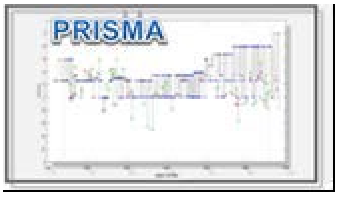
EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Teresa Messner, Fabio Rumler

Inflation heterogeneity across Austrian households. Evidence from household scanner data



No 2894

Price-setting Microdata Analysis Network (PRISMA)

This paper contains research conducted within the Price-setting Microdata Analysis Network (PRISMA). PRISMA consists of economists from the ECB and the national central banks (NCBs) of the European System of Central Banks (ESCB).

PRISMA is coordinated by a team chaired by Luca Dedola (ECB), and consisting of Chiara Osbat (ECB), Peter Karadi (ECB) and Georg Strasser (ECB). Fernando Alvarez (University of Chicago), Yuriy Gorodnichenko (University of California Berkeley), Raphael Schoenle (Federal Reserve Bank of Cleveland and Brandeis University) and Michael Weber (University of Chicago) act as external consultants.

PRISMA collects and studies various kinds of price microdata, including data underlying official price indices such as the Consumer Price Index (CPI) and the Producer Price Index (PPI), scanner data and online prices to deepen the understanding of price-setting behaviour and inflation dynamics in the euro area and EU, with a view to gaining new insights into a key aspect of monetary policy transmission (for further information see https://www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_prisma.en.html)

The refereeing process of this paper has been co-ordinated by a team composed of Luca Dedola (ECB), Anton Nakov (ECB), Chiara Osbat (ECB), Elvira Prades (Banco d'Espana), Sergio Santoro (ECB), Henning Weber (Bundesbank).

This paper is released in order to make the results of PRISMA research generally available, in preliminary form, to encourage comments and suggestions prior to final publication. The views expressed in the paper are the author's own and do not necessarily reflect those of the ESCB.

Abstract

It has been widely documented that households experience different inflation rates which are generally concealed in aggregate price indices. Using scanner data from a large household panel for Austria, we analyse price dynamics faced by individual households and try to explain the causes for the observed inflation differences. Considering not only consumption shares but also the specific product prices paid by households, we find a considerable and persistent degree of heterogeneity among household inflation rates. These are also quite variable over time, resulting from varying consumption baskets and active product substitution, allowing households to reduce their inflation exposure substantially. Factors like age and shopping behavior of households explain some of the inflation differences, whereas income does not seem to have a notable influence in normal times. However, during high inflation periods, the lowest income group is found to face higher inflation rates than other income groups.

JEL classification: D12, D30, E31

Keywords: Household inflation, heterogeneity, micro data

Non-technical summary

Monitoring price developments typically involves aggregate statistics like the Harmonized Index of Consumer Prices (HICP). However, households rarely experience price changes uniformly. Variation in inflation rates at the household level can occur because households consume different combinations of goods and services and also pay different prices for the same items depending on where and when they buy them.

Several studies have investigated how inflation affects certain groups of households differently. These studies largely analyze price changes for various consumption bundles derived from household budget surveys. They consistently find that inflation rates vary significantly across households. Typically, elderly, blue-collar, and low-income households tend to experience higher inflation rates due to their larger expenditure shares of food and energy in total consumption.

More recent research has re-examined household inflation heterogeneity using more granular data, such as household scanner data. This type of data allows researchers to account for variations in both the products households choose and the prices they pay for those products. Additionally, it provides detailed transaction information along with individual household characteristics. These analyses revealed a sizable degree of heterogeneity among household inflation rates – notably more pronounced than previously documented and, furthermore, found that inflation differences tend to persist over time, although individual household inflation rates can be quite volatile.

Using such a dataset for Austria spanning from 2008 to 2018, we explore the extent and causes of household-level inflation disparities. Our findings reveal substantial inflation heterogeneity. Over our sample period, we find that the range of inflation rates across households, measured as the interquartile range, amounted to approximately 6 percentage points on average. Despite this, the average and median household inflation rates closely mirror the aggregate inflation rate. This suggests that while aggregate price indices conceal much of the variation across households, they accurately represent the average inflation rates of all households.

Our data allow us to consider both, differences in the products bought and prices households paid for the same product, as a source of heterogeneity. We find that the majority of the variation in household inflation rates stems from price differences for identical products. Calculating household inflation rates with average product prices, as opposed to household-specific prices, substantially decreases the heterogeneity. Interestingly, we also find that a household's inflation rate is almost uncorrelated over time, suggesting that households are not necessarily stuck but move within the distribution of inflation rates from one period to another.

Beyond statistical explanations for inflation differences (i.e. price and consumption variations), we find limited evidence that sociodemographic characteristics and household shopping behavior can explain these differences. Factors such as the age of household members, search intensity, preferences for private label products, and store choices can account for a small fraction of the observed heterogeneity. Surprisingly, household income appears largely uncorrelated with individual inflation rates. Only during periods of high inflation, low-income groups experienced higher inflation rates than higher-income groups. This could be due to substitution of relatively more expensive products by cheaper ones, which is a common strategy to limit a household's inflation exposure.

On average, households are able to reduce their inflation rates by 0.4 percentage points through product substitution. However, households do not always substitute in the expected direction; a sizeable fraction of households are shown to substitute from cheaper to more expensive products. This might, however, be related to the mostly low inflation period under study, as substitution patterns become less ambiguous when inflation increases.

Our research highlights the persistent nature of inflation disparities across households and their potential impact on real individual income. The ability of households to substitute plays a critical role in mitigating their exposure to inflation.

1 Introduction

Price developments are commonly monitored using aggregate statistics such as the Harmonized Index of Consumer Prices (HICP). However, households can be affected by price changes in quite different ways. They consume different compositions of goods and services, and often pay different prices for the same good or service depending on where and when they buy it (Kaplan and Menzio, 2015).

