen Rahmen praktikable Wege zum Umgang mit den Daten geschaffen werden, die immer auch den hohen Stellenwert der statistischen Geheimhaltung im System der amtlichen Statistik berücksichtigen.

Der vorliegende Artikel diskutiert Wege zur Modernisierung einzelner Bereiche der Wirtschaftsstatistik im Zuge der Digitalisierung und nennt Beispiele bereits angestoßener Veränderungen. Bei der erwähnten Erweiterung der Datennutzungsmöglichkeiten könnten Forschungsdatenzentren künftig eine noch wichtigere Rolle einnehmen und neue Nutzungswege für die wissenschaftliche Forschung aufzeigen.

Die Federführung über die Wirtschaftsstatistik liegt innerhalb der Bundesregierung im Bundesministerium für Wirtschaft und Energie und umfasst insbesondere die Fachaufsicht über das Statistische Bundesamt, die Gesetzgebungsarbeit zur Neuschaffung oder Weiterentwicklung der wirtschaftsstatistischen Rechtsgrundlagen sowie die Arbeit in nationalen oder europäischen Gremien. Inhaltlich betrifft die Wirtschaftsstatistik vor allem die Bereiche Unternehmensstatistik, Preisstatistik, alle Konjunkturstatistiken sowie den Bereich der Volkswirtschaftlichen Gesamtrechnungen.

Digitalisierung ist eine Chance für die amtliche Statistik

Die amtliche Statistik in Deutschland hat gemäß § 1 des Bundesstatistikgesetzes die Aufgabe, „laufend Daten über Massenerscheinungen zu erheben, zu sammeln, aufzubereiten, darzustellen und zu analysieren.“ Dabei gelten die Grundsätze der Neutralität, Objektivität und fachlichen Unabhängigkeit. Die amtliche Statistik soll eine verlässliche und belastbare Datengrundlage zur Beschreibung der Lebensverhältnisse in unserem Land liefern – und dies in vielerlei Hinsicht: Sie umfasst wirtschaftliche und soziale Aspekte, aber auch viele weitere Bereiche wie zum Beispiel Wissenschaft, Kultur und Umwelt.

Die Digitalisierung bietet Gelegenheit zur grundlegenden Modernisierung der amtlichen Statistik, indem nach und nach einzelne Statistiken und die zugrundeliegenden Rechtsnormen den neuen technischen Möglichkeiten angepasst werden. Dadurch könnte in Wirtschaft und Verwaltung der Aufwand für die Erhebung und Aufbereitung von Statistiken erheblich reduziert werden. Ein Beispiel hierfür ist der Aufbau der Infrastruktur für eine registerbasierte Statistik, mit der die deutsche Registerlandschaft (z. B. Handelsregister, Melderegister, Personenstandsregister) perspektivisch weitgehend digital geführt und vernetzt werden könnte. Historisch bedingt ist die deutsche Registerlandschaft oft noch sehr dezentral und heterogen organisiert; eine grundlegende Modernisierung wird hier bereits vorbereitet. So soll der 2021 stattfindende Zensus zu einem großen Teil registerbasiert durchgeführt werden.

Eine Herausforderung einer solchen Modernisierung der Statistik ist vor allem der kurzfristige Umstellungsaufwand, der je nach konkretem Projekt erheblich sein kann. Demgegenüber stehen später potenziell deutlich reduzierte laufende Kosten und zugleich Effizienzgewinne, sowohl für die Verwaltung als auch für Bürger und Unternehmen.

Neue Technologien nutzen – Beispiel Scannerdaten

Über die beschriebenen Prozessmodernisierungen hinaus müssen mit dem technologischen Fortschritt aber auch etablierte Abläufe in der Statistik angepasst werden, um auch zukünftig ein zuverlässiges Abbild der Realität liefern zu können. Ein Beispiel ist die Erhebung der Verbraucherpreise im Rahmen der Preisstatistik.

Bislang werden viele Preise noch durch rund 600 Preiserheber individuell erfasst: Diese suchen in regelmäßigen Abständen Geschäfte des Einzelhandels und von Dienstleistern auf und zeichnen die Preise eines repräsentativen Warenkorbes auf. So werden monatlich etwa 300.000 Einzelpreise gewonnen. Zusätzlich werden im Onlinehandel mit Waren sowie für ausgewählte Dienstleistungen wie Pauschalreisen oder Fernbusse Preise zentral im Internet gesammelt. Dies erfolgt zum Teil noch manuell, zum Teil werden auch mittels Web Scraping, also dem maschinellen Auslesen von Informationen aus dem Internet, Preise automatisiert gewonnen. Daraus werden dann Preisschwankungen über die Zeit ermittelt und schließlich der Verbraucherpreisindex berechnet.

