

# Research Brief

63. Ausgabe – Januar 2024



## Wie Haushaltsscannerdaten bei der Inflationsprognose helfen

von Günter Beck, Kai Carstensen, Jan-Oliver Menz, Richard Schnorrenberger und Elisabeth Wieland

**Die Prognose der Inflationsrate für den jeweils laufenden Monat (Nowcasting) ist für Zentralbanken und Marktteilnehmer von hoher Bedeutung, insbesondere in turbulenten Zeiten. In einer neuen Studie untersuchen Forscherinnen und Forscher, ob sich der Nowcast der monatlichen Inflationsrate in Deutschland mithilfe Millionen granularer, wöchentlicher Scannerdaten von privaten Haushalten und Techniken des Maschinellen Lernens (ML) verbessern lässt.**

Die großen ökonomischen Schocks der jüngeren Zeit, die durch die COVID-19-Pandemie und die russische Invasion in der Ukraine im Februar 2022 ausgelöst wurden, erhöhten den Bedarf an verlässlichen Echtzeitinformationen über die Entwicklung der Konjunktur und der Verbraucherpreise. Da amtliche makroökonomische Statistiken typischerweise nur mit zeitlicher Verzögerung und auf monatlicher oder niedrigerer Frequenz verfügbar sind, können nichttraditionelle, höherfrequente Daten wie Web-Scraping und Transaktionsdaten einen Mehrwert liefern (vgl. z.B. Deutsche Bundesbank, 2023). In unserer Studie (Beck, Carstensen, Menz, Schnorrenberger und Wieland, 2023) zeigen wir, dass wöchentliche Haushaltsscannerdaten in Kombination mit Maschinellen Lernen (ML) Zentralbanken und anderen Marktteilnehmern ein vielversprechendes Instrument bieten, um aktuelle und möglicherweise disruptive Entwicklungen in Echtzeit zu erkennen und in solchen Situationen bessere Entscheidungen zu treffen.

### Hohe Korrelation zwischen Preisindikatoren aus Haushaltsscannerdaten und amtlichen Preisindizes

Unser Datensatz basiert auf dem GfK-Haushaltspanel. Dieses umfasst für den Zeitraum 2003 bis 2022 tägliche Einkäufe von sogenannten schnellrotierenden Konsumgütern, also solchen, die regelmäßig gekauft und schnell verbraucht werden. Die Einkaufs-Aufzeichnung beinhaltet vorwiegend Nahrungsmittel und Verbrauchsgüter wie Shampoo oder Zahnpasta, die von den Paneteilnehmenden zuhause eingescannt werden und daher Haushaltsscannerdaten heißen. Im Durchschnitt umfasst das GfK-Haushaltspanel für Deutschland rund 30.000 Haushalte, 200.000 Produkte (gemessen auf Barcode-Ebene) und 30 Millionen Beobachtungen pro Jahr. Zudem beinhaltet der Datensatz eine detaillierte Produktbeschreibung sowie eine eigene Produktkategorisierung. Diese Informationen ermöglichen eine Zuordnung zur tiefsten Aggregationsstufe in der deutschen Verbraucherpreisstatistik, nämlich gemäß der Klassifikation der Verwendungszwecke des Individualverbrauchs (Classification of Individual Consumption by Purpose – COICOP) den sogenannten

COICOP-Zehnstellern, wie beispielsweise „Butter“, „Kaffeebohnen“ und „Zahnpasta“.

Insgesamt können wir die Haushaltsscannerdaten mehr als 180 Produktgruppen des deutschen harmonisierten Verbraucherpreisindex (HVPI) zuordnen, welche rund 12 Prozent des deutschen Warenkorb sowie die üblichen Geschäftstypen wie Supermärkte und Discounter abdecken. Hieraus leiten wir mittels gängiger Indexmethoden, die Statistikämter im Zusammenhang mit Scannerdaten häufig anwenden, Preisindikatoren ab (konkret: Time-Product-Dummy Regressionen, vgl. Eurostat, 2022). Wir zeigen, dass unsere Scannerdatenbasierten Preisindikatoren den amtlichen Preisindizes recht gut entsprechen (vgl. Abbildung 1). Phasenweise Abweichungen der Raten (wie während der jüngsten Phase hoher Inflation) lassen sich unter anderem damit erklären, dass tatsächliche Transaktionsdaten im Vergleich zur amtlichen Preisstatistik verstärkt Sonderangebote sowie Produktwechsel umfassen dürften.

