DEUTSCHE
BUNDESBANK
EUROSYSTEM

Technical Paper

Empirical findings on upper-level aggregation issues in the HICP

03/2025

Julika Herzberg
Thomas A. Knetsch
Dilyana Popova
Jannik Schaller
Patrick Schwind
Sebastian Weinand

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 email: www.bundesbank.de/contact

Internet http://www.bundesbank.de

Non-technical summary

Research question

Consumer price indices usually measure the change in prices for a representative basket of goods and services over time (inflation). Measurement errors can occur in a variety of ways. If the expenditure weights of individual basket items change, for example as a result of changes in consumption patterns, these adjustments are usually incorporated into the consumer price index only after a certain time lag. As a consequence, price changes are no longer depicted in a representative manner. This paper examines two sources of mismeasurement linked to the use of expenditure weights (upper-level aggregation) when compiling the Harmonised Index of Consumer Prices (HICP), which serves as a key metric for gauging price stability, and thus for deciding the monetary policy stance adopted in the euro area.

Contribution

Measurement bias and uncertainty are quantified for the HICP between 2012 and 2021 on the basis of publicly available price indices and consumer expenditures. The analysis refers to the five biggest euro area countries, which together account for approximately 80% of the euro area, both separately and as an aggregate (Big-5). A superlative price index capturing substitution effects more accurately than the HICP is used as a benchmark for "true" inflation. In addition to this, when computing the expenditure weights of the benchmark index, revised data from national accounts are utilised. By contrast, when calculating the HICP, it is necessary over time to rely on whatever data has most recently become available, meaning that subsequent adjustments to these often provisional data are not taken into account. Any differences between the HICP and the benchmark index are attributed to measurement errors which can be divided up into a representativity component and a data vintage component. In addition, the representativity effect is also analysed over the extended period from 1997 to 2024.

Results

In times of moderate inflation, measurement errors at the upper level of aggregation are generally small for the HICP. The representativity component and the data vintage component generate on average an upward bias in the Big-5 HICP inflation rate of less than one-tenth of a percentage point, with the extent of mismeasurement engendered by each of these two components being broadly the same. The inclusion of the years, influenced by the COVID-19 pandemic (2020 and 2021), has only a small effect on the results of the analysis. In the high inflation period (2022 and 2023), however, measurement uncertainty was strongly elevated as a result of a substantial increase in the representativity component.

Nichttechnische Zusammenfassung

Fragestellung

Verbraucherpreisindizes messen üblicherweise die Preisentwicklung eines repräsentativen Warenkorbs über die Zeit (Inflation). Dabei können verschiedene Arten von Messfehlern auftreten. Ändern sich - etwa wegen Änderungen im Konsumverhalten - die Ausgabengewichte einzelner Warenkorbpositionen, fließen diese Anpassungen meist erst mit Verzögerung in den Verbraucherpreisindex ein. Die Preisentwicklung wird dann nicht mehr repräsentativ abgebildet. In der vorliegenden Arbeit werden zwei mit der Verwendung von Ausgabengewichten (obere Aggregationsebene) verbundene Messfehler für den Harmonisierten Verbraucherpreisindex (HVPI) untersucht, welcher als zentraler Indikator für die Messung von Preisstabilität und damit für die Ausrichtung der Geldpolitik im Euro-Raum dient.

Beitrag

Auf Basis öffentlich verfügbarer Preisindizes und Konsumausgaben werden Verzerrung und Messungenauigkeit für den HVPI im Zeitraum von 2012 bis 2021 quantifiziert. Die Analyse bezieht sich dabei auf die fünf größten Euro-Raum Länder, die zusammen etwa 80% des Euro-Raums ausmachen, sowohl getrennt als auch im Aggregat (Big-5). Als Referenz für die „wahre" Preisentwicklung wird ein sogenannter superlativer Preisindex verwendet, welcher Substitutionseffekte genauer abbildet als der HVPI. Zudem werden bei der Berechnung der Ausgabengewichte des Referenzindex revidierte Angaben aus den Volkswirtschaftlichen Gesamtrechnungen verwendet. Bei der Berechnung des HVPI muss dagegen im Zeitablauf auf den jeweils aktuellen Datenstand zurückgegriffen werden; spätere Revisionen dieser oft vorläufigen Daten sind demnach nicht berücksichtigt. Die Abweichungen zwischen HVPI und Referenzindex werden als Messfehler angesehen, die in eine Repräsentativitäts- und eine Datenstandskomponente aufgegliedert werden können. Zusätzlich wird der Repräsentativitätseffekt im erweiterten Zeitraum von 1997 bis 2024 analysiert.

Ergebnisse

In Zeiten moderater Inflation sind die Messfehler auf der oberen Aggregationsebene grundsätzlich klein. Die Repräsentativitäts- und die Datenstandskomponente führen zu einer durchschnittlichen Verzerrung der HVPI-Inflationsrate der Big-5 Länder um weniger als ein Zehntel Prozentpunkt nach oben, wobei sich die Beiträge der beiden Komponenten nicht allzu stark unterscheiden. Das Einbeziehen der ersten beiden Jahre der COVID-19 Pandemie (2020 und 2021) beeinflusst die Ergebnisse der Analyse nur wenig. In der Hochinflationsphase (2022 und 2023) war die Messungenauigkeit hingegen als Folge einer stark gestiegenen Repräsentativitätskomponente deutlich erhöht.

Empirical findings on upper-level aggregation issues in the HICP*

Julika Herzberg Thomas A. Knetsch Dilyana Popova

Jannik Schaller Patrick Schwind Sebastian Weinand

September 17, 2025

Abstract

We analyse potential mismeasurement of the Harmonised Index of Consumer Prices (HICP) at the upper level of aggregation, focusing on two sources of measurement error: the choice of index formula (representativity component) and the reliability of weights (data vintage component). The representativity component captures the fact that a Laspeyres-type index such as the HICP suffers from a systematic overestimation of inflation due to the disregard of changes in more recent consumption patterns. The data vintage component comprises potential mismeasurement arising from the annual updating of HICP weights based on preliminary national accounts data. With national accounts vintage data, we calculate bias and inaccuracy metrics in order to analyse mismeasurement at the upper level of aggregation in the HICPs for Germany, France, Italy, Spain and the Netherlands, as well as for the country group, over the period from 2012 to 2021. For the representativity component, the data availability allows an additional analysis of the period until 2024. Measured in terms of annual HICP rates, the total upper-level aggregation bias falls short of one-tenth of a percentage point. The representativity and the data vintage component are both found to contribute to overall bias in quite similar shares. We also find that the representativity component experienced a substantial revival during the recent high inflation period.

Keywords: Inflation measurement · Representativity bias · Updating of weights · High inflation period

JEL classification: E31 · C43

*Correspondence to: Jannik Schaller, jannik.schaller[@]bundesbank.de, Deutsche Bundesbank, Directorate General Data and Statistics, Postfach 10 06 02, 60006 Frankfurt am Main, Germany. This paper is an updated version of a background study prepared for the workstream "Inflation Measurement" of the Monetary Policy Strategy Review carried out by the Eurosystem in 2020 and 2021. Earlier versions of the paper were presented at the 17th Meeting of the Ottawa Group on Price Indices in Rome in 2022 and at the Statistische Woche 2025 in Wiesbaden.

1 Introduction

As confirmed in its most recent monetary policy strategy review, the Governing Council of the European Central Bank (ECB) considers the Harmonised Index of Consumer Prices (HICP) as "the appropriate price measure for assessing the achievement of the price stability objective" (ECB, 2025, p. 1). According to the monetary policy aim, the year-on-year percentage change of the HICP (henceforth referred to as inflation) is targeted to be at 2% over the medium term, suggesting that the Governing Council sees the need for an inflation buffer above zero. The existence of a measurement bias is amongst the reasons for this inflation buffer.

Measurement issues generally arise at various stages of HICP compilation. Mismeasurement can thus stem from different sources, including the disregard of changes in consumption patterns, belated introduction of new products, untimely account of new distribution channels and improper adjustment for quality changes (for an overview see International Monetary Fund et al., 2020, Chapter 12). When it comes to the aggregation of individual price changes over the basket of goods and services, a distinction is made between the lower and the upper level. At the lower level, prices are aggregated without any weighting information while, at the upper level, households' expenditure shares are applied to form price indices.