Die Nutzung der Daten von Scannerkassen könnte dazu beitragen, die bislang weitgehend manuelle Preiserhebung teilweise zu automatisieren und dadurch Erhebungen effizienter, kostengünstiger und bürokratieärmer zu gestalten sowie die Qualität der Preismessung zu verbessern. So könnten etwa individualisierte Preise, wie sie heute schon durch individuelle Rabattkarten gewährt werden und die sich künftig weiter verbreiten dürften, valide statistisch erhoben werden. Über die Daten der Scannerkassen würden sie als echte Transaktionspreise in großem Umfang erfasst und in den Verbraucherpreisindex einfließen. Hinzu kommt, dass mit der Datenerhebung über die Nutzung von Scannerdaten der elektronischen Kassensysteme des Einzelhandels auch die Umsätze mit erhoben werden könnten. Dies könnte zum einen die Qualität der Einzelhandelsumsatzstatistik verbessern und zum anderen den Erhebungsaufwand für diese Statistik merklich reduzieren. Die Integration von Scannerdaten in die amtliche Statistik ist jedoch methodisch und technisch anspruchsvoll[1]; vor einer vollständigen Nutzung sind daher umfangreiche Vorarbeiten und Erprobungsschritte in den Statistischen Ämtern erforderlich.

Eine solche modernisierte Preiserhebung wird in anderen Industrieländern bereits erfolgreich praktiziert. So verwenden u. a. Norwegen, die Niederlande, Dänemark und die Schweiz schon heute Scannerdaten bei der Berechnung der Verbraucherpreisentwicklung. Jüngst ebnete auch Frankreich den Weg dorthin mit besonders weitreichenden gesetzlichen Vorgaben: Dort kann der Handel verpflichtet werden, täglich Scannerdaten an die statistischen Behörden zu liefern.

Eine dynamische Preissetzung gewinnt neben den individualisierten Preisen im Einzelhandel vor allem auch durch eine wachsende Rolle des Onlinehandels an Bedeutung. Dort werden teilweise bereits mehrmals täglich Preise angepasst – dies kannten die Verbraucher bis vor wenigen Jahren fast nur von den Kraftstoffpreisen an Tankstellen. Dadurch gewinnen hier digitale Erhebungsmethoden zunehmend an Bedeutung, um die Qualität der Statistik auch zukünftig zu gewährleisten[2]. Zur Erfassung von Preisen im Onlinebereich können und werden inzwischen Techniken zur automatischen Preiserfassung im Internet (Web Scraping) eingesetzt. Jedoch gilt auch hier, dass sich die Technologie insgesamt schneller entwickelt als die entsprechende Regulierung, die nun nachziehen muss.

Für die umfassende Nutzung dieser Techniken müssten in der Wirtschaftsstatistik entsprechende Rechtsgrundlagen geändert werden, denn der Einsatz von Scannerdaten für die amtliche Statistik ist bisher nicht und jener von Web Scraping nur eingeschränkt möglich. Für eine breitere Nutzung von Web Scraping wäre eine entsprechende Rechtsgrundlage erforderlich. Zur Erprobung der Nutzung von Scannerdaten und der Entwicklung einer gesetzlichen Regelung hierfür bietet sich ein geschütztes Umfeld eng begrenzter, aber unter realen Bedingungen stattfindender Pilotprojekte (so genannte „Reallabore“) an. Derzeit werden im BMWi verschiedene solcher Pilotprojekte konzipiert.

1 Vgl. zum methodischen Ansatz der Niederlande zur Integration von Scannerdaten in den VPI: von Auer, Ludwig (2017): „Processing Scanner Data by an Augmented GUV Index“, Eurostat Review On National Accounts and Macroeconomic Indicators (EuroONA), 1/2017, 73–91.

2 Zudem werden insbesondere neue, innovative digitale Produkte in der Anfangsphase häufig nicht ausreichend von der Statistik erfasst, während etablierte Digitalprodukte hier in der Regel unproblematisch sind; vgl. Statistisches Bundesamt (2016): „Methodeninformation: Auswirkungen der Digitalisierung auf die Preisstatistik“, Wiesbaden.