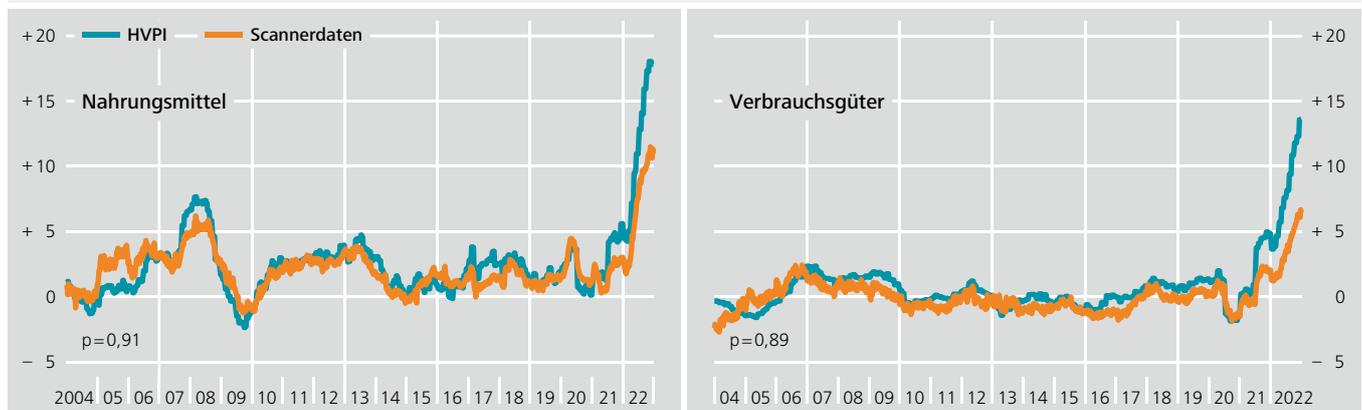
Monat, wenn nur Scannerdaten der ersten sieben Tage eines Monats vorliegen.

In einem zweiten Schritt betrachten wir die drei größeren Produktgruppen „unverarbeitete Nahrungsmittel“, „verarbeitete Nahrungsmittel“ und „nichtenergetische Industriegüter“ (darunter fallen z.B. Verbrauchsgüter für den Haushalt), deren Teuerungsraten häufig im Fokus der geldpolitischen Analyse stehen. Da diese Produktgruppen aus vielen einzelnen COICOP-Zehnstellern bestehen, ist die Verwendung des obigen MIDAS-Modell aufgrund der entsprechend großen Parameterzahl nicht mehr möglich. Daher verwenden wir diesmal ML-Schätzverfahren zur Dimensionsschrumpfung (siehe Babii, Ghysels und Striaukas, 2022), um den großen Satz an potenziellen Vorhersageindikatoren effizient in ein Zeitreihenmodell zu integrieren. Im Vergleich zum Benchmark-Modell sinkt der Prognosefehler um bis zu 25 Prozent. Auch hier verbessert unser ML-basierter Ansatz die Prognose bereits nach den ersten sieben Tagen eines Monats.