Many of the early studies on inflation heterogeneity (e.g. Michael, 1979; Hagemann, 1982; Hobijn and Lagakos, 2005; Fessler and Fritzer, 2013; and more recently Gürer and Weichenrieder, 2020) study aggregate price changes for different, group-specific consumption bundles derived from household budget surveys. These studies document sizeable differences in inflation rates across households. A common finding is that elderly, blue-collar and low-income households experience higher inflation rates in times of rising overall inflation due to higher food and energy consumption shares. The results also indicate that, overall, inflation heterogeneity is a quite persistent phenomenon, while household-specific inflation rates are rather volatile.

In contrast to the studies using price index data at the product category level and varying expenditure shares across households, Kaplan and Schulhofer-Wohl (2017) and Argente and Lee (2021) are among the few papers employing household scanner data for their analyses of US household inflation rates. Using such data enables them to (i) account for heterogeneity in the choice of specific products as well as for (ii) heterogeneity in prices paid for these products. Furthermore, (iii) such panels also provide actual transactions based on unique product identifiers and information on individual household characteristics. In particular, accounting for differences in prices paid, Kaplan and Schulhofer-Wohl (2017) find substantially larger heterogeneity among household inflation rates and more persistent deviations of individual inflation rates from the aggregate. They also confirm earlier findings on the negative correlation between income and household inflation rates in periods of rising energy and food prices. Moreover, Argente and Lee (2021) find that inflation rates between households at the lower and higher end of the income distribution diverged substantially during the Global Financial Crisis (GFC). Almost half of the gap between households at both ends of the distribution can be attributed to differences in prices paid, one third to substitution within a product category and about 13% to households' shopping behaviour. They also find evidence that wealthier households were able to change their consumption behaviour more easily during the crisis by switching to cheaper low-quality goods.

Overall, this points to inflation affecting households differently depending on their access to different supermarkets, differently priced products, and on the ability to substitute. Consequently, knowing the extent of divergence in individual inflation rates is essential for monetary policy and its attempt to control aggregate inflation.

In this paper, we also use transaction data from a large household panel in order to determine the extent and causes of inflation differences at the household level in Austria. While we find a substantial degree of inflation heterogeneity driven by the differences in consumption and specific product prices, we find it difficult to explain much of this heterogeneity with sociodemographic factors and factors related to shopping behaviour of households. We can link household inflation to age and the household's shopping behaviour, while income does not seem to play such a prominent role as in previous studies. The latter holds for most of our sample period which is characterized by relatively low and stable inflation rates. However, during periods of high inflation the lowest income group faces higher inflation rates than all other income groups. We also find a considerable degree of product substitution at the household level in response to relative price increases. Without this substitution, the average inflation rate over our sample period would have been nearly half a percentage point higher.

Our paper contributes, on the one hand, to research on distributional aspects of inflation and monetary policy. It is related to studies arguing that persistent inflation inequality across households contributes to real income inequality (e.g. Argente and Lee, 2021; Beck and Jaravel, 2020; Gürer and Weichenrieder, 2020). Persistent inflation differences across household groups can even affect the transmission of monetary policy as prices of the products consumed by high-income households are found to respond less to monetary policy shocks than those consumed by lower-income households (Cravino et al., 2020; Strasser et al., 2023). By discussing the methods used to construct price indices with scanner data and by attempting to quantify the substitution bias of a fixed-base index, our paper is also related to the literature on inflation measurement (see e.g. Ivancic et al., 2011; Handbury et al., 2013; Moulton, 2018).

The remainder of the paper is structured as follows: Section 2 introduces the dataset underlying our analysis and describes the methodology of calculating household-specific inflation rates from household scanner data. Section 3 presents comprehensive evidence on the heterogeneity among household inflation rates, followed by an attempt to explain this heterogeneity with different household-level variables in section 4. Section 5 presents further evidence on product substitution and, finally, section 6 discusses the main results and draws some conclusions.

2 Data and calculation of household inflation rates

The analyses in this paper are based on the GfK (Gesellschaft für Konsumforschung) household panel data for Austria. The dataset consists of purchases that households registered over time and includes about 22.9 million transactions of 288,261 different products uniquely identified through their barcode purchased by 13,175 households in Austria over the years 2008 to 2018.[1]

The products included in the dataset comprise so-called Fast-Moving Consumer Goods (FMCG), i.e. mainly items that can be bought in supermarkets, such as food and beverages, products for household maintenance and gardening, personal care items and pet food. In total, these products cover about 15% of the Austrian CPI basket. Table A2 in Appendix A compares spending across COICOP-4 groups in our dataset with the structure of the Austrian CPI.[2] As expected, regularly bought food items (such as milk, cheese and eggs) and non-alcoholic beverages are over-represented in the household panel data, whereas most other COICOP groups appear to be under-represented relative to the Austrian CPI. Despite the low coverage, inflation rates calculated from the GfK data are found to be broadly correlated with aggregate CPI inflation in Austria and strongly correlated with inflation of the CPI-subindex for food and beverages (see Figure 1).

Apart from information on the transactions (product price, quantity and product characteristics such as barcode, brand, manufacturer, private label), the dataset also contains information on household characteristics (age of the panelist, household size, social class, household income and the location of residence of the household) and the date, name and type of the retailer where the transaction took place.[3]

Following Kaplan and Schulhofer-Wohl (2017), we calculate household-specific inflation rates by accounting for heterogeneity in both, the composition of products consumed and in the prices paid for each product by households. To calculate year-on-year inflation rates at the household level we can only consider those products that have been purchased by the same household at least once in quarter t and the same quarter in the previous year, $t-4$.[4] If a product was purchased more than once in a quarter by a given household, we used the (volume-weighted) average price during the quarter in the calculation of the household's inflation rate. In order to reduce the sampling error we restrict our sample – in

