

How household scanner data help improve inflation forecasting*

By Günter Beck (University of Siegen), Kai Carstensen (Kiel University),
Jan-Oliver Menz (Deutsche Bundesbank), Richard Schnorrenberger (Kiel University),
and Elisabeth Wieland (Deutsche Bundesbank)

Keywords: inflation nowcasting, machine learning, scanner price data, mixed-frequency modeling.
JEL codes: E31, C55, E37, C53.

Forecasting current month inflation (“nowcasting”) is an important exercise for central banks and market participants, especially in turbulent times. This policy brief illustrates how high-frequency scanner data from households combined with machine learning (ML) can improve the nowcast of monthly inflation in Germany.

Nowcasting during turbulent times

The recent major economic shocks induced by the COVID-19 pandemic and the Russian invasion of Ukraine in February 2022 increased demand for reliable real-time information about the state of the economy and consumer prices. Since official macroeconomic statistics are typically only available with a time lag and at monthly or less frequent intervals, non-traditional, higher-frequency data such as web scraping and transaction data can provide a value added (see, for example, Deutsche Bundesbank, 2023). In our study (Beck, Carstensen, Menz, Schnorrenberger and Wieland, 2024), we show that weekly household scanner data in combination with machine learning (ML) techniques provide central banks and other market participants with a promising toolkit to monitor ongoing and potentially disruptive developments in real time and to make better-informed decisions in such situations.

*Authors’ note: The views expressed are the authors’ personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem. This policy brief is based on [ECB Working Paper Series No. 2930](#).

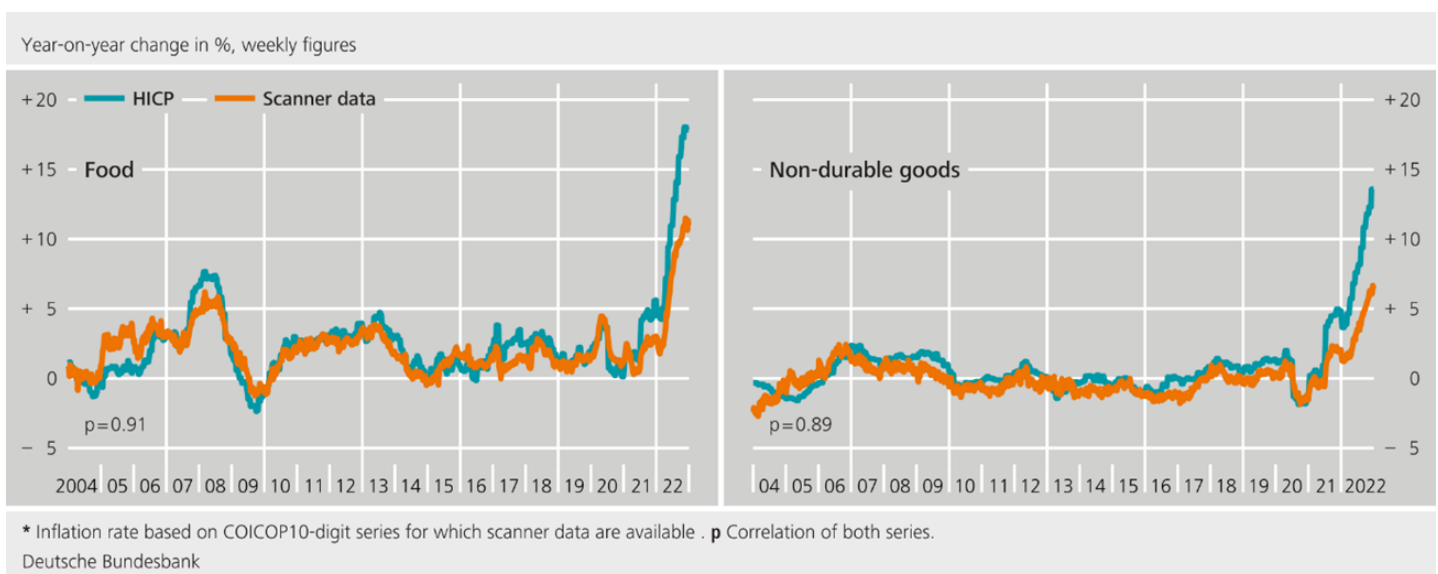
Price indices from household scanner data comove strongly with official counterparts

Our dataset comes from the household panel of the market research company GfK and contains daily purchases of fast-moving consumer goods, i.e. products that are bought regularly and consumed quickly, for the period from 2003 to 2022. The purchases covered are mainly food and non-durable goods such as shampoo or toothpaste, which are scanned by panel participants at home and therefore referred to as household scanner data. On average, the GfK household panel for Germany comprises around 30,000 households, 200,000 products (measured at the barcode level) and 30 million observations per year.

