

BIG DATA IN DER KONJUNKTURANALYSE

WELCHES POTENZIAL HABEN ALTERNATIVE DATENQUELLEN?

Zeitnahe und präzise Informationen über die wirtschaftliche Entwicklung sind entscheidend für eine effektive Wirtschaftspolitik, die möglichst zielgenau auf den jeweiligen Zustand der Volkswirtschaft abgestimmt sein sollte. Im Idealfall stünden Entscheidungsträgern tagesaktuelle Echtzeitdaten zur Verfügung, die sie in ihre Überlegungen einfließen lassen könnten. Allerdings können in der Realität viele relevante Daten nur mit deutlicher Verzögerung bereitgestellt werden: Die umfassende Prüfung und die Zusammenführung vieler Einzeldaten sind enorm aufwendig und zeitintensiv, wenn eine gewohnt hohe Qualität der amtlichen Statistik gewährleistet werden soll.

Bekanntestes Beispiel ist die Veröffentlichung der amtlichen Daten zum Bruttoinlandsprodukt (BIP): Diese sind erst 30 Tage nach Ablauf des jeweiligen Quartals als Schnellmeldung verfügbar; die Zusammenstellung der Detailmeldung benötigt fast zwei Monate. Die Konjunkturanalyse und -prognose ist daher auch auf die Auswertung von höherfrequenten Konjunkturindikatoren angewiesen, die früher vorliegen als etwa Quartalsdaten. Diese ermöglichen eine Einschätzung der aktuellen wirtschaftlichen Entwicklung deutlich zeitnäher. In der traditionellen Konjunkturbeobachtung fallen unter diese Kategorie der höherfrequenten Indikatoren sowohl monatliche Statistiken etwa zur Produktion oder zum Außenhandel als auch umfragebasierte Stimmungsindikatoren.

Neue Chancen ergeben sich in diesem Kontext durch die beeindruckenden Fortschritte der letzten Jahre im Bereich der künstlichen Intelligenz zusammen mit der Digitalisierung fast aller Lebensbereiche. Vor allem die Daten als Nebenprodukt unserer Nutzung von digitalen Diensten und Produkten eröffnen ein neues Potenzial für zeitnahe Aussagen zu unserem Verhalten und zur wirtschaftlichen Aktivität. Diese auch als „Big Data“ bekannten, großen, zunächst häufig unstrukturierten Datenquellen und Datenmengen können wertvolle neue Informationen

DIGITALE DIENSTE ERÖFFNEN ZUSÄTZLICHE DATENQUELLEN.

meist in deutlich höherer Frequenz und Aktualität enthalten. Diese neue Art von Daten weist allerdings eine hohe Bandbreite und Komplexität auf, weil sie eben nicht ursprünglich für die Analyse von ökonomischen Fragestellungen erhoben wurden, sondern vielmehr aus der Nutzung unterschiedlicher Anwendungen hervorgehen. Zur effizienten und zielgerichteten Verarbeitung solcher Daten werden daher zumeist auch neue Methoden der Auswertung nötig. Diese werden ebenfalls fortlaufend weiterentwickelt, spielen aber bisher in der makroökonomischen Analyse eher eine untergeordnete Rolle.

IN KÜRZE

Monatliche Daten und Indikatoren helfen, Analysen möglichst aktuell zu halten.



AN 5 EIGENSCHAFTEN

werden Big Data, die in verschiedenen Lebensbereichen gesammelt werden, häufig festgemacht – u. a. hohes Volumen und hohe Frequenz.

WAS GENAU IST BIG DATA?