Designed as a Laspeyres-type index, the HICP generally measures the aggregate price change of a fixed basket of goods and services (cost-of-goods index or COGI). In strict terms, HICP weights are only representative for the base period in the price comparison. A change in the consumption patterns from the base period to the comparison period may induce a source of mismeasurement which is henceforth called representativity bias.[1] A further source of mismeasurement at the upper level of aggregation stems from using preliminary national accounts data in the annual updating of weights. As the HICP is not

[1] For example, during periods of high inflation, consumers may shift their spending from more expensive goods (such as fresh produce) to cheaper alternatives (such as frozen goods). A fixed basket of goods, as used in the Laspeyres index, does not account for this shift, leading to an overestimation of inflation.

allowed to be revised in order to take account of new releases in the national accounts, a data vintage effect may impair inflation measurement, provided that national accounts are expected to converge to "true" consumption patterns from earlier to later releases.

Until 2020, the first releases of consumption expenditures from the national accounts data for the year $y-2$ were used to update the annual weights of the HICP for a calendar year y. However, for the HICP in 2021, the COVID-19 pandemic posed the challenge that the consumption patterns from 2019 ($y-2$) did not provide a realistic approximation of the consumption patterns for 2020, as consumption behaviour changed significantly during the pandemic, for example in terms of leisure activities or travel. Accordingly, for the HICP in 2021, it was necessary to use more timely information for the consumption patterns. Instead of using data from $y-2$, the statistical offices used preliminary consumption data for $y-1$ in order to better reflect the expected consumption patterns for 2021 (Eurostat, 2020). This approach was also adopted for the subsequent years after 2021, which leads to a methodological change regarding the data vintage component from 2021 on.

The focus of this paper is on estimating the extent of HICP mismeasurement at the upper level of aggregation, thereby separating out the effects of imperfect representativity and the use of preliminary data in weight compilation. While theory suggests that the representativity effect is positive, the sign of the data vintage effect is generally unknown up front. Apart from the bias, it is worth looking at the root mean squared deviation and the interdecile range as measures of inaccuracy. The analysis is similar to what Herzberg et al. (2023) studied for Germany. In general, the same formal evaluation framework is applied. Throughout this paper, we consider the HICPs of the five largest euro area countries (Germany, France, Italy, Spain and the Netherlands), as well as of the country group (henceforth called Big-5 aggregate). Given that the Big-5 aggregate covers four-fifths of the euro area HICP,[2] the empirical findings give insight into upper-level

[2] The HICP country weights are derived from the households final monetary consumption expenditure, which is part of the national accounts.

aggregation issues of the ECB's key inflation measure.[3]

In order to carry out both the representativity and the data vintage parts of HICP mismeasurement at the upper level of aggregation for this group of countries, we use the Törnqvist index as superlative index against which the HICP is evaluated. In addition, this benchmark index needs to be adjusted with regard to the weight concept. For this purpose, we construct the superlative index on the basis of weighting schemes using final, or at least revised, information from national accounts (henceforth called final NA weights) instead of preliminary national accounts data. The benchmark index is assumed to better proxy the "true" aggregate price development than the HICP, as it is formed on a symmetric weighting using timely and more mature information about households' consumption expenditures.

We consider the period from January 2012 to December 2021. The start of the sample is chosen because the annual updating of HICP weights became mandatory in 2012. The sample terminates by the end of 2021.[4] Hence, the period under consideration also includes the COVID-19 pandemic starting from 2020, which resulted in the aforementioned statistical-methodological change regarding the weight updating procedure by using more timely national accounts information ($y - 1$ instead of $y - 2$). In addition, we provide evidence for the representativity component over the whole HICP history starting in 1997 and terminating by the end of 2024. This allows us to take a long-run, albeit partial, view on HICP upper-level mismeasurement. This extended period covers a number of economic phenomena whose impacts on inflation measurement are worth studying. First, there are recessionary periods resulting from the Financial Crisis 2008/2009 and the COVID-19 pandemic. Second, the COVID-19 pandemic also involves substantial changes in consumption patterns that required modifications in the weight update procedure. Third,

[3] It is not feasible for us to carry out the analysis on the basis of a total representation of the euro area, as it has been impossible to gather national accounts vintage data for all euro area countries.

[4] More recent years could not be taken into account due to the new classifications of the consumption expenditures in the course of the national accounts general revision in 2024, making them impossible to match with the HICP data.

the time period up to 2024 allows us to examine the representativity component within the high inflation period (from 2022 to 2023).

Our main conclusions are the following. In phases of moderate inflation, the upper-level aggregation bias of the HICP, composed by the representativity and the data vintage component, is small. Measured in terms of annual HICP rates, the total upper-level aggregation bias of the Big-5 aggregate clearly falls short of one-tenth of a percentage point. The representativity and the data vintage components contribute to the overall bias in quite similar shares. As expected by theory, the representativity component is positive for all countries under consideration. The representativity effect for the euro area HICP is even markedly smaller than that for the Big-5 aggregate. Data vintage components are positive in all countries but the Netherlands. Adding the first two years of the COVID-19 pandemic (2020 to 2021) has virtually no impact on the results as it just leads to a very small decrease, in absolute terms, of the measurement bias for most of the countries as well as the country group. In contrast, the high inflation period (from 2022 to 2023) has a significantly large influence on the representativity component.

The uncertainty surrounding HICP inflation due to upper-level aggregation issues is small, too. The interdecile range of the deviations between the HICP and the benchmark amounts to about one-tenth of a percentage point for the Big-5 aggregate. For the individual countries, we evidence wider interdecile ranges, suggesting that contrary developments in country HICPs tend to balance each other out in the aggregate. In most cases, adding data from the first two years of the COVID-19 pandemic does not affect the interdecile ranges. Considering the root mean squared deviation, the results for the Big-5 aggregate as well as for most of the individual countries confirm that the representativity and data vintage components make fairly equal contributions to HICP inaccuracy at the upper level of aggregation. While, in its first two years, the COVID-19 pandemic affects the root mean squared deviation differently from country to country, it has virtually no impact for the Big-5 aggregate. For the high inflation period, however, we find a large increase in the root mean squared deviation of the representativity bias for all countries and country groups under review.

Against the backdrop of existing evidence reported mainly for the US consumer price index (CPI), the upper-level aggregation bias of the HICP turns out to be a relatively small number. However, the high inflation period has increased the represenativity bias significantly, emphasising that the representativity component is experiencing a substantial resurgence. In the report of the Boskin Commission (Boskin et al., 1998), for instance, only 0.15 of the 1.1 percentage points total bias per annum was found to be due to upper-level substitution, while a much larger portion was due to the introduction of new products and quality changes. Later research on the subject by Greenlees and Williams (2010) found the upper-level bias to be more prevalent, amounting to 0.3 percentage points per annum. Armknecht and Silver (2014) found evidence in the post-2002 US CPI that the Boskin Commission's findings on the presence of measurement bias still hold, with an upper-level aggregation bias of 0.16 percentage points.[5]

Silver and Ioannidis (1994) paid attention to potential mismeasurement caused by the use of "untimely weights" and, thus, considered a phenomenon which is quite similar to the data vintage effect studied in this paper. This is in fact not the only similarity. In addition, they expanded the range of statistical metrics by looking also at the root mean squared error, for instance, and they considered European CPIs in their empirical investigation. Herzberg et al. (2023) examined mismeasurement of the German HICP with regard to both the representativity and the data vintage components. In contrast to our paper, they calculate what they call "full-information" weights. These utilise the universe of available information at the time to which it relates.[6] In our analysis, we

[5] While plenty of research has studied the impact of upper-level aggregation bias in the US, empirical evidence for the HICP is rarely available. The report of the Boskin Commission can be credited for later on sparking further research interest in CPI measurement bias outside the US. Among the current EU members, in the nineties the topic was studied with respect to inflation in France (Lequiller, 1997), Portugal (Neves and Sarmento, 1997) and Germany (Hoffmann, 1998).

[6] For the calculation of the "full-information" weights, the information from the multi-year household budget survey (HBS) is the most relevant additional source.

use "final weights" from national accounts data instead of "full-information" weights.[7] However, it should be noted that theoretical considerations suggest that the data vintage effect may be interpreted as a lower bound if calculated using final NA weights instead of taking the universe of information into account.

The remainder of this paper is structured as follows. In the next section, the evaluation framework is sketched out. In Section 3, empirical results are presented. In Section 4, conclusions are drawn.