Amtliche Statistik stärker für wissenschaftliche Zwecke und zur Evaluation wirtschaftspolitischer Politikmaßnahmen nutzen

Eine moderne Statistik sollte mehr leisten als eine wirklichkeitsgetreue Beschreibung der Lebensverhältnisse in einem Land. Gerade statistische Daten bieten sich zur Politikevaluation durch Wissenschaft und Verwaltung an. Denn sie erfüllen viele der Anforderungen, die Empiriker heutzutage idealerweise an ihre Datengrundlage stellen.

So sollen Daten

- zuverlässig und valide erhoben sein; hier sind insbesondere amtliche Registerdaten objektiver als etwa Daten, die auf freiwilligen Befragungen beruhen;
- die zugrundeliegende Grundgesamtheit möglichst breit abdecken, so dass Stichprobenfehler minimiert werden;
- durch zeitliche Verknüpfungen von wiederholten Erhebungszeitpunkten (Paneldaten) ein hohes Analysepotenzial ermöglichen, um Effekte bestimmter Maßnahmen bewerten zu können;
- genutzt werden können, um durch Verknüpfung mit anderen Datensätzen das wissenschaftliche Analysepotenzial zu erhöhen.

Im Rahmen der amtlichen Statistik erhobene Verwaltungsdaten erfüllen prinzipiell alle vier oben genannten Punkte. Dies macht sie gerade vor dem Hintergrund des immer wichtiger werdenden Paradigmas einer evidenzbasierten Politik zu einer wichtigen Datenquelle in der Politikberatung und -evaluation.

Den Potenzialen aus Datenverknüpfungen steht aber der hohe Stellenwert des Datenschutzes und der Persönlichkeitsrechte in Deutschland gegenüber, vor allem wenn Daten von Einzelpersonen betroffen sind. Die individuellen Schutzrechte von Personen (Recht auf informationelle Selbstbestimmung) sowie die Abwehrrechte des Einzelnen gegenüber dem Staat sind sehr weit gefasst, und Einschränkungen derselben unterliegen hohen Rechtfertigungshürden. Ähnliches gilt für den Geheimschutz in der öffentlichen Verwaltung und der Wirtschaft. Neben expliziten Regelungen wie dem Steuer-, Sozial- oder Statistikgeheimnis gilt in der deutschen Verwaltung allgemein der so genannte Amtsverschwiegenheitsgrundsatz.

Soweit es zur Gewinnung von statistischen Informationen ohne zusätzliche statistische Erhebungen erforderlich ist, dürfen zwar Daten aus Wirtschafts- und Umweltstatistiken zusammengeführt werden. Aus datenschutzrechtlichen Gründen ist es allerdings nicht zulässig, Individualdaten von Personen ohne eine spezielle Rechtsgrundlage mit anderen Individualdaten zu verknüpfen. Mit Einschränkungen gilt dies auch für wahrscheinlichkeitsbasierte Verfahren, bei denen keine direkte Verknüpfung zweier identischer Merkmalsträger stattfindet, sondern statistische Verfahren (z. B. Regressionsverfahren) zur Zuordnung genutzt werden; solche Verfahren müssen im Einzelfall im Rahmen einer Risikoabschätzung auf ihre Vereinbarkeit mit dem Datenschutz geprüft werden. Damit soll ausgeschlossen werden, dass sich verschiedene Datensätze so verknüpfen lassen, dass Rückschlüsse auf die Auskunftsgebenden möglich sind. Eine Verknüpfung gleicher Befragungseinheiten ist daher generell nur zulässig, wenn eine entsprechende Einwilligung der Befragten explizit eingeholt wird oder wenn – wie bei den Zensusgesetzen – eine spezielle Rechtsgrundlage vorhanden ist. Jede potenzielle neue Rechtsgrundlage zur Verknüpfung von Individualdaten müsste vorab umfassend verfassungsrechtlich geprüft werden.

