DEUTSCHE
BUNDESBANK
EUROSYSTEM

Technical Paper

Forecasting HICP package holidays with forward-looking booking data

04/2024

Richard Schnorrenberger
(Universität Kiel)

Patrick Schwind
(Deutsche Bundesbank)

Elisabeth Wieland
(Deutsche Bundesbank)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

Non-technical summary

Research Question

Forecasting consumer prices for package holidays, which represent a major driver of the inflation rate in Germany, poses some practical challenges. With a substantial weight of currently $3\frac{1}{2}\%$ in the underlying consumer basket, the subcomponent "package holidays" in the German Harmonised Index of Consumer prices (HICP) exhibits strong seasonality, notable volatility, and several methodological breaks. This technical paper provides some guidance for forecasting the short-term dynamics of prices for package holidays and examines the predictive value of booking data enriched with forward-looking information.

Contribution

This paper assesses what is the best modelling strategy for the price dynamics of package holidays and whether it pays off to integrate forward-looking booking data into a forecasting model. We present two modelling strategies for predicting consumer prices for package holidays based on the seasonally adjusted and unadjusted target series. Moreover, we exploit the forward-looking dimension of high-frequency booking data to compile a price indicator that provides early signals about the underlying trend of the target series.

Results

Our analysis shows that the most accurate forecast is achieved with a modelling strategy that is tailored to the seasonally adjusted target series and incorporates information on the future seasonal component of the target series. Moreover, augmenting the forecasting model with the forward-looking price indicator based on booking data yields considerable gains that increase with the forecast horizon.

Nichttechnische Zusammenfassung

Fragestellung

Die Vorhersage der Verbraucherpreise für Pauschalreisen, die eine wichtige Triebkraft der Inflationsrate in Deutschland darstellen, ist mit einigen praktischen Herausforderungen verbunden. Mit einem beträchtlichen Anteil am zugrundeliegenden Warenkorb von derzeit $3\frac{1}{2}\%$ weist die Teilkomponente „Pauschalreisen" des deutschen Harmonisierten Verbraucherpreisindex (HVPI) eine starke Saisonalität, bemerkenswerte Volatilität und einige methodische Brüche auf. Dieses Technical Paper bietet Empfehlungen für die Prognose der kurzfristigen Dynamik der Preise für Pauschalreisen und untersucht die Prognosegüte von Buchungsdaten, die vorausschauende Informationen beinhalten.

Beitrag

In diesem Beitrag wird untersucht, welches die beste Modellierungsstrategie für die Preisdynamik von Pauschalreisen ist und ob es sich lohnt, vorausschauende Buchungsdaten in ein Prognosemodell zu integrieren. Wir stellen zwei Modellierungsstrategien zur Prognose der Verbraucherpreise für Pauschalreisen vor, die auf der saisonbereinigten und unbereinigten Zielreihe basieren. Zudem nutzen wir die vorausschauende Eigenschaft hochfrequenter Buchungsdaten, um einen Preisindikator zu erstellen, der frühzeitige Signale über den zugrundeliegenden Trend der Zielreihe liefert.

Ergebnisse

Unsere Analyse zeigt, dass eine möglichst exakte Prognose mit einer Modellierungsstrategie erreicht wird, die auf die saisonbereinigte Zielreihe zugeschnitten ist und Informationen zur künftigen saisonalen Komponente der Zielreihe einbindet. Darüber hinaus verbessert der vorausschauende Buchungsdaten-Preisindikator die Prognosegüte erheblich, insbesondere mit zunehmenden Prognosehorizont.

Forecasting HICP Package Holidays with Forward-Looking Booking Data*

Richard Schnorrenberger
Universität Kiel

Patrick Schwind
Deutsche Bundesbank

Elisabeth Wieland
Deutsche Bundesbank

August 21, 2024

Abstract

Forecasting consumer prices for package holidays, which represent a major driver of the inflation rate in Germany, poses some practical challenges. With a substantial share in the underlying consumer basket, prices for package holidays exhibit strong seasonality, notable volatility, and methodological breaks. We present two modelling strategies for predicting this volatile component based on the unadjusted price series and the seasonally adjusted series. Moreover, we exploit the forward-looking dimension of high-frequency booking data to compile a price indicator that provides early signals about the underlying trend of the target series. Our forecasting exercise shows that accurate forecasts are obtained with a modelling strategy tailored to the seasonally adjusted target series, alongside precise projections of the future seasonal component. Finally, augmenting the forecasting model with the forward-looking price indicator yields considerable gains that increase with the forecast horizon. Specifically, adding forward-looking information to the best-performing model increases the nowcast precision by 2.6% to 8% for short-term horizons of one to seven months, and the improvement exceeds 17% for longer horizons.

Keywords: Inflation forecasting, consumer prices, seasonality, travel booking data.

JEL codes: E31; E37; C22; C53.

*Corresponding author: Elisabeth Wieland, Deutsche Bundesbank, Mainzer Landstrasse 46, 60325 Frankfurt, Germany. Phone: +49 69 9566-37345. Email: elisabeth.wieland@bundesbank.de. We would like to thank Annette Fröhling, Christin Hartmann, Florian Kajuth, Thomas Knetsch, Jan-Oliver Menz, Daniel Ollech, and Karsten Webel (all Deutsche Bundesbank) for valuable comments and suggestions.

1 Introduction

When travelling, Germans prefer to book their hotel and flight in a package. This is reflected in a relatively high share of package holidays with currently $3\frac{1}{2}\%$ in the German Harmonised Index of Consumer Prices (HICP). Moreover, the underlying package holiday prices exhibit a strong seasonality and pronounced volatility, contributing significantly to the country's inflation dynamics. On top of that, methodological breaks in the index series pose challenges to forecasting HICP package holidays.

What is the best modelling strategy for prices of package holidays and does it pay off to integrate forward-looking booking data in the forecasting models? This technical paper provides some guidance for forecasting the short-term dynamics of HICP package holidays and examines the predictive value of booking data enriched with forward-looking information. Specifically, we conduct a pseudo-real-time forecasting exercise. We show that a modelling strategy designed for the seasonally adjusted target series is better suited to reliably anticipate the dynamics of the target series. Nevertheless, this approach must be accompanied by reliable extraction and projections of the target's seasonal component.

Moreover, we demonstrate that integrating booking data – which also enter the official price index – into the forecasting models leads to additional gains that increase with the forecast horizon. Our data stem from a travel booking system provider, featuring transaction-based prices for package holidays by travel destinations and important price determinants. In addition, the dataset encompasses a forward-looking dimension as it is organised by booking dates which are taking typically place much more in advance relative to the corresponding travel date, which corresponds to the reporting month of the HICP. Using these comprehensive features of booking data, we compile a forward-looking price indicator that closely matches the dynamics of the official HICP package holidays, particularly its underlying trend developments.

We document that incorporating our forward-looking price indicator into a forecasting model for the seasonally adjusted target yields predictive gains ranging from 2.6% for the nowcast up to 17%, on average, for longer horizons above eight months. Specifically, incorporating forward-looking information makes the model's forecasts almost one percentage point more aligned with the 7.85% average year-on-year change recorded for HICP package holidays in the evaluation period. This implies that the forward-looking price indicator provides early signals that help in spotting the underlying trend shifts of the target, especially over longer horizons. Ultimately, our model-based forecasts provide valuable support for the judgemental aspect of the forecasting process, aiding in distinguishing between transitory inflationary shocks and the persistent changes that define the long-term dynamics of inflation.