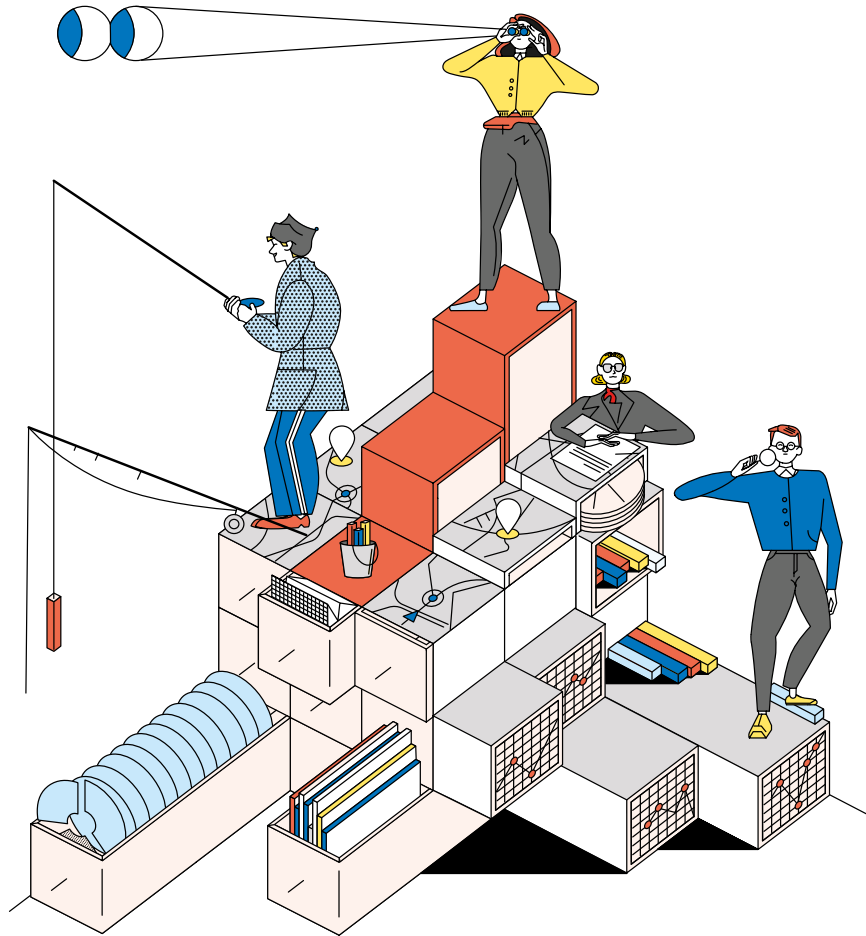
Eine erste Herausforderung, die sich im Kontext der Arbeit mit Big Data stellt, ist die klare Abgrenzung des Begriffs. Bis heute liegt keine offizielle, allgemein gültige Definition von Big Data vor. Vielmehr sind es fünf generelle Eigenschaften, an denen Big Data häufig festgemacht werden, und die in der Fachliteratur in der Regel als relevante

Kriterien anerkannt werden. Gemäß diesen fünf Eigenschaften, die auch als „5 Vs“ bekannt sind, liegen Big Data generell in hohem Volumen vor („volume“), das sich aufgrund der hohen Frequenz, mit der diese Daten generiert werden üblicherweise auch rasch vergrößert („velocity“). Aufgrund der Tatsache, dass die Daten zumeist als Nebenprodukt anderer Prozesse anfallen, sind sie in der Regel sehr viel komplexer und breiter angelegt →

IN KÜRZE

Zu den „5 Vs“ zählen neben hohem Volumen auch größere Schwankungen bei Qualität und Eigenschaften der Daten.

DIE ZAHL DER DATENQUELLEN IST IN DEN LETZTEN JAHREN STARK GEWACHSEN.



(„variety“), als es von konventionellen Daten bekannt ist. Hinzu kommt, dass oftmals auch die Datenqualität geringer ausfallen kann, als in der amtlichen Statistik üblich – nicht zuletzt, weil die Daten zum Teil verzerrt oder nicht repräsentativ sind („veracity“). Schließlich kann das technologisch und geschäftlich sehr dynamische Umfeld, aus dem Big Data zum Teil stammen, dazu führen, dass sich Eigenschaften und Abdeckung der Daten im Zeitablauf deutlich verändern oder sogar die längerfristige Verfügbarkeit der Daten in Frage gestellt ist („volatility“).