In addition, the dataset contains detailed product descriptions and has its own product classification system. These descriptions allow the data to be mapped to the most disaggregate level used in the German consumer price statistics, i.e. according to the classification of individual consumption by purpose at the ten-digit level (COICOP-10), such as “butter”, “coffee beans” and “toothpaste”.

Overall, we can map the household scanner data to more than 180 COICOP-10 items of the German Harmonised Index of Consumer Prices (HICP), which cover around 12% of the German basket of goods and typical outlet types such as supermarkets and discounters. From this, we derive price indices using common index methods often applied by statistical offices in connection with scanner data (specifically, time-product dummy regressions; see Eurostat, 2022). We show that our scanner data-based price indices match official price indices fairly closely (see Figure 1). Periodic differences between the indices (such as in the most recent period of high inflation) can be explained in part by the fact that, compared with the official price statistics, actual transaction data are likely to include more special offers and product substitutions.

Chart 1: Price indicators derived from scanner data for food and non-durable goods compared with official inflation rates



Household scanner data improve nowcasting at various levels

Our nowcasting exercise targets three hierarchy levels of inflation. First, we specify a time series model for mixed frequencies for each of the around 180 COICOP-10 items (specifically, the unrestricted mixed data sampling (U-MIDAS) model; see Foroni, Marcellino and Schumacher, 2015). Here, the weekly price indicator is used to predict the monthly inflation rate on days 7, 14, 21 and 28 of a month. We show that this approach reduces the nowcast error substantially relative to a univariate time series benchmark model. Current month inflation nowcasts improve early in the month, already after the first seven days of a month.

In a second step, we focus on the three major product groups “unprocessed food”, “processed food” and “non-energy industrial goods” (including, for example, consumer goods for the household). As these product groups consist of many individual COICOP-10 items, use of the above U-MIDAS model is no longer possible due to the correspondingly large number of parameters. We therefore resort to shrinkage estimators from the ML toolkit (see Babii, Ghysels and Striaukas, 2022) to efficiently integrate the large set of potential predictors into a time series model. Compared with the benchmark model, we achieve forecast error reductions of up to 25%. Again, our ML-based approach leads to nowcasting gains after just the first seven days of a month.