This technical paper is structured as follows. Section 2 highlights the challenges in forecasting the target variable in terms of seasonality and statistical breaks. In Section 3, we derive forward-looking price indicators for package holidays based on daily booking data and show that they co-move strongly with the official price index. Thus, our results allow a fairly precise preview at a substantial lead of the official price index. Section 4 outlines the modelling strategy and presents results of an out-of-sample forecasting exercise. Finally, Section 5 concludes.

2 Target series: HICP Package Holidays

In the HICP, prices for package holidays reflect a combination of flight and accommodation services, and are divided into domestic and international package holidays. In Germany, international package holidays correspond to more than 95% of the expenditures for package holidays. Following HICP convention, prices enter the index when the travel takes place.[1] As of January 2023, the majority of prices in the German series are derived from booking data (see Blasius, 2023, and Section 3 for more details on the computation).

Within the euro area, the price dynamics of package holidays in Germany are somewhat outstanding. Across countries, the German HICP-PACK exhibits a high volatility and a high expenditure share (see Figure A1 in the Appendix). Likewise, its price dynamics have been in the spotlight of monetary policy analysis.[2] Major statistical breaks coincide with changes in the underlying price collection, as shown in Figure 1 and listed in Table A1 in the Appendix. A significant method change was introduced in 2019 and came along with a one-off backward revision of the series until 2015 (Egner, 2019). The most recent methodological change was introduced in January 2023, when official price collection switched from using offer prices to prices from booking data (Blasius, 2023).

Price developments for package holidays are not only marked by a strong seasonality within a given year, but also by calendar effects across years. The lower panel of Figure 1 depicts the seasonal pattern by year. A common trend across years are higher prices during summer holidays (July/August) and lower prices during the winter season, except a small peak during the Christmas break in December. Nevertheless, differences between Easter (March/April) and Pentecost holidays (May/June) emerge in the period before the latest method switch in January 2023, which require a calendar adjustment during this period.[3] Moreover, we observe higher price levels over 2022 and 2023, indicating an upward trend movement after 2022.

Zooming into the drivers of prices for package holidays, a few major travel destinations explain most of their variation. With the switch to the base year 2020 of the national Consumer Price Index (CPI), the price index can be decomposed by several travel destinations, allowing for a more disaggregate, economic interpretation of price movements.[4]

[1] As stated in Eurostat, 2024, p. 28f: *"Prices for goods shall be included in the HICP for the month in which transactions can take place at that price, while the price of a service shall be included in the HICP for the month in which consumption of the service can commence."*

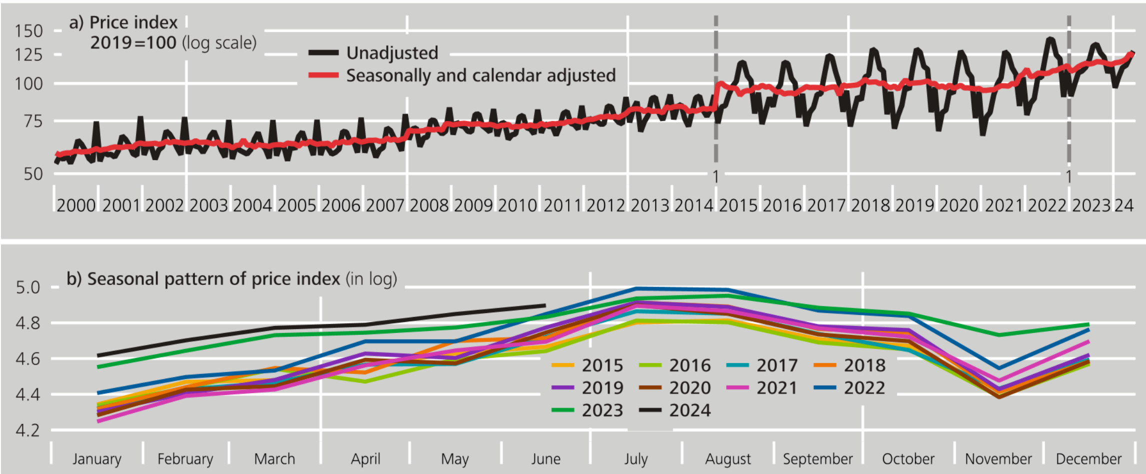
[2] See, for example, Deutsche Bundesbank (2017, 2019b,a, 2021, 2023) for Germany, as well as Eiglsperger (2019) and Dietrich, Eiglsperger, Mehrhoff, and Wieland (2021) for euro area inflation.

[3] Seasonal and calendar adjustment is performed at the Deutsche Bundesbank by means of JDemetra+, a publicly available, open source software recommended for the seasonal adjustment of official statistics in Europe. See `https://www.bundesbank.de/en/statistics/-/jdemetra--729580`. Likewise, seasonal and calendar factors for up to 48 months ahead are forecasted with JDemetra+. Note that the method switch in January 2023 reflects a structural change in seasonal volatility. Therefore, from January 2023 onwards, the seasonal adjustment procedure uses the break-free CPI series starting in 2020, as plotted in Figure 3, to extract and project the seasonal component, whereas the calendar factor is set to 100.

[4] The six subindices of the national CPI are: "0960200210 Package holidays, Canary islands", "0960200220 Package holidays, Balearic islands", "0960200230 Package holidays, Turkey", "0960200240 Package holidays, Greece", "0960200250 Package holidays, Egypt", and "'0960200300 Package holidays, other countries, cities or cruise". Note that an official decomposition by travel destinations is only available within the classification of the national CPI and as of 2020, whereas the HICP-PACK is based on the same database only as of 2023.

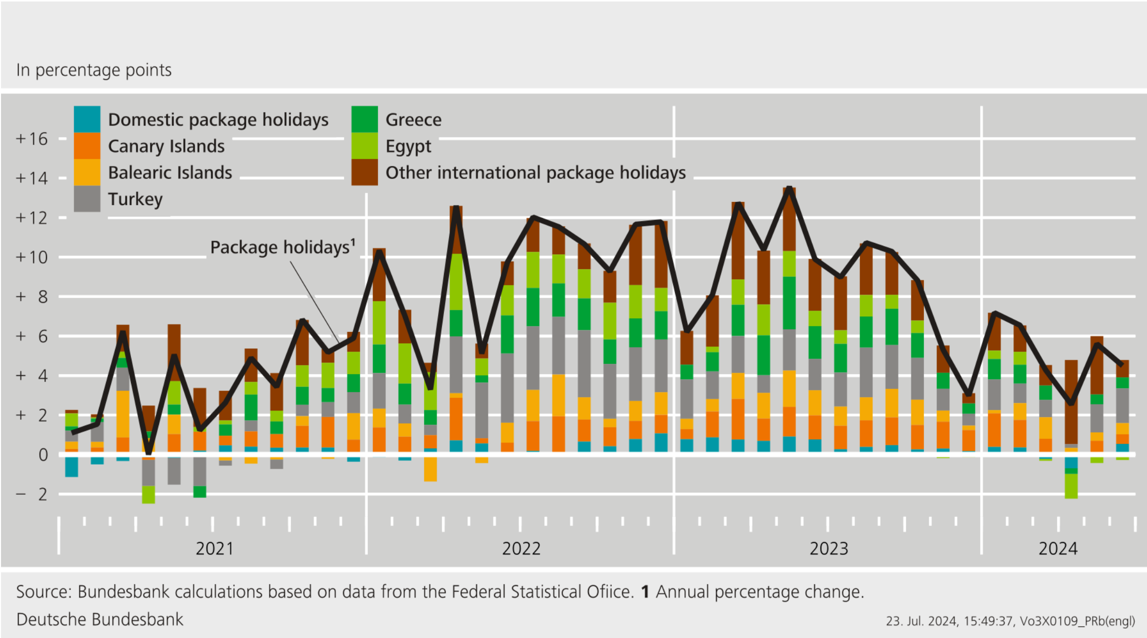
As shown in Figure 2, prices for package holidays to Spain (Canary and Balearic Islands), Turkey, Greece, and Egypt capture most of the movements in the overall price index. Following the travel bans during the Covid-19 pandemic, prices have caught up during the recent two years again, notably for Greece and Spain, but also for the residual aggregate "other package holidays", which comprise long-distance trips, city trips and cruises. During this period, prices were imputed by official price statistics, following the seasonal pattern of the previous year, as there was not enough information on prices for package holidays due to the restrictions on travel (cp. Table A1 in the Appendix).