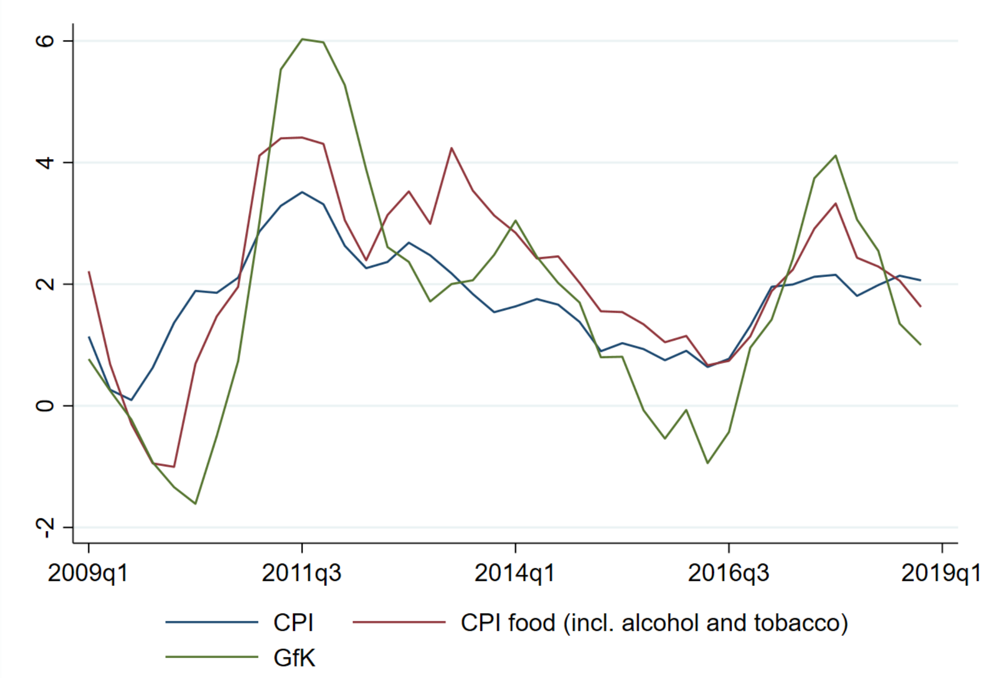
[1] See Table A1 in Appendix A for the breakdown of the dataset over time.

[2] COICOP stands for Classification of Individual Consumption by Purpose. For information on the COICOP classification see https://unstats.un.org.

[3] For more information on the data, see Appendix A.

[4] We chose to calculate inflation rates for the quarterly frequency as the number of matched products over time would have been much lower for inflation rates calculated at monthly frequency.

Figure 1: Aggregate inflation rates



Notes: Year-on-year inflation rates at quarterly frequency in percent. Data source for CPI and CPI food is Statistik Austria. Inflation based on GfK data is calculated according to equation (9).

the baseline case – to those households who bought at least five matched products in the two quarters.[5] These restrictions reduce our sample from almost 23 million transactions to around 4.4 million transactions that are included in the calculation of the household inflation rates.[6]

Let $q_{ij,t}$ be the quantity of product j bought by household i in quarter t and $p_{ij,t}$ the price paid for it, then the (gross) inflation rate between t and $t-4$ for household i according to the Laspeyres (L) and Paasche (P) indices are defined as

$$\pi_{i\ t,t-4}^{L} = \frac{\sum_j p_{ij,t} q_{ij,t-4}}{\sum_j p_{ij,t-4} q_{ij,t-4}} \tag{1}$$

and respectively

$$\pi_{i\ t,t-4}^{P} = \frac{\sum_j p_{ij,t} q_{ij,t}}{\sum_j p_{ij,t-4} q_{ij,t}}. \tag{2}$$

[5]We also calculate alternative inflation rates based on a higher minimum number of matched barcodes, but the main results remain the same.

[6]For the distribution of the matched barcodes across COICOP groups see the last column of Table A2 in Appendix A.

Inflation according to the Fisher (F) index is simply the geometric mean of the Laspeyres and the Paasche indices:

$$\pi_{i\ t,t-4}^{F} = \sqrt{\pi_{i\ t,t-4}^{L}\pi_{i\ t,t-4}^{P}} \tag{3}$$

To check whether the composition of products or the prices paid for the products by the households contributes more to inflation heterogeneity, we additionally calculate inflation rates accounting only for heterogeneity in product composition. Accordingly, these household inflation rates are calculated using household-specific consumption, but average prices at the product (i.e. barcode, BC) level instead of household specific prices:

$$\pi_{i\ t,t-4}^{L,BC} = \frac{\sum_j \bar{p}_{j,t} q_{ij,t-4}}{\sum_j \bar{p}_{j,t-4} q_{ij,t-4}} \tag{4}$$

where $\bar{p}_{j,t}$ is the volume-weighted average price for product j across all households in quarter t. Inflation rates according to the Paasche and Fisher indices are defined analogously to equations (2) and (3).

Furthermore, to be able to compare our results with studies that only use the CPI basket and CPI prices to assess inflation heterogeneity (e.g. Fessler and Fritzer, 2013; Gürer and Weichenrieder, 2020), we also calculate inflation rates that are aggregated over more broadly defined product groups (called item strata). This requires matching the products bought by the households to the respective COICOP-5 digit group of the CPI. This, however, ignores differences in prices paid for products within item strata as a source of heterogeneity, implying that differences in the allocation of consumption across item strata is the only source of heterogeneity in resulting household inflation rates.

For the average prices at the item strata level we have two possible sources: the prices derived from our dataset and the the official price indices published by Statistik Austria (at the COICOP-5 level). Let's first define the share of household i's consumption of product j at initial date $t-4$ as:

$$s_{ij,t-4} = \frac{p_{ij,t-4} q_{ij,t-4}}{\sum_j p_{ij,t-4} q_{ij,t-4}} \tag{5}$$

Inflation at the household level according to the Laspeyres index using GfK consumption shares and CPI price indices is then:

$$\pi_{i\ t,t-4}^{L,CPI} = \sum_j s_{ij,t-4} \frac{p_{k(j),t}^{CPI}}{p_{k(j),t-4}^{CPI}} \tag{6}$$

where $k(j)$ refers to the item stratum (i.e. COICOP-5 group) that contains product j and $p_{k,t}^{CPI}$ is the CPI price of COICOP group k at time t. Paasche inflation rates are defined analogously, with the difference that the final date consumption share $s_{ij,t}$ instead of the initial date consumption share is used in the aggregation.