WELCHE DATENQUELLEN SIND GEEIGNET?

Auch wenn die Abgrenzung von Big Data teilweise schwammig erscheint, zeichnen sich einige vielversprechende Datenquellen mit unterschiedlichen Einsatzgebieten ab. Darunter ist zunächst die Kategorie der Texte zu nennen. Damit gemeint sind vor allem Nachrichten aus Print- und Online-Medien sowie sozialen Medien (Twitter, Facebook etc.).

Nachrichten und Presseartikel sowie Daten aus den sozialen Medien können genutzt werden, um Indikatoren zur wirtschaftspolitischen oder Verbraucherstimmung zu bilden. Ein Nachteil ist, dass hier der Aufwand für die Schaffung einer Datengrundlage und deren Auswertung recht hoch ist. Daten aus Internetsuchanfragen sind einfacher auszuwerten und absolut zeitnah verfügbar und könnten sich ebenfalls dazu eignen, Stimmungsindikatoren zu erstellen. Allerdings ist auch hier die zentrale Herausforderung die automatisierte Beschaffung und Interpretation der Daten.

Eine weitere vielversprechende Datenquelle sind Daten von Online-Handelsplattformen oder Supermarkt-Käufen. Sie sind vor allem für Analysen von Preisen und allgemeinem Kaufverhalten nützlich. Sie sind zudem zeitnah verfügbar, umfassend und präzise. Preisdaten von Online-Händlern oder Preisvergleichsplattformen könnten mit relativ geringem Erstaufwand automatisiert gesammelt

IN KÜRZE

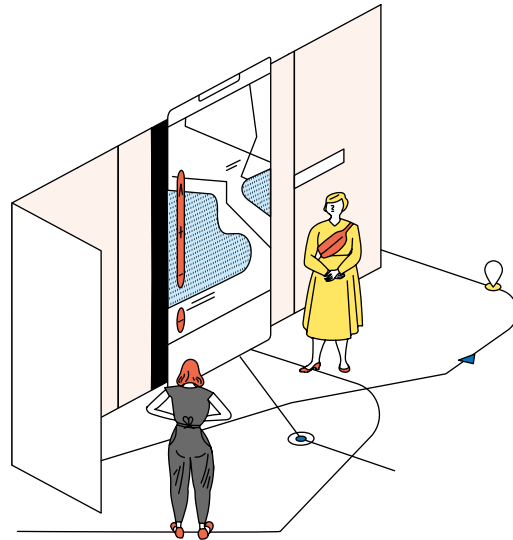
Die Auswertung von Daten aus Nachrichten, Presseartikeln und sozialen Medien ist aufwendig.

werden. Bei diesem „Web Scraping“ genannten Vorgehen werden automatisiert Webseiten ausgelesen und die relevanten Daten wie Produktname und Preis in einer Datenbank gespeichert. Auch wenn der Erstaufwand für dieses Vorgehen gering ist, so ist doch ein zeitlicher Vorlauf nötig, bis ausreichend Daten zur Verfügung stehen. Aufgrund der höheren Folgekosten, die sich aus der regelmäßigen Erfassung und der Pflege der Datenbank ergeben, wird hier weiteres Potenzial vor allem für die amtliche Statistik gesehen, die teilweise bereits entsprechende Methoden anwendet.

Transaktionsdaten des elektronischen Zahlungsverkehrs bilden wirtschaftliche Aktivität zeitnah und unmittelbar ab und haben daher im Prinzip ein sehr großes Potenzial, vor allem für das „Nowcasting“ – das heißt, die Prognose der gegenwärtigen Lage – makroökonomischer Größen. Als Hürde erweist sich hier vor allem der Zugang zu den Daten. Deshalb ergeben sich hier allenfalls Möglichkeiten für Zentralbanken und Statistikämter, für die der stetige Zugang einfacher zu erlangen sein könnte.