Figure 1: HICP Package Holidays (ECOICOP 09.2)



Source: Eurostat and Bundesbank calculations. **1** Break in HICP methodology.
Deutsche Bundesbank
12. Jul. 2024, 09:01:23, Vo3X0033_PRb(engl)

Figure 2: Contribution of travel destinations to CPI Package Holidays (COICOP 09.6)



3 Deriving Forward-Looking Price Indicators from High-Frequency Booking Data

Given that prices for package holidays are recorded in the HICP when the travel takes place, real-time booking data provides a valuable source of information for future price developments. In retrieving forward-looking information, we rely on booking data which has been analysed by Henn, Islam, Schwind, and Wieland (2019) in the context of different price index methods for package holidays, and which enter the official HICP as of 2023 (Blasius, 2023).

Our booking data stems from the Amadeus IT Group, which operates an IT system for sales and marketing in the field of travelling. The dataset for Germany contains around 25 million transaction prices for flight package holidays of German travellers from 2015 to 2023. Transactions are recorded within the AMADEUS booking systems used by online travel portals as well as traditional "offline" high street travel agencies in Germany. For each transaction, information on important price determinants such as the holiday destination, accommodation type, and number of travellers is available.[5]

In deriving forward-looking price indicators for package holidays, we closely follow Henn et al. (2019) and Blasius (2023) by estimating a double imputation model. The main idea here is to control for quality differences in a given product (e.g. package holidays to Canary Islands) by estimating a price equation for a fixed base year and the current month. Bookings for the current month to a given holiday destination (e.g. Canary Islands) are then used to estimate prices for the base year as well as for the current month. This results in a pair of imputed prices, where the estimated price for the current month is related to the estimated price for the base year. Finally, the resulting price relations for every booking are aggregated via the geometric mean to derive the price relation for the regional subindex of package holidays.

Several price determinants in the booking data enter the equation of the double imputation model. Specifically, the log regression equation for a specific travel destination is given by:

$$\begin{aligned} \ln(totalPrice_{i,t}) = \beta_0 &+ \beta_1 \ln(travellerCount_{i,t}) + \beta_2 \ln(duration_{i,t}) \\ &+ \beta_3 \ln(bookTime_{i,t}) + \beta_4 \, D(online_{i,t}) + \beta_5 \, D(holiday_{i,t}) \\ &+ \beta_7 \, D(ownArrival_{i,t}) + \beta_7 \, D(childOne_{i,t}) + ... + \beta_9 \, D(childThree_{i,t}) \\ &+ \beta_{10} \, D(starOne_{i,t}) + ... + \beta_{14} \, D(starFive_{i,t}) + \varepsilon_t, \end{aligned} \tag{1}$$

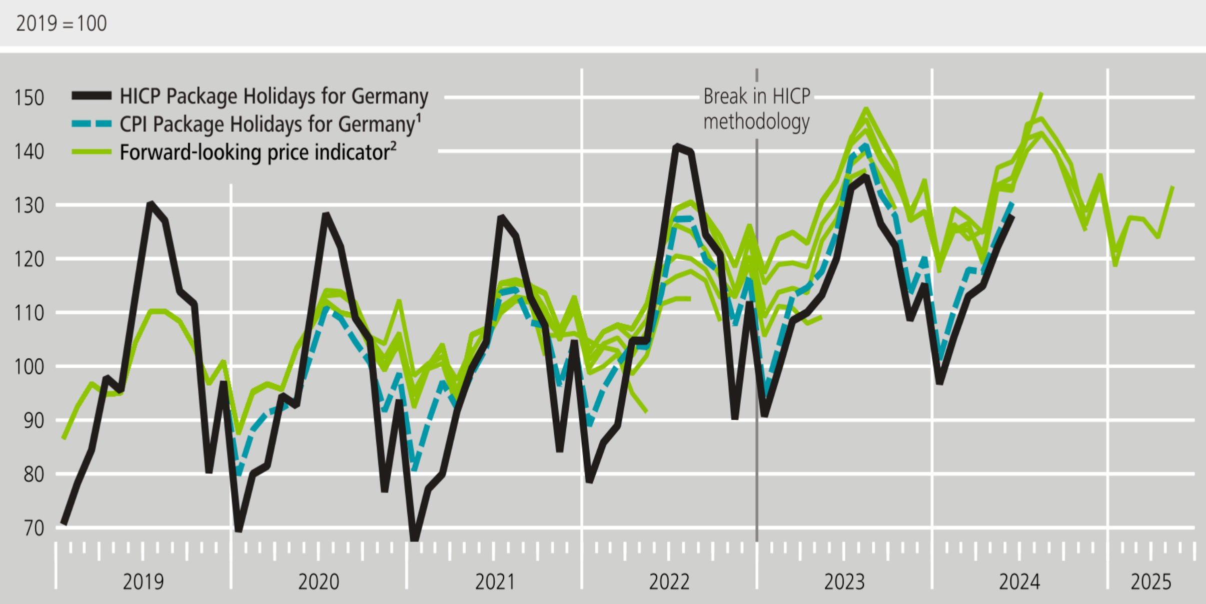
where the (log) total price for package holiday i is explained by the (log) number of travellers ("travelerCount", including children), the duration of the trip as well as the time distance between the transaction date and travel date in (log) days ("bookTime"). It also controls for online bookings, public school holidays, own arrival, whether one or up to three children join the trip as well as for the hotel category (one to five stars). Equation (1) is estimated for the (pre-pandemic) base year 2019 and for each individual

[5] See Henn et al. (2019) for a detailed description of the data set.

month from January 2016 to December 2023 separately.[6] Price indicators are derived for each travel destination (as defined by the official COICOP-10 items described in section 2, e.g. Canary Islands) separately and aggregated with the revenue share in our booking data to an overall price indicator for package holidays.