To calculate the corresponding inflation rates using stratum average prices from the GfK data instead of CPI prices, we first need to compute (Laspeyres-type) stratum level (S) inflation rates:

$$\pi_{k\ t,t-4}^{L,S} = \frac{\sum\limits_{i,j:j\epsilon k} \bar{p}_{j,t} q_{ij,t-4}}{\sum\limits_{i,j:j\epsilon k} \bar{p}_{j,t-4} q_{ij,t-4}} \tag{7}$$

which are then aggregated over item strata using the initial date shares of stratum k in household i's consumption at time $t-4$ ($s_{ik,t-4}$) as weights to arrive at the (Laspeyres) inflation rate at the household level:

$$\pi_{i\ t,t-4}^{L,S} = \sum_k s_{ik,t-4} \pi_{k\ t,t-4}^{L,S} \tag{8}$$

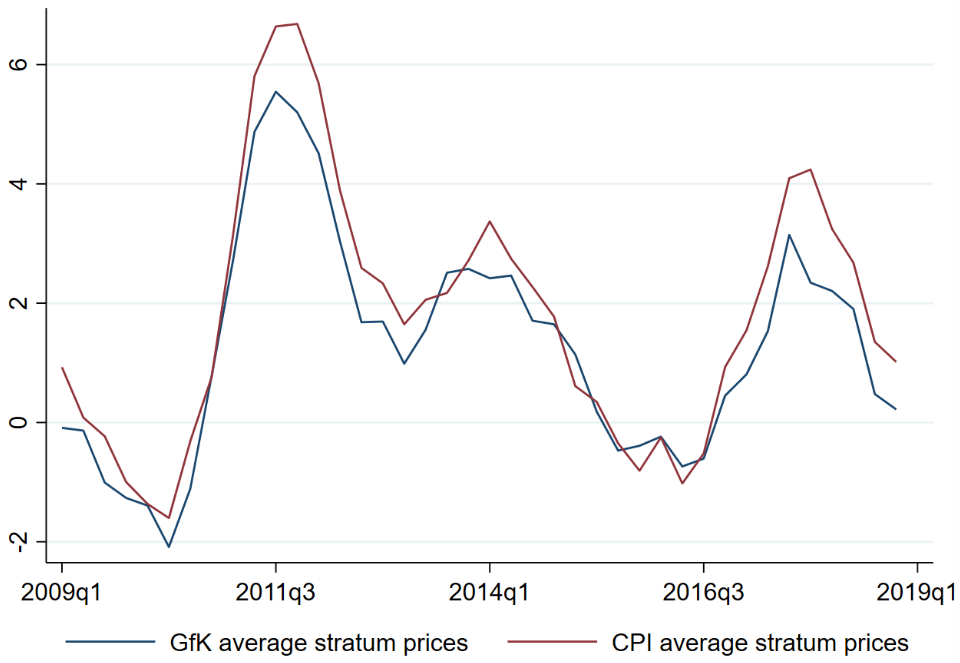
with analogous definitions of household inflation rates over item strata for Paasche and Fisher indices.

In order to assess whether the price information embedded in the GfK data is consistent with official CPI prices, we compare the inflation rate aggregated over COICOP-5 groups using GfK average stratum prices (blue line) and the inflation rate using CPI average stratum prices (red line) in Figure 2.[7] Reassuringly, in the aggregate, GfK prices and CPI prices convey the same information and are highly correlated.

So far, all calculations produce inflation rates at the household level. In order to compare the information in our dataset with the official inflation statistics from Statistik Austria, we lastly compute aggregate inflation rates using the CPI structure (i.e. COICOP groups) and CPI prices but with the consumption shares derived from our data. This amounts to

[7]Note that in contrast to Figure 1, only products that are consumed by households in quarter t and $t-4$ are included in the calculation of the inflation rates in Figure 2.

Figure 2: Inflation rates computed over CPI strata with GfK and CPI prices



Notes: The blue line corresponds to the aggregate of the household inflation rates defined in equation (8) and the red line to the aggregate of the household inflation rates defined in equation (6).

aggregating the household price indices with CPI prices $\pi_{i\ t,t-4}^{L,CPI}$ over all households:[8]

$$\pi_{t,t-4}^{L,CPI} = \frac{\sum_{i} \chi_{it-4} \pi_{i\ t,t-4}^{L,CPI}}{\sum_{i} \chi_{it-4}} \tag{9}$$

where χ_{it-4} is household i's expenditure at initial time $t-4$ of all goods that enter household i's price index:

$$\chi_{it-4} = \sum_{j} p_{ij,t-4} q_{ij,t-4} \tag{10}$$

The resulting aggregate inflation rate $\pi_{t,t-4}^{L,CPI}$ – together with aggregate CPI inflation and inflation of the CPI-subindex for food – is shown in Figure 1. As reflected by the co-movement of the two series, the correlation of GfK inflation (green line) with CPI food inflation (red line) is quite high – amounting to 0.85 – but given its standard deviation of about 2 percentage points, it is more volatile than CPI food inflation (with a standard

[8]We are using the Laspeyres formula here as the Austrian CPI is also a Laspeyres-type index.

deviation of 1.4 percentage points).

3 Distribution of household inflation rates

In this section, we investigate the variation in inflation rates across Austrian households. To get a first glance, we plot the kernel density estimates of the distributions of the household-level inflation rates across all households using household-level prices (as in equation (1)), barcode-average prices (as in equation (4)), stratum-average prices from GfK data (equation (8)) and from CPI statistics (equation (6)) for two selected quarters, namely 2011Q3 and 2018Q4, in Figure 3. We chose these two dates because 2018Q4 (the last quarter of the sample period) has the highest the number of observations and 2011Q3 features the highest inflation rate in our sample.