2 Methodology

In this section, we describe the evaluation framework. In Section 2.1, we start with a brief explanation of key HICP construction principles. We follow up with the exposition of weight concepts and index formulae in Sections 2.2 and 2.3. Finally, in Section 2.4, we introduce the statistical metrics which are employed to measure HICP mismeasurement at the upper level of aggregation.

2.1 Derivation of official HICP weights

The HICP is designed as a chain-linked Laspeyres-type index where weights are updated at the beginning of each calendar year and kept constant throughout (EU, 2020).[8] In formal terms, it may be written as:

$$P^{o}_{\text{HICP}}(y,m) = P^{o}_{\text{HICP}}(y-1,12) \cdot \sum_{i=1}^{I} w^{o}_{i}(y-1,12) \cdot \frac{p_i(y,m)}{p_i(y-1,12)} \ ,$$

[7] It goes beyond the scope of this study to calculate full-information weights that are comparable across all considered countries. Hence, final NA weights serve as a uniform basis for computing the data vintage effects throughout this paper in order to ensure comparability.

[8] In the corresponding academic literature the HICP is often referred to as a Lowe index since weight and price reference periods are different from each other (Lowe, 1823).

where $p_i(y,m)$ is the price of good i ($i = 1,\ldots,I$) in year y ($y = 1,\ldots,Y$) and month m ($m = 1,\ldots,12$). The weight of good i applied to the HICP (henceforth called "official weight" and marked by superscript "o") in year y is denoted by $w_i^o(y-1,12)$, as it refers to the price reference period which is December of the previous year. For notational convenience, however, we write $w_i^o(y-1) \equiv w_i^o(y-1,12)$ in the remainder.

According to Eurostat (2020, p. 2), Article 3.1 of the HICP Implementing Regulation (EU, 2020) means that "the expenditure shares used for the HICP in year t should be representative of year $t-1$". On the basis of these expenditure shares (referring to annual household consumption expenditure data from the national accounts), HICP weights result from an obligatory price update to December, i.e. $w_i^o(y-1) = w_i(y-1) \cdot p_i(y-1,12)/p_i(y-1)$ where $w_i(y-1)$ and $p_i(y-1)$ indicate the average expenditure share and price of good i in year $y-1$ respectively.

Until the outbreak of the COVID-19 pandemic, the weight updating procedure of the statistical offices for the HICP of calendar year y relied on national accounts data on consumption expenditures for $y-2$ (see Tab. 1). HICP legislation obliges statistical offices to review and update the expenditure shares of $y-2$ to make them representative of year $y-1$, implying a freedom of choice as regards the options "to-price-update" or "not-to-price-update" from $y-2$ to $y-1$ (Eurostat, 2024, Sections 3.5 and 8.2.3). As far as we are aware, the statistical offices of Germany, France, Italy and Spain generally made use of price-updating from 2012 to 2020,[9] whereas the Dutch statistical office generally opted not to price update from $y-2$ to $y-1$.

However, the pre-pandemic weight updating approach based on data from $y-2$ was not suitable to apply for the year 2021. This is because the COVID-19 pandemic led to a significant shift in consumption patterns from 2019 to 2020. Keeping the existing updating procedure would have meant to proxy the 2020 consumption structure with pre-pandemic national accounts data from 2019. Hence, Eurostat and national statistical

[9] In the French HICP, price-updating was applied as a general rule, while the possibility of adjusting to the previous year's expenditures was retained for exceptional cases where significant changes were identified.

offices agreed to use national accounts data for 2020 – even if they are preliminary and incomplete – as the basis for the derivation of HICP weights for the reporting year 2021 (Eurostat, 2020). This $y-1$ approach was maintained until the HICP reporting year 2024. The break in statistical methodology concerning the weight update procedure from 2021 onward is illustrated in the right-hand side of Tab. 1. In a simplifying manner, we also distinguish the two approaches with regard to the vintage dimensions. Until the reporting year 2021, the weights of the HICP are calculated based on national accounts data for $y-2$, which are available by the *end* of year $y-1$. Since 2021, however, preliminary data from $y-1$ are used, which are available at the *beginning* of year y.

2.2 Derivation of final NA weights

The procedure for updating the weights in the period under review from 2012 to 2021 differs between (i) using national accounts data from $y-2$ (until 2020) and (ii) using preliminary and incomplete data from $y-1$ (since 2021). In general, as time goes by, the information content of national accounts data becomes more adequate along the time and vintage dimensions. Instead of using first or preliminary releases for $y-2$ or $y-1$, respectively, final or at least revised data could be used. Weights formed on the basis of such final or at least revised information content are called final NA weights (henceforth indicated by superscript "f") and are generally expected to be closer to the (unknown) "true" expenditure shares needed to compile the best possible aggregate price index.

We pinpoint the difference between the official (w_i^o before and since 2021) and final

NA weights (w_i^f) by comparing the updating formulae of the two weight concepts:[10]

$$w_i^o(y-1) = \begin{cases} \bar{w}_i(y-\xi) \cdot \frac{c_i(y-2;y-1)}{c_i(y-\xi;y-1)} \cdot \frac{p_i(y-1)}{p_i(y-2)} \cdot \frac{p_i(y-1,12)}{p_i(y-1)} & \text{for } y < 2021 \\ \bar{w}_i(y-\xi) \cdot \frac{c_i(y-1;y-1)}{c_i(y-\xi;y-1)} \cdot \frac{p_i(y-1,12)}{p_i(y-1)} & \text{for } y \geq 2021 \end{cases} \tag{1a}$$

$$w_i^f(y-1) = \bar{w}_i(y-\xi) \cdot \frac{c_i(y-1;\infty)}{c_i(y-\xi;\infty)} \cdot \frac{p_i(y-1,12)}{p_i(y-1)} \tag{1b}$$

where $c_i(y;v)$ is households' consumption expenditure of good i in year y as it is reported in the national accounts vintage released in year v. Hence, y corresponds to the reporting period, while v refers to the release date. The final vintage is denoted by $v = \infty$. According to the release calendar of national accounts, detailed consumption expenditures for the year $y-2$ are available by the end of $y-1$.

In Eq. (1a), v equals $y-1$ in the case of $y \geq 2021$. This reflects the practice of statistical offices to use internal best-guess estimates of national accounts for $y-1$ at the beginning of year y though official national accounts for year $y-1$ have not yet been published. For $y < 2021$, Eq. (1a) describes the formula employed in the "to-price-update" option; in the "not-to-price-update" option, the factor $[p_i(y-1)]/[p_i(y-2)]$ does not appear. In Eq. (1b), the "to-price-update" factor is missing, too, because the final vintage comprises information about consumption expenditures of every weight reference period by definition.

Both weight updating formulae have in common that the extrapolation with consumption expenditures is anchored by, say, a hypothetical base weight $\bar{w}_i(y-\xi)$ which is derived from the universe of information about consumption patterns but is available only with a

[10] For the sake of better readibility, equations are simplified in two respects. First, the national accounts breakdown of households' consumption expenditures is not as detailed as needed for the HICP. Hence, the updating of weights is regularly impossible to be made using the same expenditure category (as displayed in these equations) but a broader one. Second, updated weights need to be scaled such that they altogether sum up to unity. This scaling factor is omitted in the equations.

time lag of $\xi > 2$ years. The knowledge of base weights is not needed for the calculation of final NA weights. By substituting Eq. (1a) in Eq. (1b), we obtain the following expression for final NA weights:

$$w_i^f(y-1) = \begin{cases} w_i^o(y-1) \cdot \frac{c_i(y-1;\infty)}{c_i(y-\xi;\infty)} \,/\, \left[\frac{c_i(y-2;y-1)}{c_i(y-\xi;y-1)} \cdot \frac{p_i(y-1)}{p_i(y-2)}\right] & \text{for } y < 2021 \\ w_i^o(y-1) \cdot \frac{c_i(y-1;\infty)}{c_i(y-\xi;\infty)} \,/\, \left[\frac{c_i(y-1;y-1)}{c_i(y-\xi;y-1)}\right] & \text{for } y \geq 2021 \end{cases} \tag{2}$$