Finally, an important feature of our price indicator arises from its booking advance, providing early signals on the price developments of HICP-PACK up to one year ahead. As stated above, in the HICP, prices for package holidays are recorded when the travel takes place, but package holidays are typically booked well in advance. For Germany, on average, one-fifth of bookings have been made half a year prior to the month of travel, and half of the bookings three months or more in advance (see Henn et al., 2019). In a real-time setting, this leads to revisions of our price indicator due to incoming bookings. Therefore, to use our price indicator in an empirical forecasting exercise, we estimate Equation (1) recursively using expanding windows starting in January 2020.

Figure 3: HICP Package Holidays for Germany and real-time forward-looking price indicator



Sources: Eurostat, Amadeus Leisure IT GmbH and Bundesbank calculations. **1** According to base year 2020 method. **2** Based on AMADEUS booking data.
Deutsche Bundesbank
Vo3X0017A_PRb(engl)

Our resulting real-time price indicators for package holidays comove fairly well with its official CPI counterpart (see Figure 3). When comparing the latest vintage of our price indicator to the CPI "Package Holidays" (blue dashed line) – which is free of statistical

[6] Note that we deviate from the official double imputation method as documented by Blasius (2023) with respect to some aspects. First, within our dataset, we cannot differentiate further into the room category (e.g. cheap vs. luxurious). Second, prices for cruises in the official price index are still based on offer data, since the Amadeus data set is missing some important price-determining characteristics such as the cabin view. Nevertheless, we use the Amadeus bookings as a proxy for prices of cruises.

breaks as of 2020 – both series generally follow a similar pattern, with a correlation coefficient of 0.79 for the year-on-year rates. Nevertheless, when comparing our price indicator to HICP-PACK (black line), the correlation coefficient drops down to 0.37. The main difference between our price indicator and HICP-PACK lies in the volatility. Our price indicator exceeds the official price index during winter months, whereas it courses below the official index in the summer months. The differences are especially caused by the subindex "Package holidays in other countries, cities or cruises", which is harder to replicate from official information.[7] Nevertheless, our overall price indicator is capturing quite well the trend of the official price measures.

4 Forecasting Exercise

In this section, we investigate several modelling choices to obtain accurate predictions of the target series HICP-PACK. To cope with the strong seasonal pattern of the target variable, our modelling strategy follows two alternative econometric frameworks. First, we focus on the unadjusted price index and model seasonal effects directly via deterministic monthly dummies. Second, we model the seasonally adjusted price index produced at the Deutsche Bundesbank using the X13 approach via JDemetra+. From there, we augment both models with our forward-looking price indicator to assess the real-time predictive value of booking data for our target variable. To this end, we conduct a recursive forecasting competition among the two approaches mentioned above (Section 4.1.1 and 4.1.2); in addition, we benchmark them to the historical Bundesbank forecasts stemming from a macroeconomic error correction model (Section 4.1.3). Our forecasting exercise uses an expanding window with monthly data from January 2015 to June 2024.[8] For the out-of-sample evaluation, we focus on the period from July 2021 to the end of our sample. This choice ensures that forecasting results are not affected by the imputation procedure implemented by the statistical office during the COVID-19 lockdown periods.[9]

4.1 Modelling strategy

4.1.1 SD-AR and SA-AR Models

To set the stage for our prediction models, let us assume a multiplicative decomposition of our target series, y_t, that separates its seasonal component, S_t, from the remainder

[7] Across holiday destinations, the correlation of year-on-year rates of our disaggregate booking indicator is the lowest with $\rho = 0.29$ for "other countries, cities and cruises". In contrast, for the remaining five holiday regions, the comovement is strong, with a correlation coefficient ranging between 0.81 to 0.9. See Figure A2 in the Appendix.

[8] We take a stance on the initial point of the sample (January 2015) due to the structural break that introduces a more volatile seasonal pattern to HICP-PACK (see Table A1 in the Appendix for more details).

[9] For the periods 2020M4-2020M6 and 2020M9-2021M6, official price statistics imputed travel prices according to the previous year's month-on-month rate to preserve the seasonal pattern of HIPC-PACK. Since this imputation procedure was publicly known in advance at that time, we decided not to include months before 2021M7 in the evaluation sample.

components characterised by trend and irregular movements.[10] Within this time series structure, some econometric remedies can be adopted to project the future dynamics of a target variable marked by pronounced seasonal effects. Specifically, the seasonal component in y_t, can be treated as an integral part of the modelling process by incorporating seasonal dummies or using seasonal lag operators in the model. Alternatively, S_t can be addressed separately through seasonal adjustment procedures (see, e.g., Ghysels, Osborn, and Rodrigues, 2006; Ghysels and Marcellino, 2018).

Our first competing model essentially focuses on the unadjusted log series, $\ln(y_t)$, assuming an autoregressive (AR) structure augmented with monthly seasonal dummies and a fixed moving-average correction term:[11]

$$\phi(L)\ln(y_t) = \alpha_0 + \sum_{s=1}^{13}\gamma_s\, d_{s,t} + \frac{1}{2}\sum_{j=1}^{2}\hat{\varepsilon}_{t-j} + \varepsilon_t. \tag{2}$$

Therefore, the seasonal-dummy autoregressive (SD-AR) model presented in Equation (2) assumes deterministic seasonality and directly accounts for $\ln(S_t)$ through 11 monthly seasonal dummies $d_{1,t},\ldots,d_{11,t}$, complemented by an Easter dummy $d_{12,t}$ and a Pentecost dummy $d_{13,t}$. These seasonal dummies capture the specifics of the German public holiday season affecting the prices of package holidays.[12] Moreover, Equation (2) adjusts the forecast by the observed average error in the previous $J = 2$ months because time-invariant dummies cannot account for changes in the price level over time. This means that recent forecast errors convey information that provides additional lifts in forecasting performance. Finally, we specify a conventional autoregressive polynomial $\phi(L)$ with lags $\{1, 12\}$ to account for the temporal dependence in y_t.[13]

Next, we turn to the dynamics of the seasonally adjusted log series, $\ln(y_t/S_t)$; whereas the historical series of the seasonal component S_t is estimated beforehand via JDemetra+ using the HICP-PACK vintage available at the end of the sample, hence the pseudo-real-time aspect of our study.[14] This model captures the dynamics that happen beyond the regular seasonality, which follow a simple AR process in the first-difference:

$$\phi(L)\,\Delta\ln(y_t/S_t) = \alpha_0 + \varepsilon_t, \tag{3}$$

where $\phi(L)$ is specified with lags $\{1, 12\}$. Equation (3) is labeled as the seasonally adjusted autoregressive (SA-AR) model and produces forecasts for the underlying trend patterns

[10] For convenience, we assume that S_t comprises both the seasonal and calendar components of y_t. Concerning HICP-PACK, the calendar component is estimated to have no impact as of January 2023 (see footnote 3).

[11] See Beck, Carstensen, Menz, Schnorrenberger, and Wieland (2023) for a similar modelling approach of HICP-PACK in terms of a bottom-up inflation nowcasting exercise.

[12] The Easter dummy measures how many days of the two Easter weeks are in March and April, while the Pentecost dummy measures how many of the three Pentecost days (Saturday to Monday) are in May and June.

[13] Throughout the forecasting analysis, we choose AR lags in $\phi(L)$ based on out-of-sample performance.