## Scannerdaten-basierte Preisindikatoren für Nahrung und Verbrauchsgüter im Vergleich zu den amtlichen Teuerungsraten<sup>\*)</sup>

Abbildung 1

Veränderung gegenüber Vorjahr in %, Wochenwerte



\* Teuerungsraten basierend auf COICOP-Zehnstellern, für welche Scannerdaten vorliegen. **p** Korrelation beider Reihen.  
Deutsche Bundesbank

## Haushaltsscannerdaten verbessern Nowcast auf verschiedenen Ebenen

Zunächst wenden wir für jeden der rund 180 COICOP-Zehnstellern ein Zeitreihenmodell für gemischte Frequenzen an (konkret: Unrestricted Mixed Data Sampling (U-MIDAS) Modell, vgl. Forni, Marcellino und Schumacher, 2015). Hierbei wird anhand des wöchentlichen Preisindikators die monatliche Teuerungsrate an den Tagen 7, 14, 21 und 28 eines Monats vorhergesagt. Wir zeigen, dass dieser Ansatz den Nowcast-Fehler im Vergleich zu einem gängigen univariaten Zeitreihenmodell erheblich reduziert. Der Nowcast der Teuerungsrate im laufenden Monat verbessert sich hierbei bereits früh im

Schließlich erstellen wir Nowcasts für die Gesamtinflation eines Monats. Hierzu betrachten wir sechs Unterkomponenten des deutschen HVPI separat: unverarbeitete Nahrungsmittel, verarbeitete Nahrungsmittel, Energie, Pauschalreisen, nichtenergetische Industriegüter und Dienstleistungen (ohne Pauschalreisen). Unser Indikatorenset erweitern wir neben den Scannerdaten zudem um wöchentliche Preisindikatoren für die beiden volatilen Komponenten Energie (aus dem Weekly Oil Bulletin der Europäischen Kommission) und Pauschalreisen (basierend auf tatsächlichen Buchungen über die AMADEUS-Plattformen, vgl. Henn et al., 2019). Konkret schätzen wir zum einen ein gemischtfrequentes ML-Modell,

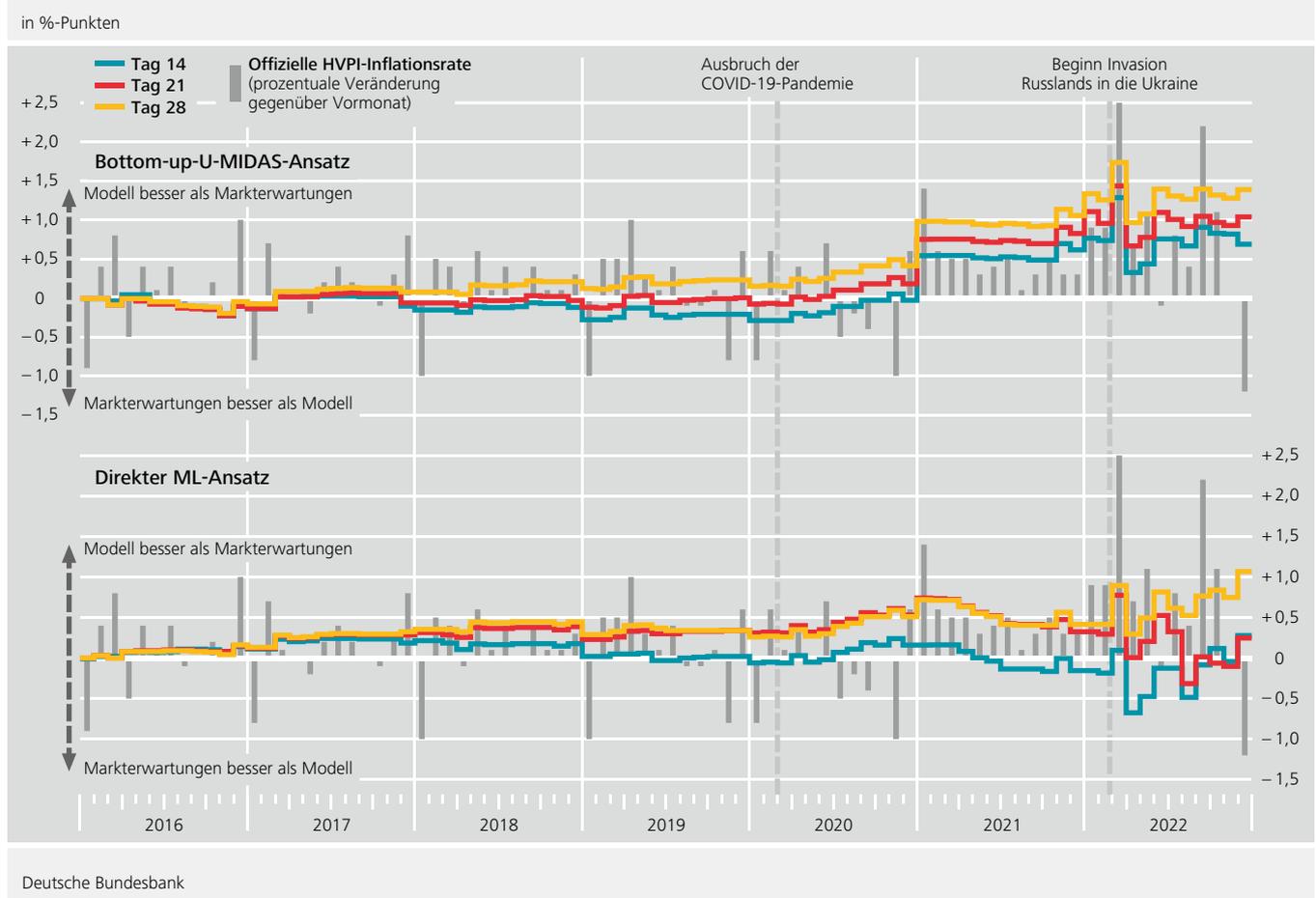
das für jede der sechs Komponenten die relevanten Vorhersageindikatoren direkt selektiert („Direct ML“), sowie zum anderen ein „Bottom-Up-U-MIDAS“ Modell für jeden COI-COP-Zehnsteller, deren entsprechende Nowcasts wir anhand des offiziellen HVPI-Wägungsschemas zunächst zu den Komponenten und dann zur Gesamtrate aggregieren.