If we assume $c_i(y-\xi; y-1) = c_i(y-\xi; \infty)$ for ξ sufficiently large because of the frontloading of current revisions, we end up with the relation:

$$\frac{w_i^f(y-1)}{w_i^o(y-1)} = \begin{cases} \frac{c_i(y-1;\infty)}{c_i(y-2;y-1)\cdot[p_i(y-1)/p_i(y-2)]} & \text{for } y < 2021 \\ \frac{c_i(y-1;\infty)}{c_i(y-1;y-1)} & \text{for } y \geq 2021 \end{cases} \tag{3}$$

suggesting that, for $y < 2021$, the ratio between the official and final NA weight of some good i is equal to the ratio between the final release for households' consumption expenditure of good i in the weight reference period $y-1$ and the price-updated first release referring to one year prior. For $y \geq 2021$, it is suggested that the ratio between the official and final NA weight for good i is equal to the ratio between the final and the preliminary release for the households' consumption expenditure of good i in the weight reference period $y-1$. In the stylised vintage dataset displayed in Tab. 1, the entries used as final releases are denoted by "f" and the first releases by "o".

		vintage available at the...												
HICP	NA*	end of year										beginning of year		
reporting period		block A		block B					block C					
		2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
	y_0	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	2009	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
2012	2010	o	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
2013	2011		o	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	f
2014	2012			o	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	f
2015	2013				o	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	f
2016	2014					o	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	f
2017	2015						o	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	f
2018	2016							o	$\times$	$\times$	$\times$	$\times$	$\times$	f
2019	2017								o	$\times$	$\times$	$\times$	$\times$	f
2020	2018									o	$\times$	$\times$	$\times$	f
	2019										$\times$	$\times$	$\times$	f
2021	2020											o	$\times$	f
2022	2021												o	f
2023	2022													o/f

Note: Entries in the vintage dataset are denoted by "$\times$" with two exceptions. The first releases which are employed in the calculation are marked by "o". The entries which are taken as final releases are marked by "f". Note that for Germany, vintages are also available at the beginning of 2024.

*NA: national accounts

Table 1: Stylised vintage dataset.

Apart from current revisions which result from capturing lagged statistical information, national accounts are subject to benchmark revisions. In multi-year intervals, often every five years and harmonised among European countries, conceptual and methodological enhancements are introduced. In addition, benchmark revisions are often an occasion initiating the account of new data sources. Benchmark revisions generally alter the complete time series of households' consumption expenditures. This generally impairs the comparability across the vintage dimension.

The period under consideration includes benchmark revisions in 2013, 2018 and 2024. In contrast to the benchmark revisions in 2013 and 2018, the revision in 2024 involved the implementation of a new classification for consumption expenditures. This transition of national accounts to COICOP 2018 makes it impossible to adequately match with the HICP data, which is still classified according to COICOP 1999.[11] Since later releases are based on the new classifications, the vintage dimension for France, Italy, Spain and the Netherlands ends with the release available at the beginning of 2023. For Germany, consumption expenditures classified according to COICOP 1999 are available until the beginning of 2024.

In Tab. 1, the various accounting regimes are denoted by "A", "B" and "C". The benchmark revisions in 2013 and 2018, however, did not comprise significant changes in the classification of consumption expenditures. Therefore, it is possible to account for the benchmark revisions in 2013 and 2018 via vintage transformations.

Vintages before and after a benchmark revision are generally made comparable by a vintage transformation. This is carried out following Knetsch and Reimers (2009). In particular, we run bivariate cointegrating regressions:

$$\ln c_i(y, 2017) = \alpha_i^B + \beta_1^B \ln c_i(y, 2023) + \epsilon_i^B(y) \tag{4a}$$

$$\ln c_i(y, 2012) = \alpha_i^A + \beta_1^A \ln c_i(y, 2023) + \epsilon_i^A(y) \tag{4b}$$

where α_i^R, β_i^R $(R = A, B)$ are regression coefficients and ϵ_i^R are covariance stationary

[11] COICOP: Classification of Individual Consumption According to Purpose.

residuals. The samples cover the periods from y_0 to 2017 in block "B" and from y_0 to 2011 in block "A".

The quality of vintage transformation functions is checked using the coefficient of determination (R^2). If $R^2 \geq 0.8$, the first releases of the blocks "A" and "B" are made comparable with the final releases, applying the following transformation:

$$\hat{c}_i(y; y+1) = \exp(\alpha_i^A)[c_i(y; y+1)]^{\beta^A}, \quad y = 2010, 2011, \tag{5a}$$

$$\hat{c}_i(y; y+1) = \exp(\alpha_i^B)[c_i(y; y+1)]^{\beta^B}, \quad y = 2012, ..., 2016. \tag{5b}$$

Otherwise, first releases of blocks "A" and "B" remain untransformed and the latest vintages before the respective benchmark revision are taken as final releases.[12]

The vintage data sets available for the five countries under consideration are not entirely homogeneous but differ in the number of available vintages, the time length of the data in each vintage and lack of information for some positions. Data are retrieved from the ECB's Statistical Data Warehouse and double-checked with, and occasionally complemented by, information from the national accounts data repositories of the national central banks of the countries under review. The very few remaining gaps in the vintage data sets are filled by estimates.[13]

2.3 Index formulae

In the following analysis, monthly year-on-year price relatives are aggregated to make summary metrics interpretable as a source of mismeasurement of inflation, i.e. comparable to the year-on-year percentage change of a price index and measured in percentage points (see Herzberg et al., 2023, footnote 10). Thus, the aggregate price relative representing

[12] This alternative induces the data vintage effect to be systematically distorted downward because the impact of later current revisions is not captured. Hence, it is no more than a surrogate in the rare cases where the estimation of proper vintage transformation functions fail.

[13] The vintage data sets and detailed meta information are available upon request.

the HICP is defined as:

$$P_L^o(y,m) = \sum_{i=1}^{I} w_i^o(y-1) \cdot \frac{p_i(y,m)}{p_i(y-1,m)}\,. \tag{6}$$

We indicate this index with subscript "L" because it is of a Laspeyres type and superscript "o" because it is based on official weights.

The benchmark index against which P_L^o is evaluated is designed by a superlative index which symmetrically incorporates the weights of both the base period and the comparison period. Superlative indices are Fisher, Törnqvist or Walsh indices, for example. In this study, we restrict the exposition of results to the Törnqvist index, where arithmetic averages of the value shares in the two periods are used as weights.[14] The aggregate price relative representing the benchmark index is defined as:

$$P_{T\ddot{o}}^x(y,m) = \prod_{i=1}^{I} \left[\frac{p_i(y,m)}{p_i(y-1,m)}\right]^{\frac{1}{2}\left[w_i^x(y-1)+w_i^x(y)\right]}, \quad x = f, o\,. \tag{7}$$

In our analysis, we calculate Törnqvist indices using official and final NA weights.

2.4 Bias and inaccuracy metrics

HICP mismeasurement at the upper level of aggregation is evaluated using a number of statistical metrics building on the deviation of the Laspeyres-type index based on official weights (P_L^o) from the Törnqvist index based on final weights ($P_{T\ddot{o}}^f$). In order to disentangle representativity and data vintage effects, the deviation is decomposed in the following way:

$$\frac{P_L^o}{P_{T\ddot{o}}^f} = \frac{P_L^o}{P_{T\ddot{o}}^o} \cdot \frac{P_{T\ddot{o}}^o}{P_{T\ddot{o}}^f}\,. \tag{8}$$

The first term relates a Laspeyres-type index with a Törnqvist index, both using official HICP weights. This is a measure of the representativity effect. The second term consists

[14] In Herzberg et al. (2023), statistical mismeasurement metrics are reported using Fisher, Törnqvist or Walsh indices, confirming the well-known result that metrics are rather insensitive to the choice of the superlative index formula.

of two Törnqvist indices, the one based on official and the other on final NA weights. This ratio captures the data vintage effect.

To evaluate the quality of the current HICP measurement, we focus on bias measured by the mean deviation (MD) and on inaccuracy measures employing the root mean squared deviation (RMSD) and the interdecile range (IDR). Complementing the MD, the RMSD and the IDR offer a comprehensive view of the uncertainty surrounding HICP inflation.