[14] It is worth emphasising that real-time vintages of the seasonal component are not available. Nevertheless, short-run projections for the seasonal component do not suffer significant changes as new data becomes available. Thereby, we assume that the latest estimate of S_t based on data vintage of June 2024 provides a good proxy for the real-time estimates of S_t throughout the evaluation period.

of the target variable.[15] In a second step, we construct the forecast y_{t+h} of the unadjusted target as follows. The seasonal treatment of y_t using JDemetra+ also delivers forecasts for S_{t+h}. This predicted seasonal component is then added back to the forecast produced by the SA-AR model.

4.1.2 Supplementing the models with our forward-looking price indicator

To assess the real-time predictive content of booking data, we incorporate our aggregate forward-looking price indicator, x_t, based on advance bookings at the forecast horizon $t + h$ (see Section 3) as an explanatory variable in the model. Hence, we exploit the booking advance, which associates bookings with travel dates up to 10 months into the future. For the nowcast horizon, we take advantage of the high-frequency feature of the booking dataset and thus incorporate the forward-looking price indicator compiled with information up to day 21 of the reporting month.

First, we augment the SD-AR model with the log values of the forward-looking price indicator, hereafter labelled as the SD-ARX model and given by:

$$\phi(L)\ln(y_t) = \alpha_0 + \beta\ln(x_t) + \sum_{s=1}^{13}\gamma_s\, d_{s,t} + \frac{1}{2}\sum_{j=1}^{2}\hat{\varepsilon}_{t-j} + \varepsilon_t. \tag{4}$$

Note that within model (4), the forward-looking price indicator might help capture seasonal patterns beyond the regular seasonalities determined via the dummies $d_{s,t}$.

Second, we feed the 3-month moving average of monthly log-differences of the forward-looking price indicator, alongside its lagged series at orders $\{1, 5\}$, into a SARIMA model for the seasonally adjusted log series:[16]

$$\phi(L)\,\Delta_{s=12}\ln(y_t/S_t) = \sum_{p\in\{0,1,5\}}\beta_{p+1}\left(\frac{1}{3}\sum_{j=0}^{2}\Delta\ln(x_{t-j-p})\right) + \psi\,\varepsilon_{t-12} + \varepsilon_t. \tag{5}$$

We label this model as SA-SARIMAX, which is implemented with a seasonal lag operator $s = 12$ for the dependent variable and autoregressive lags $\{1, 4\}$. Note that SA-SARIMAX implies a year-on-year transformation to a seasonally adjusted target; nevertheless, in our application, it proved superior predictive performance compared to the SA-AR framework when incorporating information from our forward-looking price indicator.

[15] Note that first-differences in the SA-AR equation account for the non-stationary behavior of $\ln(y_t/S_t)$ stemming from the upward trend component, especially after 2022 (cp. Figure 1a).

[16] We also investigated the predictive power of other past lags and moving average filters of the forward-looking price indicator. However, the chosen contemporaneous and lagged series in Equation (5) provided the most robust signal for the target in various out-of-sample scenarios. Moreover, we also tested the use of the seasonally adjusted forward-looking price indicator, which did not improve the forecasting performance in comparison to the model in Equation (5). Here, the seasonality of the forward-looking price indicator is attenuated to some extent by using the 3-month moving average. Note that calendar effects resulting from shifts in public school holidays are already taken into account in calculating the price indicator (see Equation (1) in Section 3).

4.1.3 Macroeconomic benchmark model

We compare the four models presented before against historical Bundesbank forecasts within the Eurosystem's Narrow Inflation Projection Exercise (NIPE), hereafter BBK-NIPE.[17] These forecasts are primarily model-based but incorporate some degree of informed judgement, and updates occur every quarter.[18] The BBK-NIPE model follows an error-correction framework fitted to the seasonally adjusted log series, $\ln(y_t/S_t)$. A cointegration relationship is assumed between the target and a set of macroeconomic variables (oil prices, disposable income, the unemployment rate, and the USD/EUR exchange rate).

4.2 Out-of-sample results

4.2.1 Forecasting performance

In Table 1, we document the Root Mean Squared Error (RMSE) of the monthly forecasts up to 10 months ahead (including the nowcast horizon, $h = 0$) produced by the competing models for the evaluation period spanning from July 2021 to June 2024. The results are reported relative to the BBK-NIPE benchmark, where a value below 1 means that the forecast is doing better than the benchmark.[19] RMSE values are based on year-on-year inflation rates. For readability, each cell in Table 1 is coloured in a heatmap style, with darker colours indicating improved forecasting performance compared to the benchmark. To test whether forecasts generated by the competing models statistically outperform the benchmark, the results of the Diebold and Mariano (1995) test are also reported.

The results underscore the superior performance of modelling strategies designed for the seasonally adjusted target series across all forecast horizons. This is evident from the strong outperformance of SA-AR and SA-SARIMAX, contrasting with the less favourable outcomes attributed to SD-AR and SD-ARX when compared to the benchmark.[20] The outperformance of SA-AR and SA-SARIMAX forecasts is consistent across forecast horizons but is particularly pronounced in shorter horizons. Notably, these nowcasts exhibit a significant advantage of about 55% relative to the benchmark. In very short horizons up to $h = 3$, forecasting gains slightly decay to levels around 43% and 45% for SA-AR and SA-SARIMAX respectively, though retaining statistical significance – mostly at the 1% level – compared to the benchmark. Subsequently, these improvements stabilise at significant average gains of 35% to 43% for $4 \leq h \leq 7$, and approximately 33% to 45% for $h \geq 8$.[21]

[17] See ECB (2016) for a description of the NIPE. In particular, NCB staff provides four times a year short-term forecasts over a horizon of 11 months for overall HICP inflation and some key components, which are aggregated by the ECB to obtain euro area inflation projections.

[18] Hereby we use the latest BBK-NIPE projections to complete the missing monthly forecasts.

[19] The corresponding absolute RMSE values are presented in Table A2 of the Appendix.

[20] We also implement SARIMA-type models to the unadjusted target series to account for potential stochastic seasonalities. Similarly, they show significant underperformance compared to the benchmark across most horizons considered. These results are available upon request.

[21] To account for the missing real-time vintages of the seasonal component historically produced at the Deutsche Bundesbank, we implement a recursive real-time seasonal adjustment using default X13 settings in JDemetra+. Using our real-time projections of the seasonal component within the SA-SARIMAX framework, we consistently keep the edge relative to the benchmark up $h \geq 6$. Nevertheless, predictive

Table 1: RMSE for HICP-PACK: Competing models relative to BBK-NIPE benchmark

	SD-AR	SD-ARX	SA-AR	SA-SARIMAX
$h = 0$	1.209	1.05	0.453^{**}	0.441^{**}
$h = 1$	1.216	1.108	0.554^{**}	0.531^{**}
$h = 2$	1.593	1.325	0.567^{*}	0.546^{**}
$h = 3$	1.872	1.45	0.582^{*}	0.552^{*}
$h = 4$	1.804	1.421	0.627^{*}	0.576^{**}
$h = 5$	1.791	1.489	0.655^{**}	0.58^{**}
$h = 6$	1.671	1.422	0.654^{**}	0.569^{**}
$h = 7$	1.585	1.441	0.634^{*}	0.545^{**}
$h = 8$	1.557	1.42	0.683^{*}	0.565^{**}
$h = 9$	1.389	1.291	0.7	0.572
$h = 10$	1.171	1.119	0.622	0.518

Note: The table shows the heatmap of RMSE values for the competing models relative to the BBK-NIPE benchmark at forecast horizons $h = 0, 1, \ldots, 10$ for the period 2021M7-2024M6. RMSE values are based on year-on-year rates of the HICP-PACK. Results for the Diebold and Mariano (1995) test in case of outperformance relative to the benchmark are indicated by the symbols * (5% level) and ** (1% level).