Both panels show that inflation rates are most dispersed, when considering the actual prices paid by the households (blue line), followed by the household inflation rates that are calculated using average prices at the product level (red line). When aggregating over item strata using average prices at the COICOP-5 level – either from the GfK data or from CPI statistics – the distribution of household inflation rates becomes much narrower. Looking more closely at the high inflation period in the right panel of Figure 3, reveals a generally wider distribution of household inflation rates than in the left panel for all four measures, although the order in terms of extent of heterogeneity remains the same.

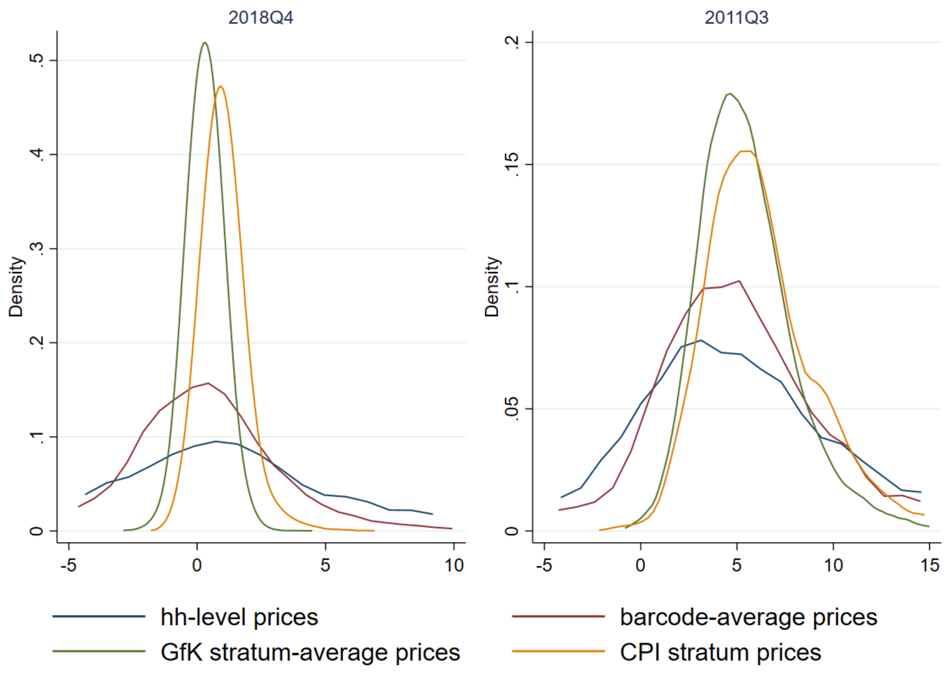
An interesting further finding from the figure is that the distribution of household inflation rates with CPI average prices is located to the right of the one calculated with GfK average prices. This indicates that the products actually consumed by households in our sample, on average, have lower inflation rates than the products surveyed in official CPI statistics.[9]

The dispersion measures calculated over the entire sample period, shown in Table 1, confirm the above results. The interquartile range of the inflation rates with household-level prices is almost double of the one using barcode-average prices and more than three times higher than when using stratum-average prices.

But where does this heterogeneity come from? By calculating the ratio of the variances between the different concepts of household inflation rates (bottom of Table 1), we find

[9]This result is not specific to the time periods chosen or to Austria (see Strasser et al., 2023). An explanation of this might be related to sales and the fact that Statistical Institutes – in accordance with conventions of constructing the CPI – are only allowed to take into account price discounts and promotions granted to all customers but not personal discounts, such as those in loyalty programs and the like; see the documentation of the HICP methodology by Eurostat.

Figure 3: Distribution of household-level inflation rates



Notes: Kernel density estimates using Epanechnikov kernel function with a bandwith of 0.5 on 100 points. Data include 2,365 households with at least 5 matched barcodes between 2017Q4 and 2018Q4 (left panel) and 2,035 households with matched barcodes between 2010Q3 and 2011Q3 (right panel). The plot is truncated at -5% and 10% (15%) in the left (right) panel.

that most of the variation in the household inflation rates comes from differences in the prices consumers pay for the same product. To put this in numbers, the variance of inflation rates calculated by allowing only the allocation of consumption between barcodes to vary, but not the prices, is about one third of the case when both are allowed to vary. When the allocation of consumption between more broadly defined item strata are allowed to vary, but not the respective prices, the variance of inflation rates is only 7% of the benchmark case when prices can vary within barcodes.

Plotting the interquartile range over time (Figure 4), we can track the evolution of inflation heterogeneity. While we do observe some cyclical movements in inflation dispersion that seem to be correlated with aggregate inflation, there is no trend movement over time, confirming the persistent nature of inflation heterogeneity documented in previous studies (e.g. Argente and Lee, 2021). This holds true regardless of how household inflation rates are calculated: The interquartile range of inflation rates with household-level prices is consistently about 50% higher than with barcode-average prices and three times as high as those with stratum-average prices and CPI stratum prices. Concerning the two concepts

Table 1: Dispersion of household inflation rates (average 2009Q1 to 2018Q4)

	mean	s.d.	min	max
Interquartile range				
household-level prices				
Laspeyres	6.09	0.43	5.18	7.31
Paasche	6.10	0.39	5.34	7.18
Fisher	5.92	0.38	5.09	7.05
barcode-average prices				
Laspeyres	3.98	0.52	3.25	5.42
Paasche	3.96	0.49	3.18	5.25
Fisher	3.85	0.49	3.17	5.16
stratum-average prices				
Laspeyres	1.78	0.74	0.69	3.82
Paasche	1.80	0.75	0.70	3.44
Fisher	1.74	0.72	0.66	3.47
CPI stratum prices				
Laspeyres	1.84	0.76	0.85	4.18
Paasche	1.87	0.78	0.87	3.97
Fisher	1.81	0.74	0.82	3.88
Ratio of variance of inflation rates with household-level prices to variance with...				
barcode-average prices				
Laspeyres	0.36	0.05	0.26	0.50
Paasche	0.37	0.05	0.27	0.50
Fisher	0.37	0.05	0.28	0.50
stratum-average prices				
Laspeyres	0.07	0.05	0.01	0.27
Paasche	0.07	0.05	0.01	0.25
Fisher	0.07	0.05	0.01	0.27
CPI stratum prices				
Laspeyres	0.07	0.06	0.01	0.32
Paasche	0.08	0.06	0.01	0.31
Fisher	0.08	0.06	0.01	0.32