The mean deviation (MD) is defined by:

$$MD_{\text{Total}} = \frac{1}{T} \sum_{t=1}^{T} \ln \left(P_L^o(t) / P_{T\ddot{o}}^f(t) \right). \tag{9}$$

According to Eq. (8), the measurement bias can be additively decomposed into a representativity component and a data vintage component:

$$\begin{aligned} MD_{\text{Total}} &= MD_{\text{Representativity}} + MD_{\text{Data vintage}} \\ &= \frac{1}{T} \sum_{t=1}^{T} \ln \left(P_L^o(t) / P_{T\ddot{o}}^o(t) \right) + \frac{1}{T} \sum_{t=1}^{T} \ln \left(P_{T\ddot{o}}^o(t) / P_{T\ddot{o}}^f(t) \right). \end{aligned} \tag{10}$$

The root mean squared deviation ($RMSD$) is separately calculated for the total effect as well as for the representativity and the data vintage sources of mismeasurement:

$$RMSD_{\text{Total}} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \ln \left(P_L^o(t) / P_{T\ddot{o}}^f(t) \right)^2} \tag{11a}$$

$$RMSD_{\text{Representativity}} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \ln \left(P_L^o(t) / P_{T\ddot{o}}^o(t) \right)^2} \tag{11b}$$

$$RMSD_{\text{Data vintage}} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \ln \left(P_{T\ddot{o}}^o(t) / P_{T\ddot{o}}^f(t) \right)^2} \tag{11c}$$

In contrast to the bias, $RMSD_{\text{Representativity}}$ and $RMSD_{\text{Data vintage}}$ do not sum up to $RMSD_{\text{Total}}$.

An additional uncertainty measure is the interdecile range:

$$IDR_{\text{Total}} = P_{90} - P_{10} \ , \tag{12}$$

where P_{10} and P_{90} are the 10th and the 90th percentiles of $\ln\left(P_L^o(t)/P_{T\ddot{o}}^f(t)\right)$.

3 Results

The empirical study is based on price indices and weights for 76 product groups, containing COICOP positions at the two, three or four digit level. In terms of number and breakdown of detailed HICP data, the data sets are uniform for the five countries under consideration.

As shown in Tab. 1, for all countries except Germany, the vintage dimension ends with the release available at the beginning of 2023 due to new classifications of consumption expenditures in the course of the national accounts general revision in 2024. Hence, for the reporting period 2022, no revised or final weights are available since the vintages released in 2024 are already based on the new classifications. With regard to the data vintage component (see Eq. (8)), this restricts the time period under consideration and implies that our sample terminates in 2021, since final or at least revised weights for the year y are required for calculating $P_{T\ddot{o}}^f$. For Germany, however, the vintage dimension ends with the release available at the beginning of 2024, which allows an analysis up to the reporting period 2022. We provide the results for the extended period in Appendix A.

The bias and inaccuracy measures of the full-fledged analysis separating representativity and data vintage components are reported in Section 3.1. In Section 3.2, we additionally provide results for the representativity component over the whole HICP history starting in 1997 and terminating by the end of 2024. The long sample consists of 336 monthly observations.

3.1 Bias and inaccuracy

Bias and inaccuracy metrics are calculated on the basis of the logarithmic deviation of a Laspeyres-type index based on official weights from the benchmark index, which is a Törnqvist price index using final NA weights. From a mathematical perspective, the logarithmic deviation is measured as a percentage of the benchmark index. Owing to the construction of the price indices, monthly deviations and bias estimates may be interpreted as mismeasurement in HICP inflation, measured in percentage points, and inaccuracy metrics as uncertainty measures surrounding HICP inflation, also measured in percentage points.

Deviation over time. Fig. 1 displays the monthly logarithmic deviations for all five countries as well as the Big-5 aggregate from January 2012 to December 2021 and their decomposition according to Eq. (8). Overall, the time series profiles observed for Germany, France, Italy and Spain exhibit quite similar characteristics. These are found in the Big-5 aggregate, too. In the case of the Netherlands, however, the pattern is distinctly different.

In total over all countries, we observe that the monthly deviations range from -0.16 to 0.30 percentage points. The ranges vary from one country to another. While it is relatively low for France (from -0.06 to 0.19), it is quite high for Spain (from -0.11 to 0.28) and the Netherlands (from -0.16 to 0.24). Both Germany and Italy have their minimum at -0.04, but the maximum of Germany (0.30) is about 25% higher than that of the Netherlands. The range of the deviations of the Big-5 aggregate is substantially lower compared to the country ranges. Until 2019 it is almost completely located above the zero line.

As regards the decomposition of total deviations, we observe almost exclusively positive contributions from the representativity effect for all countries and the country group until 2019. After 2019, notable negative representativity effects can be observed in France, Spain, the Netherlands and in the country group. A striking feature is that, in the Netherlands, the components mostly have different signs, whereas in the other countries and in the country group, they are mostly the same. This implies that, in arithmetical terms,

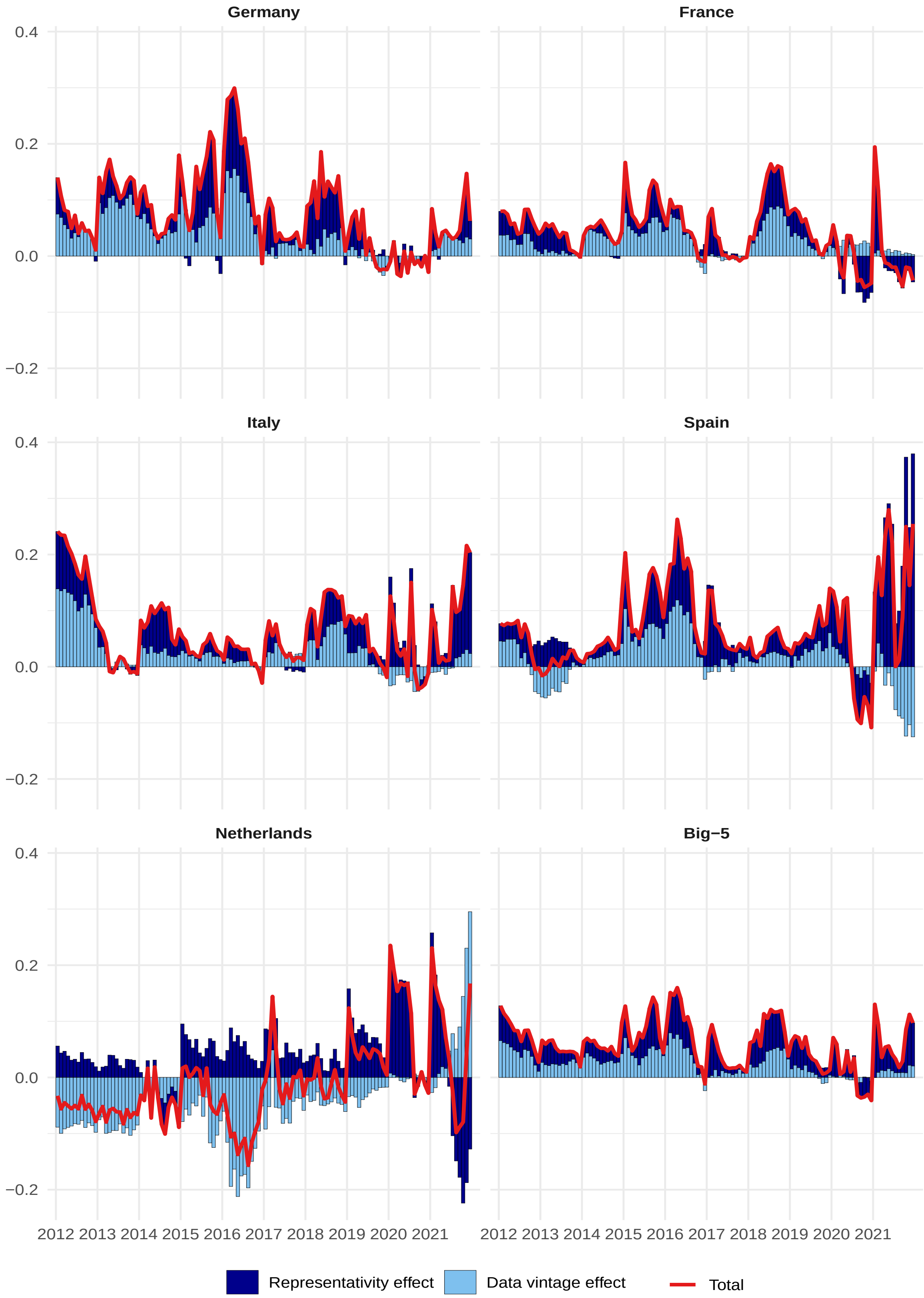

Figure 1: Monthly deviations (in percentage points p.a.) in the year-on-year change rates between official Laspeyres-type index and superlative Törnqvist index with final NA weights, decomposed into representativity and data vintage effects.

representativity and data vintage components are typically compensating each other in the case of the Netherlands, while they are reinforcing each other in the former group of countries.