IMMER MEHR DATEN WERDEN AUTOMATISIERT ERFASST UND GESAMMELT.

Zu einer weiteren Datenkategorie gehören Satellitendaten (allgemeiner: „Fernerkundungsdaten“), Mobilfunkdaten sowie experimentelle Ansätze wie Messungen der Lkw-Fahrleistung aus Maut-Daten, des Containerumschlags, des Umsatzes an Zollstellen oder auch die Analyse von Schiffspositionsdaten. Diese Verkehrsdaten haben das Potenzial, Informationen über die laufende wirtschaftliche Entwicklung zu liefern oder Handelsaktivität abzubilden. Insbesondere Maut-Daten aus dem Lkw-Verkehr werden



bereits vom Bundesamt für Güterverkehr und vom Statistischen Bundesamt produziert und gelten als geeignet für die Prognose beispielsweise der Industrieproduktion.

LKW-VERKEHRSDATEN KÖNNEN WICHTIGE INFORMATIONEN LIEFERN.

Fernerkundungsdaten können Informationen zu Infrastruktur oder Witterungsbedingungen liefern. Zukünftig könnte sich durch eine zeitaktuelle Erfassung und Auswertung von Satellitenbildern die Möglichkeit ergeben, die laufende wirtschaftliche Aktivität besser zu erfassen (beispielsweise durch Analyse von Bauaktivität auf Baustellen). Allerdings ist damit aus heutiger Sicht ein sehr hoher Aufwand verbunden. Vor allem in industrialisierten Volkswirtschaften sind diese Informationen eher nicht dazu geeignet, kurzfristige Schwankungen der wirtschaftlichen Aktivität abzubilden. Das Problem sind hier vor allem die nicht zeitnahe Erfassung und die technisch anspruchsvolle „Interpretation“ der Satellitendaten durch Algorithmen. Hier ist fraglich, ob ein Mehrwert gegenüber der amtlichen Statistik, besteht, die den Ausstoß im Baugewerbe misst.

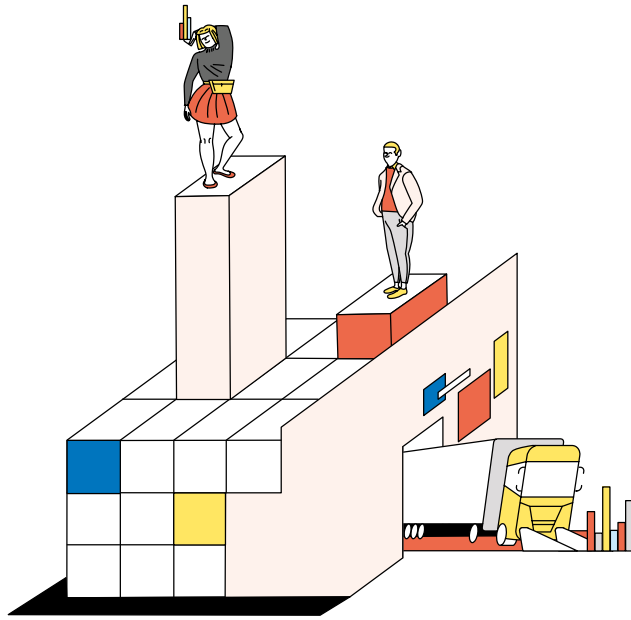
Mobilitätsdaten, die bei der Verwendung von Smartphones entstehen, können fein auflösende Standortdaten und damit Informationen zu Mobilität oder Pendlerbewegungen liefern oder Eingang in die regionale Bevölkerungsstatistik finden. Das Potenzial für die makroökonomische Analyse erscheint dagegen hier begrenzt. Eine Ausnahme stellte die Corona-Krise da, in der Bewegungsdaten (Google-Mobilitätsdaten) einen Anhaltspunkt für den Mobilitätsrückgang und damit auch für Nachfragerückgänge in Gastronomie und Einzelhandel lieferten. →

IN KÜRZE

Die Auswertung und Interpretation von Satellitenbildern ist komplex und technisch anspruchsvoll.