Ein wichtiges Mittel, um die Verknüpfung von Individualdaten praktikabel zu gestalten, wäre die Verwendung von Personenkennziffern als gemeinsames Identifikationsmerkmal (Identifier). Hierzu führt ein aktuelles Gutachten des Normenkontrollrates aus, dass Personenkennziffern, die bereichsspezifisch, also an einen sachlichen Erhebungszweck gebunden sind, grundsätzlich auch in Deutschland verfassungskonform implementiert werden können[3]. Damit würden sich die Möglichkeiten für rechtlich zulässige Datenverknüpfungen beispielsweise im Rahmen eines registerbasierten Zensus deutlich erweitern.

Breiteren Datenzugang in den Forschungsdatenzentren schaffen

Künftig sollten Forschungsdatenzentren des Bundes und der Länder eine prominentere Rolle einnehmen, um berechtigten Nutzern den Datenzugang zu erleichtern. Schon heute ermöglichen die Forschungsdatenzentren der statistischen Ämter Wissenschaftlern einen Zugang zu den Mikrodaten der amtlichen Statistik und stellen damit einen wesentlichen Teil der Forschungsinfrastruktur in Deutschland. Das Statistische Bundesamt sowie die Statistischen Landesämter haben Forschungsdatenzentren an verschiedenen Standorten eingerichtet.

Künftig sollte der Rahmen der Datennutzung erweitert werden, um unter bestimmten Voraussetzungen für Daten der amtlichen Statistik zeitgemäße empirische Methoden der Datenaufbereitung und -verknüpfung nutzbar zu machen. So ermöglichen etwa große Mikrodatensätze hoher Qualität quasi-experimentelle Forschungsdesigns mit einem hohen kausalen Erklärungsgehalt. Um diese Möglichkeiten zu nutzen, müssten Mikrodaten aber horizontal, also zwischen verschiedenen Datensätzen, und vertikal, also über die Zeit, verknüpfbar sein. Die möglichen Anwendungsfälle für die Politikberatung und -evaluation sind zahlreich: Ein Beispiel ist die Evaluierung politischer Vorhaben, bei der erst Datensätze verschiedener amtlicher Statistiken zusammen ein aussagekräftiges Gesamtbild ergeben. Somit würde ein besserer Datenzugang auch der evidenzbasierten Wirtschaftspolitik und -beratung zugutekommen, für die Kausalzusammenhänge (die sich mit einer Verknüpfung von Daten darstellen lassen) besonders wertvoll sind.[4]

Eine prominentere Rolle der Forschungsdatenzentren für einen solchen Datenzugang müsste dabei flankiert werden von einer entsprechenden Anpassung der rechtlichen Rahmenbedingungen. Damit unter sich weiterentwickelnden technischen Vorzeichen gewährleistet werden kann, dass

3 Vgl. Martini, Mario; Wagner, David; Wenzel, Michael (2017): „Rechtliche Grenzen einer Personen- bzw. Unternehmenskennziffer in staatlichen Registern“, Speyer 2017.

4 Vgl. etwa nachstehendes Gutachten im Auftrag des BMAS: Löffler, Max; Peichl, Andreas; Wittneben, Christian; Neisser, Carina (2015): „Möglichkeiten zur Verbesserung der statistischen Datengrundlage zur Beschreibung höchster Einkommen und Vermögen“, Bundesministerium für Arbeit und Soziales, Berlin.

eine moderne Statistik auch weiterhin den strengen Anforderungen des deutschen Datenschutzes gerecht wird, hätten die Forschungsdatenzentren künftig entsprechende prozedurale und technische Maßnahmen zur faktischen Anonymisierung speziell großer Mengen von Mikrodaten in Strukturen komplexer Verknüpfungen zu treffen. Schon heute ermöglichen Forschungsdatenzentren wissenschaftlichen Nutzern Wege der Nutzung amtlicher Daten, insbesondere den Zugang zu anonymisierten Einzeldaten, die anderweitig nicht zugänglich gemacht werden könnten. Wissenschaftler können dort bereits jetzt an speziell geschützten Arbeitsplätzen direkt mit Mikrodaten arbeiten (On-Site-Nutzung). Diese On-Site-Nutzung sollte perspektivisch ausgebaut werden, um anders nicht mögliche Analyse- und Verknüpfungsmethoden für Mikrodaten in einem solch geschützten Umfeld zu ermöglichen. Alternative Nutzungswege sollten bestehen bleiben und ebenfalls ausgebaut werden, wie etwa die Übermittlung vorab erstellter Auswertungsroutinen, so dass der berechtigte Nutzer lediglich entsprechend anonymisierte oder aggregierte Ergebnisse zurückerhält.