Wir zeigen, dass beide Ansätze äußerst wettbewerbsfähige Nowcasting-Modelle liefern, die ähnlich wie oder sogar besser als die Markterwartungen gemäß der Bloomberg-Umfrage abschneiden (vgl. Abbildung 2). Die Markterwartungen liegen nur für die Gesamtinflationsrate vor und werden typi-

scherweise in der zweiten Monatshälfte abgefragt. Sie gelten in der Prognoseevaluation als schwer zu übertreffender Maßstab (vgl. Bańbura, Leiva-León and Menz, 2021). Hinsichtlich des gewählten Verfahrens zeigt sich, dass der direkte ML-Ansatz in normalen Zeiten, gemessen anhand einer relativ niedrigen Volatilität der Inflationsrate, schwer zu schlagen war, dafür aber in Zeiten höherer Volatilität schlechter als die Markterwartungen abschnitt. Dagegen führte der Bottom-up-Ansatz in turbulenten Zeiten zu einer Verbesserung des Inflationsnowcasts, insbesondere mit dem Auslaufen der temporären Mehrwertsteuersenkung im Januar 2021 und während des starken Inflationsanstiegs im Jahr 2022.

Kumulierter relativer Prognosefehler im Vergleich zu Markterwartungen

Abbildung 2



Quellen: GfK-Haushaltspanel; Weekly Oil Bulletin der Europäischen Kommission; AMADEUS; Bloomberg-Umfrage; eigene Berechnungen.

Hinweis: Die Abbildung zeigt auf der linken Achse die (invertierte) kumulative Summe der quadrierten Prognosefehlerdifferenz zwischen unserer Modellprognose (Bottom-up-U-MIDAS-Ansatz im oberen Panel und direkter ML-Ansatz im unteren Panel) und den Bloomberg-Markterwartungen für den Umfragezeitraum (Tage 14, 21 und 28). Die grauen Balken zeigen die amtliche Inflationsrate gegenüber dem Vormonat.