Turning to the value added by booking data, feeding a smoothed set of series derived from the forward-looking price indicator into the predictor set of our SA-SARIMAX approach always leads to additional lifts in performance relative to its pure AR counterpart, the SA-AR. While predictive gains at very short horizons are relatively marginal ranging from 2.6% for the nowcast up to 5% for 3 months ahead, the contribution of our forward-looking price indicator substantially increases at longer horizons. Specifically, the improvement in forecasting accuracy is above 8% for $h \geq 4$ and reaches an average peak of 17% for $h \geq 8$. Based on the absolute RMSE (see Table A2 in the Appendix), this translates into an average higher accuracy of 92 basis points for tracking the year-on-year dynamics of HIPC-PACK over longer horizons. More precisely, we expect the model supplemented with forward-looking information to be almost one percentage point closer to the average 7.85% year-on-year change of HICP-PACK during the evaluation period. Thereby, the forward-looking price indicator indeed carries relevant real-time predictive content on the underlying trend dynamics of the target. These are particularly helpful in spotting trend shifts at longer horizons.

It is also worth noting that predictive gains of our forward-looking price indicator are more pronounced when comparing SD-ARX against its pure AR counterpart at shorter horizons. These contributions average 15% for $0 \leq h \leq 3$, although their forecast precision for the target variable still significantly lags behind that of SA-type models. Since the SD-ARX framework focuses on the unadjusted target series, these outcomes likely stem from the fact that booking data helps to identify future seasonal patterns beyond those captured by the seasonal dummies, thereby enhancing the model's implied projections for the seasonal component.

gains at the nowcast horizon here reach the 35% level while an average of 18.5% prevails for $1 \leq h \leq 6$. For more details, the complete results can be provided upon request.

4.2.2 Tracking prediction accuracy over time

Finally, we assess whether our best-performing model generates projections that sufficiently track the dynamics of HICP-PACK since 2021. Figure 4a illustrates the month-on-month predicted rates along with the observed target. Notably, SA-AR and SA-SARIMAX forecasts (depicted in purple and green) closely follow the strong seasonal trajectory of HICP-PACK, which might vary from -20% to +20% on a month-to-month basis. Although forecast precision marginally diminishes with the horizon, these projections consistently align with the seasonal movements. Moreover, it becomes evident that models based on seasonal dummies (red and yellow lines related to SD-AR and SD-ARX) struggle to anticipate the target dynamics during the Easter holiday season, and to a lesser extent, Pentecost. On the other hand, these calendar effects do not impact considerably the predictive performance of modelling strategies that rely on a seasonal adjustment procedure.

It is also worth zooming into the poor performance of the seasonal-dummy approaches after the methodological change in HICP-PACK of 2023, leading to much lower seasonal volatility compared to past years (see Figure 3). These unstable seasonal movements disorient estimates for monthly seasonal dummies, translating into imprecise forecasts of SD-AR and SD-ARX as of 2023.

Figure 4b depicts the implied year-on-year projections constructed with the same competing models. It exhibits a clear outperformance of SA-AR and SA-SARIMAX. In particular, these models clearly fare well over the second half of the evaluation period. The effectiveness of the SA modelling strategy is not only restricted to the nowcasting scenario – where projections are spot-on with the target in numerous instances – but such high precision is extended to longer horizons. Therefore, it is a modelling strategy designed for the seasonally adjusted target series, complemented by the forward-looking nature of the booking dataset with respect to trend shifts, that enables us to anticipate the dynamics in HICP-PACK with reasonable accuracy when faced with methodological changes.

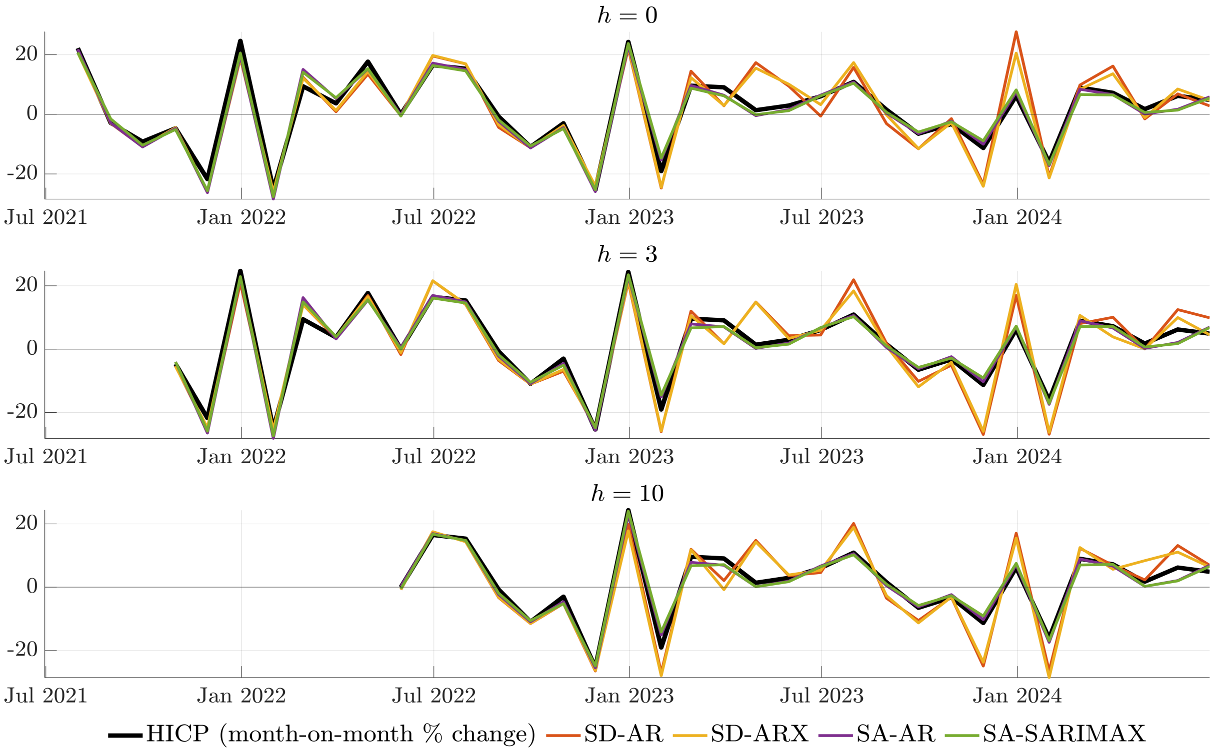
5 Conclusion

This technical paper highlighted some challenges when forecasting prices for package holidays in Germany, which carry a substantial weight in the country's inflation dynamics. These challenges are mostly associated with the series' strong seasonality, notable volatility, and methodological breaks. Within a pseudo real-time forecasting exercise, we provide a comprehensive analysis of forecasting strategies that integrate booking data as a forward-looking price indicator, which has shown to offer considerable predictive value for capturing the short-term dynamics of package holiday prices. The results underscore the effectiveness of a modelling strategy tailored to the seasonally adjusted series, which, when augmented with our forward-looking price indicator, significantly enhances forecast accuracy, especially for longer forecast horizons.