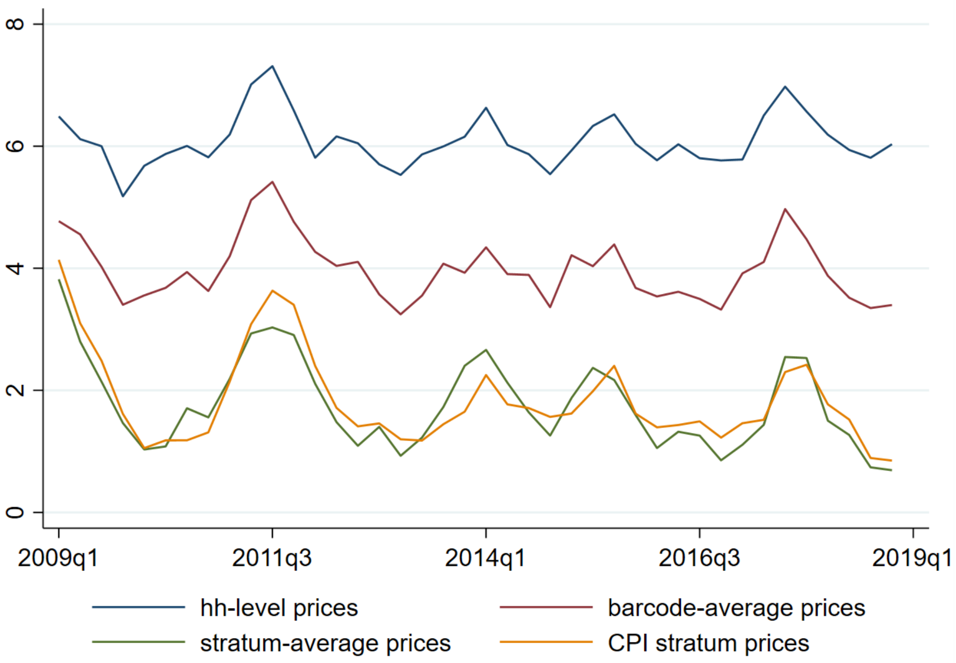
Notes: Mean, standard deviation, minimum and maximum of dispersion measures for each quarter over 2009Q1–2018Q4.

using item strata, inflation dispersion using GfK data and CPI data is very similar over time.

Figure 5 shows the distribution of the household-level inflation rates over time. It suggests that the mean and median household inflation rates closely correspond to the aggregate inflation rate calculated using CPI prices (defined in equation (9)), fluctuating between 0% and 5% most of the time. The mean, however, conceals a large amount of heterogeneity with household inflation rates ranging from +14% at the 90^{th} to -7% at the 10^{th} percentile. Furthermore, the entire distribution seems to shift roughly in parallel with aggregate inflation, indicating that inflation heterogeneity shows little dependence on the overall level of inflation.

While overall inflation heterogeneity appears to be stable and persistent over time, does the same also apply to the position of individual households in the distribution of inflation rates or do households move rather wildly within this distribution? The former scenario would imply that a household is essentially "stuck" in the distribution of inflation rates and always experiences either above-average or below-average inflation. The latter scenario,

Figure 4: Interquartile range of household inflation rates over time



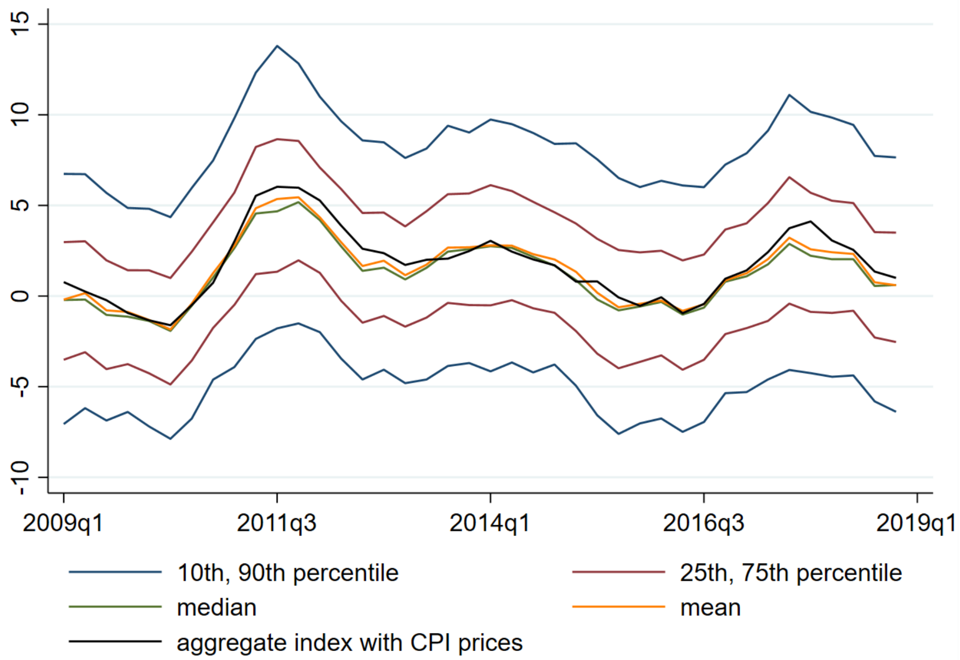
Notes: Interquartile range calculated from Laspeyres-type indices.

however, implies that at the household level a relatively high inflation rate in one period is followed by a relatively low inflation rate in the next period. To explore this, we calculate the one-year and two-year serial correlation of inflation rates for each household and plot the average of these serial correlations across households over time in Figure 6.