WAS BRINGT BIG DATA FÜR DIE MAKROÖKONOMISCHE ANALYSE UND PROGNOSE?

Traditionell ist die amtliche Statistik der Hauptlieferant an Daten für die makroökonomische Analyse und Prognose. Sowohl die Zielgrößen für Prognosen als auch die erklärenden Variablen werden in den meisten Fällen durch statistische Ämter zur Verfügung gestellt. Allerdings sind neben der bereits angesprochenen zeitlichen Verzögerung der gemeldeten Daten auch die Revisionsanfälligkeit und die Limitierung auf bekannte wirtschaftliche Sachverhalte Herausforderungen, mit denen die amtliche Statistik und im erweiterten Sinne auch die makroökonomische Analyse und Prognose umgehen müssen.



**BIG DATA KANN
WIRTSCHAFTS-
STATISTIKEN
DEUTLICH
ERWEITERN.**

DIE KOMBINATION VERSCHIEDENER DATENQUELLEN ALS „KÖNIGSWEG“.

Vor diesem Hintergrund kann Big Data die regelmäßig veröffentlichten Wirtschaftsstatistiken um umfassendere Informationen erweitern und sie so verbessern. Neben der möglicherweise höheren Geschwindigkeit und Zeitnähe, mit der Analysen aus Big Data zur Verfügung stehen, können sie die amtliche Statistik ggf. auch um neue Sachverhalte bereichern, so dass insgesamt ein runderes, vollständigeres Bild entsteht. Vor allem die Kombination verschiedener Datenquellen wie beispielsweise von Bewegungsdaten mit Transaktionsdaten kann noch präziseren Aufschluss über Größen wie etwa den Privaten Konsum oder andere wirtschaftliche Zielvariablen geben.

Die amtliche Statistik arbeitet daher längst am Thema Big Data und untersucht eine Reihe von alternativen Datenquellen, um deren Potenzial zu erkunden. So führt das Statistische Bundesamt unter der Rubrik „Experimentelle Daten“ neue, innovative Datenprojekte, die größtenteils unter den Begriff „Big Data“ fallen. Bekannte Beispiele umfassen etwa den bereits erwähnten Lkw-Maut-Fahrleistungsindex, der aus den Mauterhebungsdaten berechnet wird und frühzeitige Anhaltspunkte für die Industrieproduktion geben kann.

Ein limitierender Faktor für den Einsatz von Big Data sind allerdings die hohen Anforderungen an die Datenqualität, die für die amtliche Statistik üblich und auch für eine verlässliche makroökonomische Analyse und Prognose erforderlich sind. Darüber hinaus stellen auch Datenschutzanforderungen hohe Hürden für den Einsatz bestimmter Datenquellen, etwa wenn über die Kombination verschiedener Daten Rückschlüsse auf Einzelpersonen möglich werden sollten. Zusätzlich ist der Aspekt der Verstetigung sowohl für die Statistikämter als auch für die Nutzer dieser

IN KÜRZE

Die Datenqualität ist nicht immer so hoch wie gewünscht und erforderlich.

KONJUNKTURFORSCHUNG ERHÄLT DURCH MODERNE ANALYSE-METHODEN NEUE IMPULSE.

Daten von zentraler Bedeutung. Aufgrund der schnelllebigsten Natur vieler Big Data-Quellen stehen viele Datenreihen unvorhergesehen entweder in abgeänderter Form oder gar nicht mehr zur Verfügung, was insbesondere für die kontinuierliche Konjunkturbeobachtung und -prognose problematisch ist.

Dennoch ist insgesamt davon auszugehen, dass die Bedeutung von Big Data für die amtliche Statistik zukünftig weiter steigen wird. Die Datenmenge dürfte sich durch das stetige Hinzukommen neuer digitaler Anwendungen kontinuierlich erhöhen. Dabei erscheint es wahrscheinlich, dass die neuen Daten in vielen Bereichen komplementär zu den traditionellen, amtlichen Statistiken verwendet werden, diese jedoch nicht ersetzen können.