Bias. We estimate the total upper-level aggregation bias by averaging the monthly deviations over the complete sample as well as for the pre-COVID period from January 2012 to December 2019 (see Tab. 2), which is not affected by COVID-related aspects such as significant changes in consumption patterns or a recessionary period. For the Big-5 aggregate, the bias is positive but small in both cases. It falls short of one-tenth of a percentage point. It is also positive for Germany, France, Italy and Spain, whereas it is negative in the case of the Netherlands (especially until 2019). Among the countries reporting a positive bias, in both periods the largest bias is observed for Germany, while France shows the lowest positive bias. Regarding the timeline, the first two years of the COVID-19 pandemic lead, in absolute terms, to a small drop in the bias in the country group and in all countries except Spain.

	Representativity	**Data vintage**	**Total**	**Period**
Germany	0.037	0.041	0.079	
France	0.018	0.026	0.044	
Italy	0.039	0.025	0.064	
Spain	0.055	0.015	0.070	2012-2021
Netherlands	0.039	-0.047	-0.008	
Big-5	0.035	0.025	0.060	
Euro Area	0.019	-	-	
Germany	0.044	0.049	0.093	
France	0.027	0.029	0.056	
Italy	0.031	0.034	0.066	
Spain	0.042	0.025	0.067	2012-2019
Netherlands	0.040	-0.069	-0.029	
Big-5	0.037	0.030	0.067	
Euro Area	0.022	-	-	

Table 2: *MD* (in percent of a Törnqvist index with final NA weights), 2012-2021 and 2012-2019.

The representativity component of the bias is positive for the Big-5 aggregate and all individual countries in this group. The data vintage component of the bias is also positive in the Big-5 aggregate as well as in Germany, France, Italy and Spain. For the Netherlands, however, the data vintage contribution to the bias is negative and, in absolute terms, strongest among all countries under review.

In the Big-5 aggregate as well as in Germany and France, the representativity and data vintage components contribute to the total bias in roughly equal shares. In the case of Italy this is only true until 2019. For Spain, the data vintage component is significantly smaller than the representativity component, particularly over the complete sample period. In the case of the Netherlands, we observe that, until 2019, the (positive) representativity contribution to the bias is, in absolute terms, nearly half the size of the (negative) data vintage contribution.

Including the years 2020 and 2021 in the analysis leads to an alignment of both components. Concerning the data vintage component of all countries and the Big-5 aggregate, the inclusion of those years ends up in a decrease, in absolute terms, of the data vintage component. This is due to the change of the updating procedure when calculating the HICP weights (see Section 2.2). In contrast, the effect on the representativity component differs from country to country. In Germany and France the representativity component decreases, whereas it increases in Italy and Spain. In the Netherlands and the Big-5 aggregate, no substantial change between the time periods can be observed.

The systematic difference between the Netherlands and the remaining countries under review deserves some explanation. It turns out to be related to the fact that before 2021 the Dutch statistical office chose the not-to-price-update option whereas the other statistical offices generally applied the price-updating one (recall Section 2.2). At first glance, the waiver of price updating from $y-2$ to $y-1$ turns out to be advantageous as the resulting negative data vintage effects "correct" the positive representativity bias while, with price updating, data vintage effects tend to reinforce mismeasurement at the upper level of aggregation. Theory suggests that it is a matter of own and cross-product price elasticities how the expenditures of individual goods and services respond to price changes. The assessment of whether the presence or absence of price updating is justified, thus, generally requires (empirical) knowledge about the responsiveness of all goods and services to price changes. In addition, the internal consistency of statistical procedures may also be an aspect to be considered given that, according to HICP weight updating rules, the expenditures for the annual average $t-1$ have to be price-updated to December $t-1$.

Inaccuracy. The uncertainty surrounding inflation measurement due to upper-level aggregation issues is measured by two statistical metrics, the total *RMSD* and the interdecile range (see Tab. 3). As a key takeaway from this analysis, it might be worth memorising that the interdecile range of the Big-5 aggregate is about one-tenth of a percentage point for each period under review. While the interdecile range of the aggregate is nearly unaffected by the inclusion of 2020 and 2021, it increases at the country level by adding the COVID-19 period (especially in the Netherlands, where it almost doubles its size). The interdecile range is highest for the German HICP in the period 2012-2019 and the Dutch HICP shows the highest range for the period 2012-2021, which includes the first two years of the COVID-19 phase. While Italy and Spain follow quite closely, the distance to France is more marked.

	RMSD			***IDR***	**Period**
	Representativity	Data vintage	Total		
Germany	0.057	0.058	0.106	0.181	
France	0.042	0.036	0.068	0.135	
Italy	0.057	0.046	0.091	0.158	
Spain	0.091	0.047	0.102	0.177	2012-2021
Netherlands	0.080	0.083	0.078	0.206	
Big-5	0.045	0.033	0.073	0.104	
Euro Area	0.030	-	-	-	
Germany	0.062	0.064	0.116	0.166	
France	0.035	0.040	0.070	0.115	
Italy	0.043	0.051	0.090	0.145	
Spain	0.052	0.045	0.088	0.149	2012-2019
Netherlands	0.052	0.081	0.060	0.125	
Big-5	0.043	0.037	0.076	0.102	
Euro Area	0.028	-	-	-	

Table 3: *RMSD* and *IDR* (in percentage points p.a.), 2012-2021 and 2012-2019.

For the Big-5 aggregate as well as for Germany, France and Italy, we observe that the *RMSD* of the representativity component and the *RMSD* of the data vintage component do not differ markedly in size for both periods under consideration. Until 2019, this also applies to Spain but not to the Netherlands, whereas it applies to the Netherlands but not to Spain over the complete sample until 2021. With regard to the total *RMSD*, no substantial difference can be observed for Germany, France, Italy and the country group between the two periods 2012-2021 and 2012-2019. For Spain and the Netherlands, however, the total *RMSD* increases from 0.088 to 0.102 (Spain) and from 0.060 to 0.078 (Netherlands) when adding the COVID-19 period to the analysis. The first years of the COVID-19 pandemic do not have similar effects on the *RMSD* of the bias in the countries under review. While the data vintage component declines slightly for most of the countries (except Spain and the Netherlands) as well as for the Big-5 aggregate, the representativity component rises (except for Germany) by including the years 2020 and 2021.

3.2 Long-run evidence on the representativity component

As stated in the previous section, the full-fledged analysis is restricted to the period until 2021 due to the implementation of the new COICOP 2018 classifications in the course of the national accounts benchmark revision in 2024. In addition, the full-fledged analysis cannot be extended to the years before 2012 because it is infeasible to calculate final NA weights. In principle, it would be very cumbersome but not per se impossible to gather the vintage data sets for the years prior to 2012. The infeasibility rather comes as a consequence of the flexibility statistical offices were granted in the compilation of HICP weights at that time. From the outset of the HICP until 2011, European regulation had imposed minimum standards while the harmonised weight updating procedure with a systematic use of national accounts data was implemented only in 2012 (Eiglsperger and Schackis, 2009; ECB, 2012). Of course, without this weight updating rule, final NA weights are impossible to construct, as they result from plugging in timely and more mature national accounts data in a weight updating formula which was given birth as a general standard only with the 2012 methodological change.

A look at the representativity part of upper-level aggregation over the entire HICP history is worthwhile nonetheless. Complementing the evidence of Herzberg et al. (2023), we report the results for bias and inaccuracy measures for the Big-5 aggregate and all individual countries in this group. Three issues are in the spotlight. The first is the iden-