Bereits heute unterhalten in Deutschland abseits der Statistischen Ämter verschiedene weitere Behörden Forschungsdatenzentren. So betreibt die Bundesagentur für Arbeit über ihre Forschungseinrichtung, das Institut für Arbeitsmarkt- und Berufsforschung (IAB), ein auch international beachtetes Forschungsdatenzentrum (Standorte in Deutschland, im Vereinigten Königreich und in den Vereinigten Staaten). Hier wird der Wissenschaft unter Wahrung deutscher Datenschutzstandards Zugang zu den Mikrodaten der Bundesagentur für Arbeit gewährt. Eine Datenverknüpfung ist jedoch auch innerhalb des IAB nur mit ausdrücklicher Zustimmung der Befragten möglich. Weitere Forschungsdatenzentren existieren beispielsweise innerhalb der Deutschen Bundesbank sowie bei der Deutschen Rentenversicherung Bund. Thematisch decken diese Forschungsdatenzentren einen breiten Bereich ab, der neben der Wirtschaftspolitik viele weitere Politikbereiche umfasst, etwa die Arbeitsmarkt- und Sozialforschung sowie die Forschung zur Familien- und Bildungspolitik. Gerade in letztgenanntem Bereich unterstützt das Bundesministerium für Bildung und Forschung die entsprechenden Forschungsdatenzentren auch finanziell. Mit dem Rat für Sozial- und Wirtschaftsdaten (RatSWD) fördert das BMBF zudem ein Gremium, das sich für weitere Verbesserungen der Dateninfrastruktur einsetzt; hier sind Forschende verschiedener Disziplinen sowie Datenproduzenten vertreten.

Rechtlichen Rahmen für mehr Kooperation zwischen Forschungsdatenzentren anpassen

Da die bestehenden Zentren bereits einen Großteil der amtlichen Datenquellen abdecken, geht es nicht in erster Linie darum, zusätzliche Forschungsdatenzentren zu gründen. Die Statistischen Ämter erlauben über ihre Forschungsdatenzentren den Zugang zu einer Breite regionaler und überregionaler Statistiken; im Statistischen Bundesamt liegen auch die für die Wirtschafts- und Sozialwissenschaften häufig besonders erkenntnisreichen Daten aus den steuerlichen Veranlagungen als Einzeldatenmaterial vor. Weitere überregionale Daten zur Beschäftigung führt die Bundesagentur für Arbeit.

Vor diesem Hintergrund gilt es, den rechtlichen Rahmen für eine stärkere Kooperation zwischen den bestehenden Forschungsdatenzentren zu schaffen. So wäre eine technisch bereits heute mögliche Verknüpfung von Steuerdaten mit Beschäftigungsdaten auf Mikrodatenebene (etwa über den Abgleich von Namen und Adresse oder perspektivisch über einen „Identifier“ wie die Steueridentifikationsnummer) auch rechtlich abgesichert und könnte durch entsprechenden Datenaustausch in einem Forschungsdatenzentrum des Statistischen Bundesamtes oder der Bundesagentur für Arbeit stattfinden[5]. Die technischen Anpassungen, um innerhalb der Forschungsdatenzentren für die weiter oben umrissenen komplexen empirischen Analysen Geheimhaltung und Anonymität zu gewährleisten, erscheinen realisierbar. Es müssten jedoch eine Infrastruktur für Datenaustausch und -verknüpfung sowie ein entsprechender rechtlicher Rahmen geschaffen werden. Die gesetzlichen Vorschriften müssen so angepasst werden, dass unter Wahrung des deutschen Datenschutzes ein solcher Austausch möglich und zulässig ist. So sind insbesondere individuelle Steuerdaten durch das Steuergeheimnis besonders streng geschützt. Für eine Nutzung müssten entsprechend hohe Anforderungen gestellt werden. Die Forschungsdatenzentren der amtlichen Statistik könnten ein Muster liefern, wie eine solche Kooperation technisch und organisatorisch zu leisten ist.

Für Wissenschaft sowie politische Evaluation und Beratung würde sich ein deutlich breiteres Analysepotenzial erschließen.

Kontakt: Matthias Pannhorst
Referat: Wachstum, Demografie, Statistik

5 Vgl. auch das Petitum in Peichl, Andreas (2017): „Bessere Daten für die Wissenschaft“, ifo Schnelldienst, 13/2017, 55–56.