We observe a slightly negative one-year within-household serial correlation of inflation rates hovering around -0.1 and a two-year serial correlation of around zero. This means that a household's inflation rate in a given period is little related to its inflation rate one year ago and virtually unrelated to the one two years ago. In other words, any persistence in a household's position in the distribution of inflation rates dissipated after two years. Households face very different inflation rates over time and are not necessarily stuck in a specific part of the distribution of household inflation rates. This result is in line with Hobijn and Lagakos (2005) who also find little persistence in household inflation rates using more aggregate data.

One explanation for the strongly varying inflation rates at the household level could also be related to the specific structure of the data. Given the irregular nature of our sample obtained by matching identical products purchased by households at different points in

Figure 5: Distribution of household-level inflation rates over time

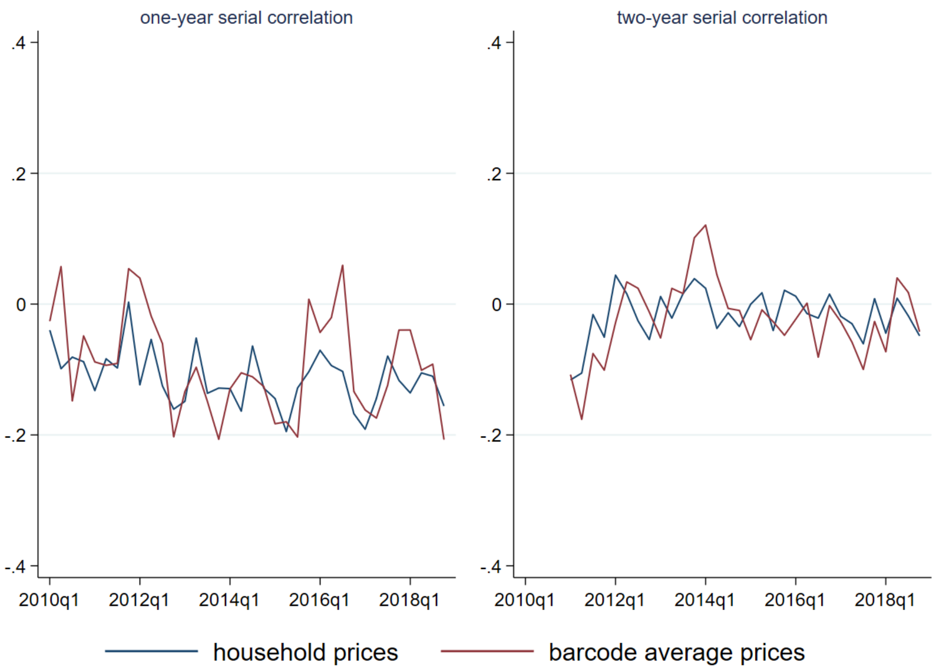


Notes: The mean and median are calculated on data between the 1st and 99th percentile of the distribution to remove the effect of extreme outliers.

time, resulting inflation rates are based on varying consumption baskets. Although this correctly captures consumption behavior at the very disaggregate level, it can render the time series of household-specific inflation rates quite unstable. In order to limit the variation from this source, we alternatively calculate household-specific inflation rates based only on households for which we can match at least five products in each period. The results (presented in the next section) are little affected by this restriction.

Overall, the evidence presented so far corroborates previous findings that household inflation rates are very heterogeneous, and even more so, the more granular the underlying consumption and price data. Moreover, this heterogeneity is – at least in times of relatively stable inflation rates as in the period 2008-2018 – a rather persistent phenomenon. However, individual households are not necessarily stuck in the distribution of inflation rates. Instead, they can alter their personal inflation rates by adjusting consumption and substituting products.

Figure 6: Within-household serial correlation of household inflation rates



Notes: Serial correlation of a household's inflation rate in quarters t and $t-4$ (left panel) and in quarters t and $t-8$ (right panel), averaged across households considering only data between the 1st and the 99th percentile of the distribution.

4 Explaining inflation heterogeneity across households

In the previous section we documented that most of the variation in household inflation rates arises from differences in prices paid, followed by differences in consumption, i.e. products chosen. This section further explores determinants of inflation heterogeneity beyond price and consumption differences.

It is a well-established fact that effective inflation rates differ across household groups according to their sociodemographic characteristics such as income, age, household size and gender (see e.g. Michael, 1979; Hagemann, 1982; Hobijn and Lagakos, 2005; Gürer and Weichenrieder, 2020; Fessler et al., 2022). The most prominent factor in this literature is income, with lower-income households often facing higher inflation rates due to higher consumption shares of more volatile components such as energy and food products. But also elderly households commonly face higher inflation rates due to their relatively higher consumption share of health products.

Making use of the sociodemographic information of surveyed households in our dataset, we calculate inflation rates for different household groups in terms of income[10], age, household size and place of residence of the household. Table 2 shows average yearly inflation rates for the entire sample period, broken down into 5 income groups, 6 age groups, 5 household size groups and 3 regional groups. For income, comparing the top (above 90,000 EUR) with the bottom (below 20,000 EUR) income group, we find higher average inflation rates for the lowest income group, but the difference is rather small. Moreover, inflation does not fall monotonically with increasing income, as the medium income group (40,000–59,999 EUR) faces a higher average inflation rate than the income group below (20,000–39,999 EUR).

For age, we find a tendency of relatively younger households facing considerably lower average inflation rates compared to elderly households. Inflation rates increase almost monotonically with the age of the household head, resulting in an inflation difference of 0.66 percentage points between the youngest (below 30) and oldest (above 70) age group. Regarding household size, on the other hand, we find a tendency of lower average inflation rates the larger the household, but differences are rather small. Particularly, single-person households stand out as the household type with higher inflation rates than all other households. Finally, households in Western Austria (Vorarlberg, Tyrol and Salzburg) face somewhat higher inflation rates than households living in the South and North-East.