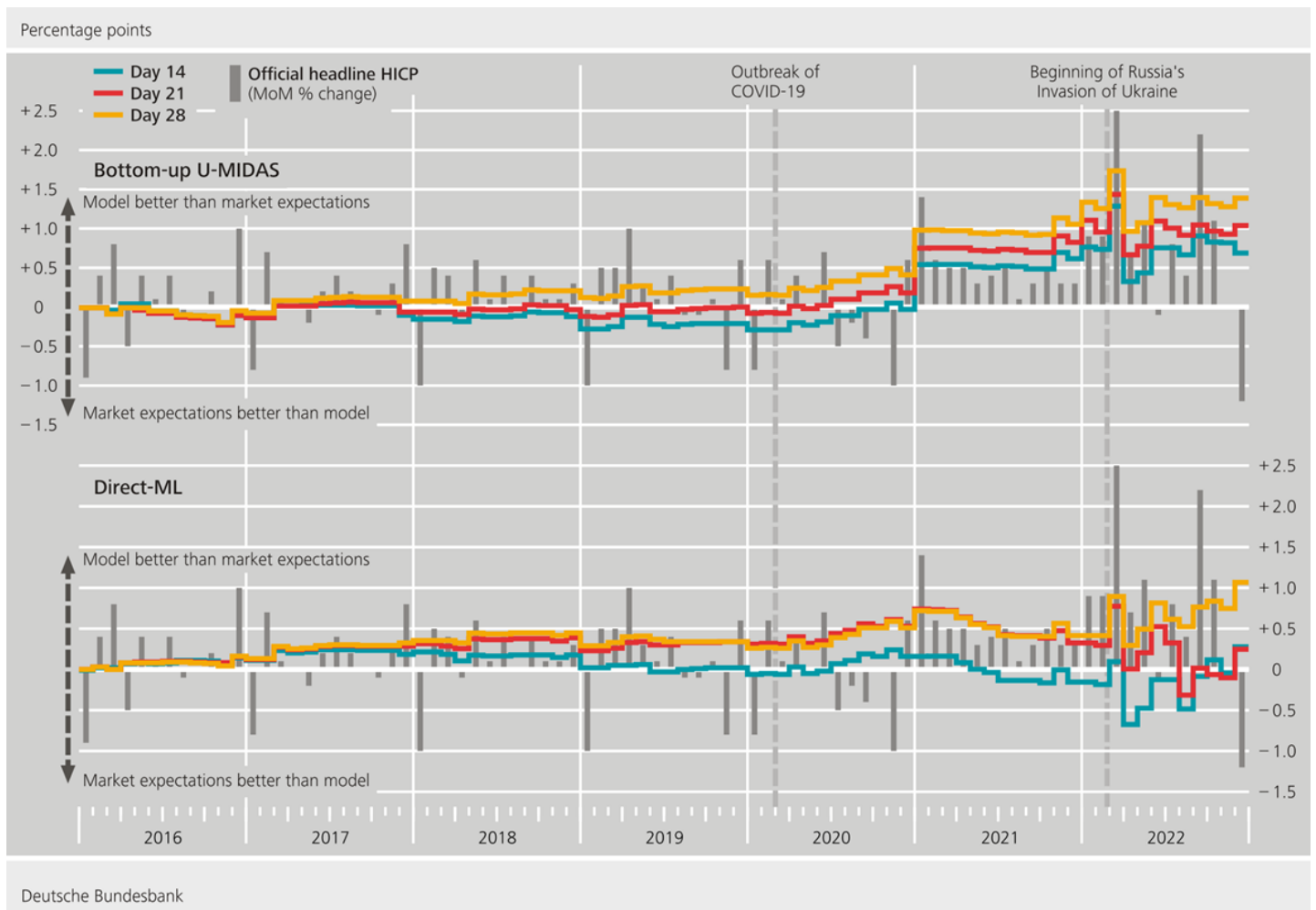
In the final step, we nowcast headline inflation. To this end, we consider six subcomponents of the German HICP separately: unprocessed food, processed food, energy, package holidays, non-energy industrial goods, and services (excluding package holidays). In addition to the scanner data, we expand our indicator set to include weekly price indicators for the two volatile components of energy (from the European Commission’s Weekly Oil Bulletin) and package holidays (based on actual bookings via the AMADEUS platforms, see Henn et al., 2019). Specifically, we estimate a mixed-frequency ML model that directly selects the relevant predictors for each of the six components (“direct ML”) and a bottom-up U-MIDAS model for each COICOP-10 item, the nowcasts of which we aggregate first to the components and then to headline inflation using the official HICP weighting scheme.

We show that both approaches produce highly competitive nowcasting models that are on par with, or even outperform, Bloomberg market expectations (see Figure 2). Market expectations are only available for headline inflation and are typically obtained in the second half of the month. When forecasting inflation, their predictions can be considered as a benchmark that is hard to beat (see Bańbura, Leiva-León and Menz, 2021). In terms of the approach chosen, it can be seen that the direct ML approach was superior in normal times, as measured by relatively low inflation volatility, but performed worse than market expectations in times of higher volatility. By contrast, the bottom-up approach improved inflation nowcasting in turbulent times, particularly with the expiry of the temporary VAT cut in January 2021 and during the sharp rise in inflation in 2022.

Concluding Remarks

In this policy brief, we have illustrated that weekly price indices from household scanner data can significantly improve monthly inflation nowcasts at various hierarchy levels of inflation. This is clearly evident at the level of major subcomponents of German inflation and is generally already the case after the first seven days of a month. Our nowcasting approach to headline inflation produces highly competitive models that are on par with, or even outperform, market expectations. In terms of the choice between bottom-up U-MIDAS models or direct ML models, the latter seem to be hard to beat in normal times, but do not necessarily adapt quickly enough to large shocks. As a result, neither of the two nowcasting methods analysed consistently outperforms the other. Rather, the use of higher-frequency scanner data and their careful transformation into representative price indices seems to improve nowcasting performance as compared to standard approaches.

In summary, our strategy to combine high-frequency data with flexible ML tools turned out to provide us with accurate inflation nowcasts at different levels of aggregation. Our approach thereby exploits the virtue of granular data to provide disaggregate and high-frequency real-time information underlying overall inflation.