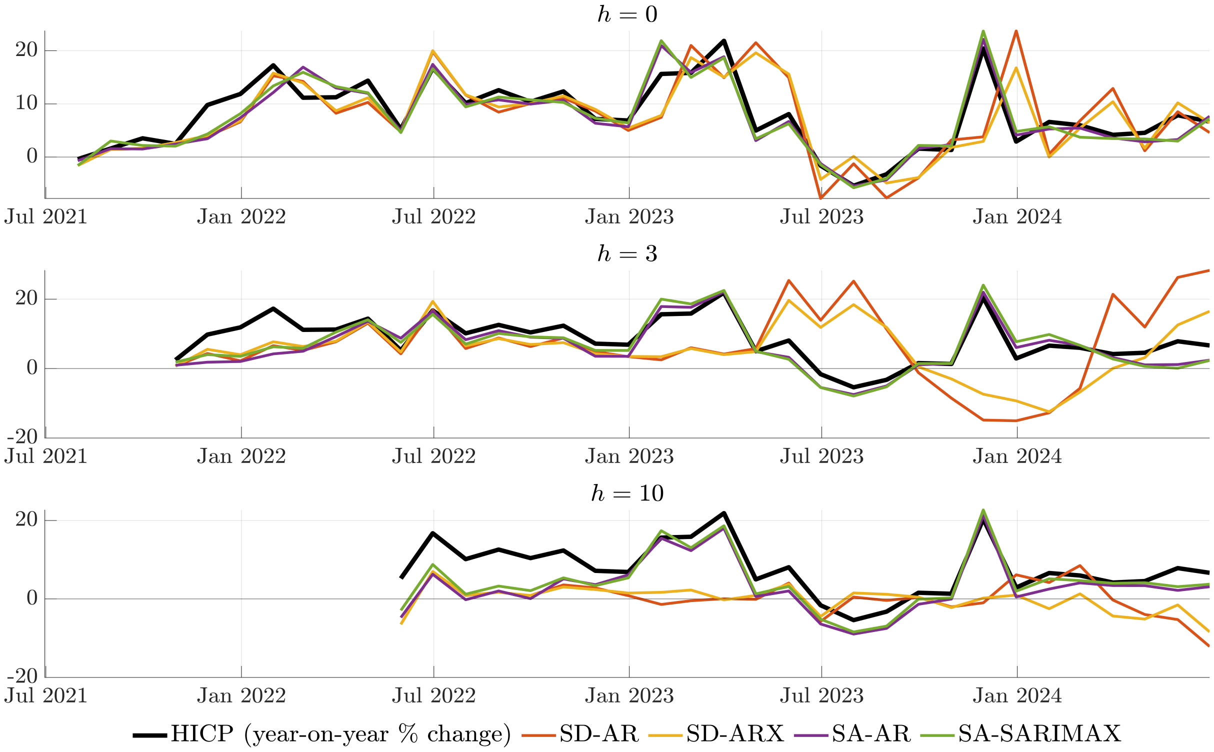
In summary, the forward-looking nature of travel booking data offers early signals of the underlying price trend, contributing to a more robust forecasting process that centres on the seasonally adjusted target series. This analysis showcases the potential of digital data sources in improving economic forecasts. Our findings encourage further exploration of digital data sources and advanced methods to refine the adaptability of forecasting models to strong seasonal volatility and methodological changes like those experienced by HICP-PACK.

Figure 4: Realised HICP-PACK and projections at horizons $h = 0, 3, 10$

(a) Month-on-month projections



(b) Year-on-year projections



References

Beck, G., K. Carstensen, J.-O. Menz, R. Schnorrenberger, and E. Wieland (2023). Nowcasting consumer price inflation using high-frequency scanner data: Evidence from Germany. Discussion Paper, Deutsche Bundesbank, No. 34/2023.

Blasius, A. (2023). Digitalisierung in der Preisstatistik – Nutzung von Reisebuchungsdaten. *Wirtschaft und Statistik 4/2023*, 34–45.

Destatis (2022). Impact of the corona crisis on the CPI/HICP price collection. 13 July 2022.

Deutsche Bundesbank (2017). The Volatility of the Traditional Core Inflation Rate in Germany. *Bundesbank Monthly Report November 2017*, 49–51.

Deutsche Bundesbank (2019a). Dampening special effect in the HICP in July 2019. *Bundesbank Monthly Report*, 57–59. August 2019.

Deutsche Bundesbank (2019b). The Revision of the Sub-Index for Package Holidays and its Impact on the HICP and Core Inflation. *Bundesbank Monthly Report*, 8–9. March 2019.

Deutsche Bundesbank (2021). One-off effects relating to COVID-19 in the HICP in 2021. *Bundesbank Monthly Report February 2021*, 63–66.

Deutsche Bundesbank (2023). Impact of the new HICP/CPI weights on inflation developments in Germany. *Bundesbank Monthly Report March 2023*, 8–11.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Dietrich, A., M. Eiglsperger, J. Mehrhoff, and E. Wieland (2021). Chain linking over December and methodological changes in the HICP: view from a central bank perspective.

ECB (2016). A guide to the Eurosystem/ECB staff macroeconomic projection exercises.

Egner, U. (2013). Verbraucherpreisstatistik auf neuer Basis 2010. *Statistisches Bundesamt, Wirtschaft und Statistik*, 329–344. May 2013.

Egner, U. (2019). Verbraucherpreisstatistik auf neuer Basis 2015. *Statistisches Bundesamt, Wirtschaft und Statistik*, 86–106. May 2019.

Eiglsperger, M. (2019). A new method for the package holiday price index in germany and its impact on hicp inflation rates. *Economic Bulletin Box 2*.

Eurostat (2024). Harmonised Index of Consumer Prices (HICP) - Methodological Manual - 2024 edition.

Ghysels, E. and M. Marcellino (2018). *Applied Economic Forecasting Using Time Series Methods*. Oxford University Press.

Ghysels, E., D. R. Osborn, and P. M. Rodrigues (2006). Forecasting Seasonal Time Series, Chapter 13. Volume 1 of *Handbook of Economic Forecasting*, pp. 659–711. Elsevier.

Henn, K., C.-G. Islam, P. Schwind, and E. Wieland (2019). Measuring Price Dynamics of Package Holidays with Transaction Data. *EURONA 2/2019*, 95–132.