## Fazit

Wöchentliche Preisindikatoren aus Haushaltsscannerdaten können die Prognose der Inflationsrate des jeweils laufenden Monats („Nowcast“) auf verschiedenen Ebenen deutlich verbessern. Dies zeigt sich deutlich auf der Ebene wichtiger Teilkomponenten der deutschen Inflationsrate und gilt generell bereits nach den ersten sieben Tagen eines Monats. Unser Nowcasting-Ansatz für die Gesamtinflationsrate führt zu äußerst kompetitiven Modellen, die mit den Markterwartungen auf Augenhöhe sind oder diese sogar übertreffen. Wenn solche Daten künftig kontinuierlich verfügbar gemacht werden könnten, erscheint es daher vielversprechend, sie in der laufenden Konjunkturanalyse zu nutzen. Hinsichtlich der Methodenwahl zeigt sich, dass ML-Modelle in normalen Zeiten schwer zu übertreffen sind, aber nicht unbedingt schnell genug auf große Schocks reagieren. Insgesamt schneidet somit keine der beiden betrachteten Nowcasting-Methoden kontinuierlich besser als die andere ab. Vielmehr scheint die Nutzung von höherfrequenten Scannerdaten und deren sorgfältige Verarbeitung zu repräsentativen Preisindikatoren die Prognosefähigkeit im Vergleich zu Standard-Ansätzen zu verbessern.

## Referenzen

- Beck, G., K. Carstensen, J.-O. Menz, R. Schnorrenberger und E. Wieland (2023). Nowcasting Consumer Price Inflation Using High-Frequency Scanner Data: Evidence from Germany. Deutsche Bundesbank Discussion Paper No 34/2023.
- Babii, A., E. Ghysels und J. Striaukas (2022). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics* 40(3), S. 1094-1106.
- Bańbura, M., D. Leiva-León und J.-O. Menz (2021). Do Inflation-Expectations Improve Model-Based Inflation Forecasts? ECB Discussion Paper No 2604.
- Deutsche Bundesbank (2023). Modelle zur kurzfristigen Konjunkturprognose während der jüngsten Krisen, Monatsbericht, September 2023, S. 63-81.
- Eurostat (2022). Guide on Multilateral Methods in the Harmonised Index on Consumer Prices (HICP) - 2022 edition.
- Foroni, C., M. Marcellino und C. Schumacher (2015). Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials. *Journal of the Royal Statistical Society* 178(1), S. 57-82
- Henn, K., C.-G. Islam, P. Schwind und E. Wieland (2019). Measuring price dynamics of package holidays with transaction data. *Eurona* 2/2019, S. 95-132.



**Günter Beck**  
Professor für Volkswirtschaft  
an der Universität Siegen,  
Miggrprices



**Kai Carstensen**  
Professor für Ökonometrie an der  
Christian-Albrechts-Universität zu  
Kiel



**Jan-Oliver Menz**  
Ökonom  
im Zentralbereich Volkswirtschaft  
der Deutschen Bundesbank



**Richard Schnorrenberger**  
Doktorand und wissenschaftlicher  
Mitarbeiter am Institut für Statistik  
und Ökonometrie der Christian-  
Albrecht-Universität zu Kiel



**Elisabeth Wieland**  
Ökonomin  
im Zentralbereich Volkswirtschaft  
der Deutschen Bundesbank

## Neuigkeiten aus dem Forschungszentrum

### Veröffentlichungen

“Estimation of Heterogeneous Agent Models: A Likelihood Approach” von Juan Carlos Parra-Alvarez (University of Aarhus), Olaf Posch (University of Hamburg) und Mu-Chun Wang (Deutsche Bundesbank) wird im *Oxford Bulletin of Economics and Statistics* erscheinen.

“Indeterminacy and Imperfect Information” von Thomas A. Lubik (Federal Reserve Bank of Richmond), Christian Matthes (Indiana University) und Elmar Mertens (Deutsche Bundesbank) wird im *Review of Economic Dynamics* erscheinen.

### Veranstaltungen

7. – 8. Mai 2024

“Spring Conference 2024 – Structural Changes and the Implications for Inflation” gemeinsam organisiert von der Dänischen Nationalbank, der Norges Bank und der Deutschen Bundesbank

---

### Haftungsausschluss:

Die hier geäußerten Ansichten spiegeln nicht zwangsläufig die Meinung der Deutschen Bundesbank oder des Eurosystems wider.