NEUE INDIKATOREN UND DERZEITIGE BMWI-PROJEKTE

Am BMWi läuft derzeit ein aus mehreren Teilprojekten bestehender Forschungsauftrag, der sich mit verschiedenen Aspekten der Big Data-Thematik befasst. Zunächst ging es darum, in einer Übersichtsstudie ein Kompendium der verschiedenen möglichen Daten und Analyseansätze zu erstellen. Ziel war es auch, Projektnehmer, weitere Wissenschaftler und Ministerien in Workshops zu vernetzen und so die Big Data-Forschungslandschaft in Deutschland insgesamt zu stärken.

Weitere Projekte beschäftigen sich mit unterschiedlichen volkswirtschaftlichen Kennzahlen. So werden in zwei Forschungsvorhaben die Prognosen der Investitionen bzw. der Preise durch Big Data untersucht. In einem weiteren Projekt werden die Anwendungsmöglichkeiten von Texten als Datenquelle für die Konjunkturanalyse in den Blick genommen.

Für die Investitionsprognose scheinen ersten Ergebnissen zufolge vor allem experimentelle Indikatoren wie der Lkw-Fahrleistungsindex oder der Umsatz an Zollstellen erfolgversprechend. Im Bereich der öffentlichen und privaten Bauinvestitionen könnten öffentliche Online-Ausschreibungen oder Daten für die Vermittlung von Handwerkerleistungen Anhaltspunkte für die zukünftige Entwicklung bieten.

Für die Preisprognose bieten sich interessante neue Möglichkeiten. So können mittels Auswertung von regionalen Liefer-Angeboten von Einzelhandels-

ketten Preisunterschiede zwischen Regionen oder Räumen (bspw. Stadt oder Land) analysiert werden. Das lässt wichtige Rückschlüsse auf die unterschiedlichen Lebenshaltungskosten in den verschiedenen geographischen Räumen zu. Außerdem wird daran gearbeitet, die Angebotsmieten aus Wohnungsportalen mit Daten aus amtlichen Befragungen wie dem Mikrozensus zu ergänzen, um weitere Rückschlüsse auf Bestandsmieten zu ziehen.

FAZIT

Stark datengetriebene Analysen können Expertenwissen nicht ersetzen. Vor allem Sonderfaktoren und qualitative, politische Maßnahmen können im Modell nur schwer erfasst werden. Während der Corona-Krise zeigte sich dies beispielsweise bei der Anordnung notwendiger Eindämmungsmaßnahmen, die ex-ante in Modellen nur schwer abzubilden sind. Umso mehr Bedeutung kommt jedoch der Interpretation der Daten und Ergebnisse zu.

Das BMWi lotet derzeit anhand eines Forschungsprojekts das Potenzial neuer Methoden für die Beobachtung, Analyse und Projektion der konjunkturellen Entwicklung in Deutschland aus. Mit dem Projekt soll auch der deutschen Konjunkturforschung ein Impuls gegeben werden, sich weiterhin intensiv mit den Themen Digitalisierung und Big Data zu beschäftigen. —

IN KÜRZE

Sonderfaktoren und politische Maßnahmen sind (in Modellen) meist nur schwer erfassbar.

MEHR ZUM THEMA

Institut für Weltwirtschaft (2021), „Big Data in der makroökonomischen Analyse“ in: Kieler Beiträge zur Wirtschaftspolitik Nr. 32

Experimentelle Daten des Statistischen Bundesamts: t1p.de/EXDAT

KONTAKT

DR. JIN-KYU JUNG & DR. CHRISTIAN WITTNEBEN
Referat: Beobachtung, Analyse und Projektion der gesamtwirtschaftlichen Entwicklung

schlaglichter@bmwi.bund.de