Chart 2: Cumulative relative forecast error: models versus market expectations

References

- Beck, G., K. Carstensen, J.-O. Menz, R. Schnorrenberger and E. Wieland (2024). [Nowcasting Consumer Price Inflation Using High-Frequency Scanner Data: Evidence from Germany](#). ECB Working Paper Series No. 2930.
- Babii, A., E. Ghysels and J. Striaukas (2022). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics*, 40(3), pp. 1094-1106.
- Bañbura, M., D. Leiva-León and J.-O. Menz (2021). Do Inflation Expectations Improve Model-Based Inflation Forecasts? ECB Working Paper Series No 2604.
- Deutsche Bundesbank (2023). Models for Short-Term Economic Forecasting During the Recent Crises. *Monthly Report*, September 2023, pp. 61-78.
- Eurostat (2022). Guide on Multilateral Methods in the Harmonised Index on Consumer Prices (HICP) – 2022 edition.
- Forni, C., M. Marcellino and C. Schumacher (2015). Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials. *Journal of the Royal Statistical Society*, 178(1), pp. 57-82.
- Henn, K., C.-G. Islam, P. Schwind and E. Wieland (2019). Measuring Price Dynamics of Package Holidays with Transaction Data. *Eurona* 2/2019, pp. 95-132.

About the authors

Günter Beck is Professor for Macroeconomics at the University of Siegen and Research Director of Miggroprices. He studied Economics at the Universities of Bonn and Berkeley and received his PhD (Doctorate) in economics from the Goethe University Frankfurt. In his research, he primarily focusses on the use of micro price/scanner price data to gain new insights in the fields of macroeconomics, monetary economics and international economics. He has published in leading academic journals including the *Journal of Monetary Economics*, the *Journal of International Economics*, the *Journal of the European Economic Association* or the *Journal of Applied Econometrics*.

Kai Carstensen is Professor for Econometrics at Kiel University and External Research Professor at the Ifo Institute in Munich. From 2007 to 2014 he was Professor for Economics at the University of Munich and Head of the Business Cycle and Surveys Department of the Ifo Institute where he coordinated the business surveys on which the Ifo Business Climate indicator is based. From 2018 until 2024 he served as member of the European Statistical Governance Advisory Council (ESGAB). His research focuses on empirical macro, business cycle analysis and forecasting. He has published, inter alia, in the *International Journal of Forecasting*, the *Journal of Money, Credit and Banking*, and the *Journal of Business and Economic Statistics*.

Jan-Oliver Menz is a Senior Economist in the Macroeconomic Analysis and Projections Division of the Deutsche Bundesbank. He has worked in various expert groups within the European System of Central Banks on the drivers of (low) inflation, on the exchange rate pass-through, on inflation expectations (in the wake of the strategy review) and in the Price-setting Microdata Analysis Network (PRISMA). In the Bundesbank, he has contributed to establishing the Bundesbank Online Panel (BOP) used to measure inflation expectations of households and firms. He has published in peer-reviewed journals (*American Economic Journal: Macroeconomics*, *European Economic Review*) and holds a PhD in Economics from the University of Hamburg, Germany.

Richard Schnorrenberger is a PhD candidate and teaching assistant at the Institute for Statistics and Econometrics of the Kiel University. His research interests are mostly linked to time series econometric models for macro-finance forecasting, including Bayesian methods and mixed-frequency data.

Elisabeth Wieland is a Senior Economist in the Macroeconomic Analysis and Projections Division of the Deutsche Bundesbank. She has worked in various expert groups within the European System of Central Banks on digitalisation and the Price-setting Microdata Analysis Network (PRISMA). Her research focus is on micro prices, inflation measurement and analysis; she has published in peer-reviewed journals (*American Economic Journal: Macroeconomics*, *Macroeconomic Dynamics*) and holds a PhD in Economics from the University of Munich, Germany.

SUERF Publications

Find more **SUERF Policy Notes and Briefs** at www.suerf.org/publications/suerf-policy-notes-and-briefs/



SUERF is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy.

SUERF's events and publications provide a unique European network for the analysis and discussion of these and related issues.

SUERF Policy Briefs (SPBs) serve to promote SUERF Members' economic views and research findings as well as economic policy-oriented analyses. They address topical issues and propose solutions to current economic and financial challenges. SPBs serve to increase the international visibility of SUERF Members' analyses and research.

The views expressed are those of the author(s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.

All rights reserved.

Editorial Board

Ernest Gnan
David T. Llewellyn
Donato Masciandaro
Natacha Valla

SUERF Secretariat
c/o OeNB
Otto-Wagner-Platz 3
A-1090 Vienna, Austria
Phone: +43-1-40420-7206
www.suerf.org • suerf@oenb.at