Appendix

Figure A1: Standard deviation of growth rates and expenditure weight for HICP Package Holidays (ECOICOP 09.6) by country

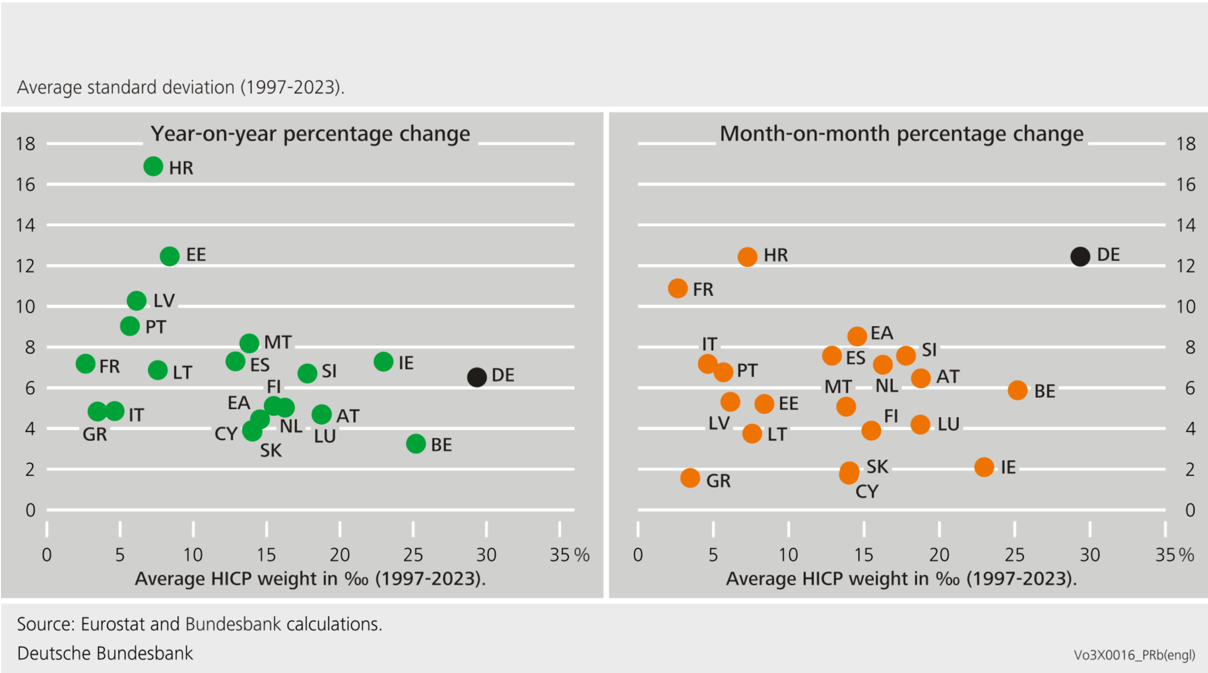


Figure A2: CPI Package holidays and own price indicator by travel destination

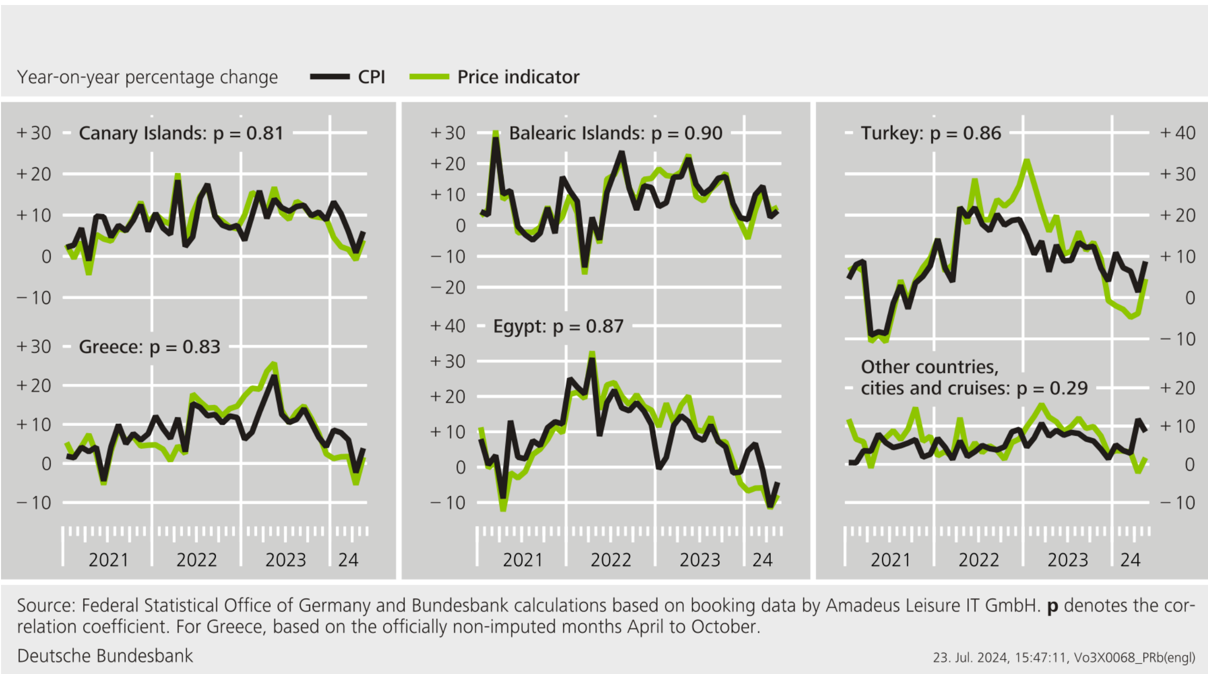


Table A1: Overview of major structural breaks in the German HICP Package Holidays

Date	Index date	Description
2013M1	2013M1	Methodological change following the introduction of the CPI base year 2010 (see Egner, 2013).
2019M2	2015M1-2022M12	Methodological change following the introduction of the CPI base year 2015 to better reflect the seasonality of package holiday prices throughout the year. By chain-linking both methods through the December 2014 HICP value, this led to distortions of annual rates of changes in 2015 (see Deutsche Bundesbank, 2019b).
2020M4-2021M6	2020M4-2021M6	Travel restrictions due to the Covid-19 pandemic led to imputed prices for most travel destinations of package holidays. Specifically, prices were carried forward with the month-on-month rate of change of the previous year, to sustain the seasonal pattern (see Destatis, 2022).
2023M1	2023M1 - today	Methodological change following the introduction of the CPI base year 2020. Price collection switched from offer prices to booking data (see Blasius, 2023).

Table A2: Absolute RMSE of competing models for HICP-PACK

	SD-AR	SD-ARX	SA-AR	SA-SARIMAX
$h = 0$	6.569	5.703	2.462^{**}	2.398^{**}
$h = 1$	7.054	6.427	3.214^{**}	3.077^{**}
$h = 2$	10.815	9.001	3.847^{*}	3.706^{**}
$h = 3$	13.46	10.425	4.186^{*}	3.971^{*}
$h = 4$	13.553	10.672	4.712^{*}	4.323^{**}
$h = 5$	13.461	11.192	4.921^{**}	4.358^{**}
$h = 6$	12.6	10.723	4.928^{**}	4.29^{**}
$h = 7$	11.577	10.525	4.632^{*}	3.982^{**}
$h = 8$	11.224	10.24	4.921^{*}	4.075^{**}
$h = 9$	10.513	9.775	5.301	4.332
$h = 10$	10.568	10.102	5.615	4.677

Note: The table shows the heatmap of absolute RMSE values for the competing models at forecast horizons $h = 0, 1, \ldots, 10$ for the period 2021M7-2024M6. RMSE values are based on year-on-year rates of the HICP-PACK. Results for the Diebold and Mariano (1995) test in case of outperformance relative to the benchmark are indicated by the symbols * (5% level) and ** (1% level).