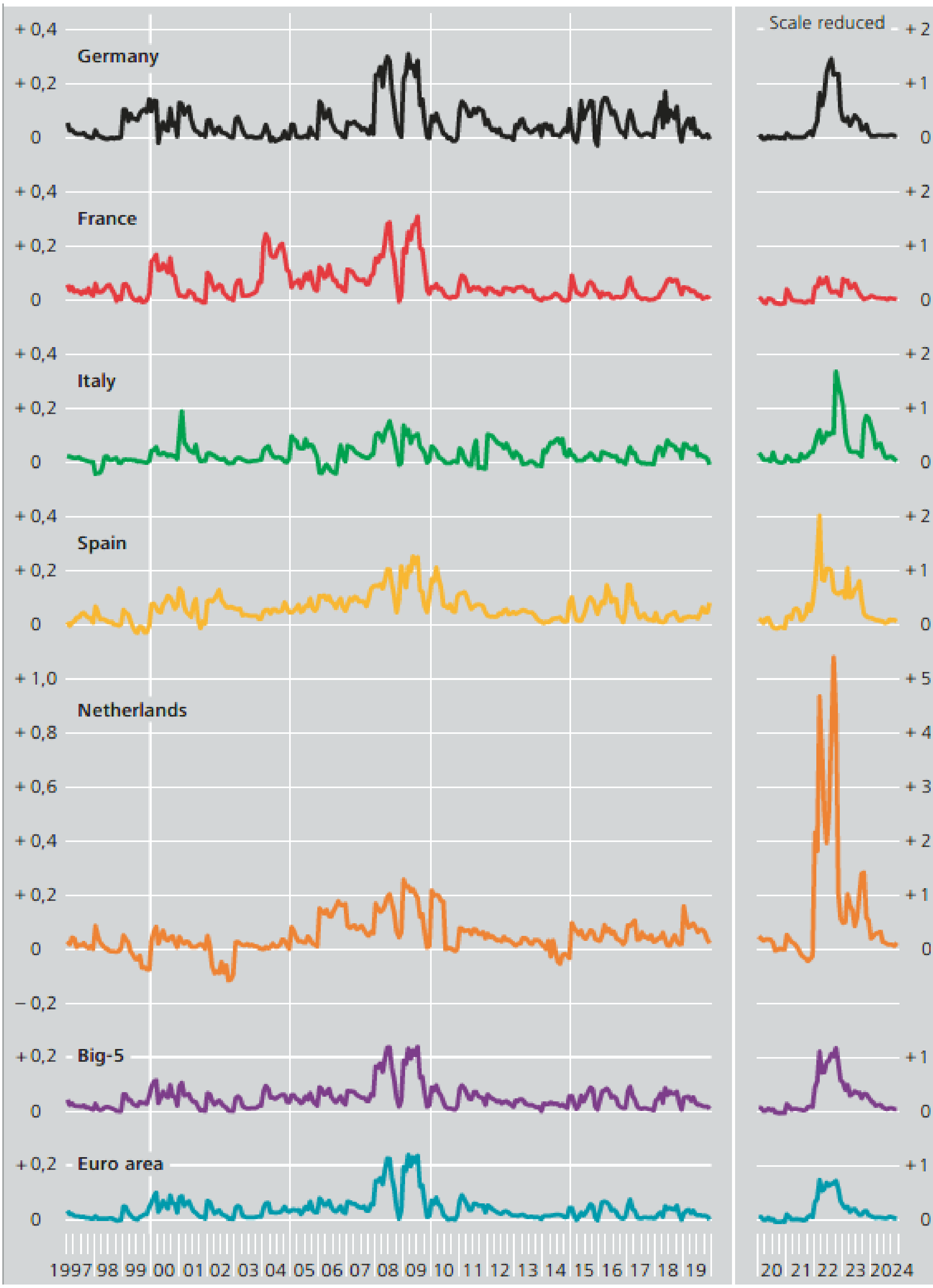

Figure 2: Representativity component (in percentage points p.a.) of several HICPs, 1997-2024.

tification of cyclical patterns in upper-level HICP mismeasurement. The second addresses the question whether, and to what extent, the 2012 methodological change has led to a reduction in the representativity bias, while the third is related to the effect of the high inflation period.

In Fig. 2, the time series of the monthly logarithmic deviation of the HICP from the Törnqvist index using official weights are plotted for the euro area, the Big-5 aggregate and all individual countries in this group over the period from 1997 to 2024. The plots for the euro area and the Big-5 aggregate seem to closely resemble each other. Across individual countries, some differences in the time profiles are observed. For the period 2020-2024, however, we reduced the scale due to the substantially higher values for the representativity component.

Until 2019, with (local) maxima uniformly detected in the years of the Financial Crisis 2008/2009, however, a visible commonality is worth reporting, too. In this period, inflation mismeasurement due to imperfect representativity peaked at about one-quarter of a percentage point in the euro area and the Big-5 aggregate. With three-tenths of a percentage point, the peak was markedly higher in Germany and France whereas, in Italy, it was lower at around one-sixth of a percentage point. The multitude and size of relative price shifts which appear in economically turbulent times such as the Financial Crisis may well explain why the representativity bias was above average. The high inflation period, which was primarily driven by energy prices, led to extraordinary peaks in the representativity component. The highest peak can be observed for the bias of the Dutch HICP. Among all five countries, the Netherlands experienced by far the highest inflation rates for energy during the high inflation period 2022 to 2023, apparently resulting in a substantially higher representativity bias for the Netherlands compared to the other countries.

In Tab. 4, the mean deviations, averaged over the complete sample from 1997 to 2024, as well as for selected subsamples of interest are reported. The comparison of the period 1997 to 2011 with the period from 2012 to 2019 reveals that the 2012 methodological change reduced the size of the representativity bias in the country groups and all countries under review except Italy. The most significant reduction of bias can be observed for the French HICP, whereas the smallest reduction is found for the Dutch HICP. The counterintuitive result reported for Italy seems to be due mainly to the extraordinarily low mean deviation in the pre-2012 period while the estimate for the representativity bias in the period since 2012 turns out to fit the results of the other countries well. There might have been special factors in the compilation of the Italian HICP in the pre-2012 period that dampened the size of the representativity bias.

As already shown in Fig. 2, the high inflation period in 2022 and 2023 has a significant influence on the representativity bias. During this period, the mean bias is highest in the Netherlands, reaching almost two percentage points. For the Big-5 aggregate as well as for Germany and Italy, the bias is about 0.6 percentage points, while it amounts to 0.7 percentage points in Spain and only to 0.2 percentage points in France. Over the complete sample, the representativity bias is about one-tenth of a percentage point for all countries and country groups, except for the Netherlands, where it is nearly one-fifth of a percentage point due to the substantial increase in bias during the high inflation period.

Periods of high inflation represent a unique economic environment, typically resulting in higher distortions of relative prices (Adam et al., 2024) where consumers may substitute more expensive goods with cheaper alternatives. This leads to more significant shifts in consumption patterns. These changes, however, are not captured in the fixed weights of the Laspeyres index, resulting in a larger representativity bias.

	Germany	**France**	**Italy**	**Spain**	**Netherlands**	**Big-5**	**Euro Area**
1997-2024	0.086	0.061	0.078	0.111	0.177	0.088	0.059
1997-2011	0.057	0.073	0.026	0.071	0.048	0.056	0.047
2012-2019	0.044	0.027	0.031	0.042	0.040	0.037	0.022
2020-2024	0.242	0.082	0.305	0.345	0.780	0.267	0.153
2022-2023*	0.578	0.207	0.580	0.718	1.832	0.594	0.361

*High inflation period

Table 4: *MD* (in percent of a Törnqvist index) of representativity component, 1997-2024.

The results for the *RMSD* of the representativity component, reported in Tab. 5, reveal that the uncertainty surrounding the HICP was reduced by the 2012 methodological change. This was quite substantial in the case of the euro area. Progress in the

French HICP contributed the most while a more moderate decline is observed for the German, the Spanish and the Dutch HICP. In these countries, the pre-2012 levels had been comparatively high.

In addition, the uncertainty surrounding the HICP is also affected by the high inflation period. For the Big-5 aggregate, as well as for Germany, Italy and Spain the *RMSD* fluctuates between 0.66 and 0.82 percentage points during the period from 2022 to 2023, while it is about 0.2 percentage points in France. Once again, the highest value is observed for the Dutch HICP. Regarding the country groups, the uncertainty for the Big-5 aggregate is about 50% higher than that for the euro area.

	Germany	**France**	**Italy**	**Spain**	**Netherlands**	**Big-5**	**Euro Area**
1997-2024	0.206	0.101	0.204	0.235	0.635	0.189	0.126
1997-2011	0.091	0.100	0.047	0.089	0.088	0.075	0.069
2012-2019	0.062	0.035	0.043	0.052	0.052	0.043	0.028
2020-2024	0.455	0.157	0.472	0.529	1.495	0.424	0.271
2022-2023*	0.719	0.240	0.710	0.816	2.355	0.664	0.427

*High inflation period

Table 5: *RMSD* (in percentage points p.a.) of representativity component, 1997-2024.

4 Conclusion

We present evidence on HICP mismeasurement at the upper-level of aggregation for the five largest euro area countries and the country group as a whole. During periods of moderate inflation, the upper-level aggregation bias of the HICP, consisting of the representativity and the data vintage component, is generally small. In the recent high inflation period 2022 and 2023, however, the bias was very large as a result of the substantial increase in the representativity component.

The results of the full-fledged analysis from 2012 to 2021 show that neither the Laspeyres formula nor the updating of weights with preliminary national accounts data are major sources of bias or inaccuracy in moderate inflation phases. For the Big-5 aggregate, the mean deviation of the HICP from a superlative index, based on weights being updated with timely and more mature national accounts data, clearly falls short of one-tenth of a percentage point and the interdecile range of the deviations has a length of about one-tenth of a percentage point. During the recent high inflation period, the representativity bias was on average about three-fifths of a percentage point for the Big-5 aggregate and about one-third of a percentage point for the euro area, and thus substantially higher than in all other periods before 2022. This underlines that the representativity component remains an important factor to be considered in inflation measurement.

The upper-level of aggregation is one stage of HICP production where mismeasurement might occur. Hence, the results of this study enlighten only a part of a multi-faceted picture and, according to existing knowledge in this field, this part is likely to be quantitatively less important than other sources of mismeasurement. In ECB (2021, chapter 3), a comprehensive evaluation framework is sketched out and patchy evidence is presented. However, an overall assessment in the style of the Boskin Report is still lacking for the HICP.

As regards mismeasurement at the upper level of aggregation, it is generally insufficient to look at the representativity component only. As already argued by Herzberg et al. (2023), the use of more current weights through the annual updating procedure implemented in 2012 has come at the cost of relying on preliminary national accounts data. This paper quantifies this additional source of mismeasurement for a country group representing more than four-fifths of the euro area HICP. However, the benchmark can only be specified in terms of final NA weights, implying that the estimates of the data vintage components may be interpreted as a lower bound for the "true" impact of preliminary data in weight updates (which would be better proxied using full-information weights as carried out in Herzberg et al. (2023) for the German HICP).

The results of this paper let us conclude that the representativity and data vintage components contribute in fairly equal parts to the total upper-level aggregation bias and inaccuracy in the period from 2012 to 2021. The starting date for the full-fledged analysis is forced by the concept of final NA weights which only makes sense to be applied in the period since the implementation of the annual updating of weights. With regard to the effects of this methodological change, we can therefore provide partial evidence, namely that the representativity component has been reduced. Knowledge about weight updating practice in the pre-2012 era when weight compilation methods were rather non-harmonised, apart from imposing some minimum standards, would be needed to extend the analysis in this direction.

The results also comprise first COVID-19 evidence for the mismeasurement at the upper-level of aggregation. The effect on the representativity component varies across countries. In the Big-5 aggregate, the inclusion of these years has no impact on this sort of mismeasurement. Concerning the data vintage component for all countries and the Big-5 aggregate, the inclusion of the years 2020 and 2021 results in a slight decrease, in absolute terms, of the data vintage component. This could be due to the 2021 modification of the updating procedure using preliminary national accounts data on consumption expenditures from $y-1$ as a response to the strong consumption shifts during the COVID-19 pandemic. Since 2025, some national statistical offices returned to basing their weights primarily on national accounts data for consumption expenditures from $y-2$, while others continued including data for $y-1$ in this annual weight update (Eurostat, 2024, Chapter 3.5). In this regard, our results suggest that using preliminary data from $y-1$ does not introduce an additional source of error for inflation measurement but even slightly reduces the data vintage bias. Consequently, a standardised approach to the weight update procedure based on preliminary but more recent data for $y-1$ could be implemented. This would not only lead to a better comparability but could also potentially result in slightly lower susceptibility to inflation mismeasurement.

References

Adam, K., Alexandrov, A., and Weber, H. (2024). Inflation distorts relative prices: Theory and evidence. Deutsche Bundesbank Discussion Paper 23/2024, Deutsche Bundesbank.

Armknecht, P. and Silver, M. (2014). Post-Laspeyres: The case for a new formula for compiling consumer price indexes. Review of Income and Wealth, 60(2):225–244.

Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z., and Jorgenson, D. W. (1998). Consumer prices, the consumer price index, and the cost of living. Journal of Economic Perspectives, 12(1):3–26.

ECB (2012). New standards for HICP weights. Monthly bulletin, European Central Bank. April 2012, Box 3, 36-39.

ECB (2021). Inflation measurement and its assessment in the ECB's monetary policy strategy review. Occasional Paper Series 265, European Central Bank.

ECB (2025). The ECB's monetary policy strategy statement. https://www.ecb.europa.eu/mopo/strategy/strategy-review/ecb.strategyreview202506_strategy_statement.en.pdf. Accessed: 15 July 2025.

Eiglsperger, M. and Schackis, D. (2009). Weights in the harmonised index of consumer prices: Selected aspects from a user's perspective. 11th Ottawa Group Meeting.

EU (2020). Commission Implementation Regulation (EU) 2020/1148 of 31 July 2020 . Official Journal of the European Union.

Eurostat (2020). Guidance on the compilation of HICP weights in case of large changes in consumer expenditures. https://ec.europa.eu/eurostat/documents/10186/10693286/Guidance-on-the-compilation-of-HICP-weights-in-case-of-large-changes-in-consumer-expenditures.pdf. Accessed: 9 September 2025.

Eurostat (2024). Harmonised Index of Consumer Prices (HICP): Methodological Manual. European Union, Luxembourg.

Greenlees, J. S. and Williams, E. (2010). Reconsideration of weighting and updating procedures in the US CPI. Jahrbücher für Nationalökonomie und Statistik, 230(6):741–758.

Herzberg, J., Knetsch, T., Schwind, P., and Weinand, S. (2023). Quantifying bias and inaccuracy of upper-level aggregation in the Harmonised Index of Consumer Prices for Germany and the euro area. Review of Income and Wealth, 69(3):581–605.

Hoffmann, J. (1998). Problems of inflation measurement in Germany. Discussion Paper 01/1998, Deutsche Bundesbank.

International Monetary Fund, International Labour Organization, Organisation for Economic Co-operation and Development, European Union, United Nations, and The World Bank (2020). Consumer price index manual: Concepts and methods.

Knetsch, T. A. and Reimers, H.-E. (2009). Dealing with Benchmark Revisions in Real-Time Data: The Case of German Production and Orders Statistics. Oxford Bulletin of Economics and Statistics, 71(2):209–235.

Lequiller, F. (1997). Does the French consumer price index overstate inflation? INSEE Série des documents de travail de la Direction des Etudes et Synthèses Économiques, 3.

Lowe, J. (1823). The present state of England in regard to agriculture, trade, and finance. Second Edition. London: Longman, Hurst, Rees, Orme and Brown.

Neves, P. D. and Sarmento, L. M. (1997). The substitution bias of the consumer price index. Banco de Portugal Economic Bulletin, pages 25–33.

Silver, M. and Ioannidis, C. (1994). The measurement of inflation; untimely weights and alternative formulae: European evidence. The Statistician, 43(4):551–562.

A Extended results for Germany

The results in Tab. A1 and A2 include the period from 2012-2022 for Germany. Regarding the measurement bias, Tab. A1 reveals a substantially higher overall bias for the period 2012-2022 compared to the period until 2021. While the data vintage component decreased when including the year 2022, the value for the representativity component increased significantly, rising from 0.04 for the period 2012-2021 to 0.12 for the period up to 2022. Apparently, the high inflation phase beginning in 2022 had a substantial impact on the representativity bias and, consequently, on the overall bias.

	Representativity	**Data vintage**	**Total**	**Period**
	0.115	0.036	0.151	2012-2022
Germany	0.037	0.041	0.079	2012-2021
	0.044	0.049	0.093	2012-2019

Table A1: *MD* (in percent of a Törnqvist index with final NA weights) with 2012-2022 period for Germany.

Similar findings can be observed for the uncertainty surrounding the HICP as shown in Tab. A2. While the RMSD for the data vintage component remains nearly constant when adding the year 2022 to the analysis, the RMSD for the representativity bias increases considerably. Consequently, the total RMSD for Germany is also higher compared to the period ending in 2021. Additionally, the interdecile range increases by approximately one-tenth of a percentage point when the year 2022 is included in the analysis.

	RMSD			***IDR***	**Period**
	Representativity	Data vintage	Total		
	0.300	0.056	0.304	0.277	2012-2022
Germany	0.057	0.058	0.106	0.181	2012-2021
	0.062	0.064	0.116	0.166	2012-2019

Table A2: *RMSD* and *IDR* (in percentage points p.a.) with 2012-2022 period for Germany.