Summing up the breakdown of inflation rates by household type, we find that relatively older households and those living in the West of Austria are subject to higher average inflation rates compared to others, while inflation differences across income and household size groups are relatively minor.

To account for other explanatory factors and to capture also variation over time, we proceed by examining the role of household demographics and other explanatory variables for inflation heterogeneity in a simple regression framework. We first regress the (Laspeyres) inflation rate of household i at time t relative to average inflation based on the GfK dataset in that period ($\pi^L_{i,t} - \pi^L_t$) on a number of household-specific sociodemographic observables summarized in $Z_{i,t}$:

$$\pi^L_{i,t} - \pi^L_t = \alpha + \gamma_t + \beta Z_{i,t} + \varepsilon_{i,t} \tag{11}$$

where γ_t are time (quarter) fixed-effects and $Z_{i,t}$ contains dummies for income, age and

[10] Panelists report total monthly net household income based on all income sources in 18 equally-spaced ranges starting from 0 up to 3,600 EUR and more. Income is reported once per year and, thus, may vary for individual households over time. For comparability with other studies we converted and summarized the 18 ranges of net monthly income into gross yearly income for 5 income groups shown in Table 2.

Table 2: Average yearly inflation rates by household groups (2009Q1-2018Q4)

Household group		Mean
Income (in EUR)		
	below 20,000	1.202
	20,000–39,999	1.141
	40,000–59,999	1.148
	60,000–89,999	1.147
	above 90,000	1.136
Age		
	below 30	0.896
	30–39	0.961
	40–49	1.106
	50–59	1.187
	60–69	1.173
	above 70	1.556
Household size	(persons)	
	1	1.228
	2	1.165
	3	1.124
	4	1.067
	5 and more	1.038
Regions		
	North-East	1.111
	South	1.176
	West	1.296
Overall		1.150

Notes: Income groups refer to gross yearly income of all household members from all income sources. Age refers to the age of the household head. Household size is the number of persons living in the household. Regions are defined as North-East: Upper Austria, Lower Austria, Vienna, Burgenland; South: Carinthia, East Tyrol and Styria; West: Vorarlberg, North Tyrol and Salzburg.

household size groups and the region where the household is located (North-East, South, West). The equation is estimated with OLS and robust standard errors assuming $\varepsilon_{i,t}$ is a household-specific error term.

The results of this baseline specification are reported in column 1 of Table 3. Overall, the household characteristics mentioned before are only able to explain a negligible fraction of the variance of inflation rates across households. In line with findings in earlier studies (Lieu et al., 2004; Hobijn and Lagakos, 2005), elderly households, particularly those with household heads above the age of 50, tend to face significantly higher inflation rates than younger households.[11] Furthermore, households in the touristic West of Austria face on

[11] This finding also resembles the results on the positive relationship between age and individuals' inflation expectations found in the survey literature (see e.g. Blanchflower and MacCoille, 2009; Rumler and Valderrama,

average higher inflation rates than those in the North-East. In contrast, household size and income do not explain any differences in inflation rates across households.

The latter seems to be at odds with part of the literature that suggests a negative link between income and household inflation rates. However, our results are based on FMCG, which represent only about 15% of the consumption basket and are thus not fully representative for the whole basket. Apart from that, many studies, in particular those using scanner data, also fail to find a clear and unambiguous relationship between income and inflation. In these studies, the relationship between the two variables is often found to be sample-dependent with a negative link in periods of high energy and food price inflation, but no or even a positive link in other periods (see Hobijn and Lagakos, 2005; Kaplan and Schulhofer-Wohl, 2017; Strasser et al., 2023; Fessler et al., 2022). In our sample, the lowest income group (below 20,000 EUR) experiences the highest inflation rates during periods of overall high inflation. As illustrated in Figure B1 in Appendix B, the lowest income group experienced notably higher inflation rates than all other income groups around the turning points of the inflation surges in 2011, 2014 and 2018, while in most other periods there were no systematic differences across income groups.

In column 2 we extend the baseline specification by variables that proxy the shopping behaviour of the households. Shopping behaviour, i.e. which products and varieties households buy, how they react to price increases, how often and where they buy, has been shown to determine the extent to which households are affected by inflation (Griffith et al., 2009; Jaravel and O'Connell, 2020; Nevo and Wong, 2019). Specifically, we include the number of shopping trips as a proxy for the household's intensity of search for the best price,[12] the share of private label products a household regularly buys, the share of a household's purchases in different store types, and the share of purchases of different product groups (at the COICOP-3 level).

The estimation results indicate that the shopping frequency (measured as the number of the household's visits in any supermarket over the whole quarter) of a household has a dampening impact on its inflation rate: the negative coefficient for the number of shopping trips in the current quarter and the positive sign for the number of shopping trips 4 quarters ago implies that a household that increases its shopping trips between the base period and the current period has a significantly lower inflation rate than other households. This finding is in line with Kaplan and Schulhofer-Wohl (2017) and suggests that households with a higher search intensity (proxied by shopping frequency) not only

2020 for Austria).

[12]According to consumer choice theory, the search intensity should be negatively related to the price the household pays (Diamond, 1993).

Table 3: Explaining household inflation rates with household characteristics and shopping behavior

	Baseline		Baseline + shopping behavior		Weighted regression	
Income (rel. to below 20,000)						
20,000-39,999	-0.050	(0.092)	-0.093	(0.092)	-0.066	(0.081)
40,000-59,999	0.028	(0.090)	-0.034	(0.090)	-0.041	(0.079)
60,000-89,999	0.015	(0.092)	-0.064	(0.093)	-0.045	(0.081)
above 90,000	0.033	(0.097)	-0.081	(0.098)	-0.057	(0.085)
Age (rel. to under 30)						
30-39	0.033	(0.131)	0.051	(0.131)	0.066	(0.125)
40-49	0.184	(0.127)	0.229*	(0.128)	0.209*	(0.122)
50-59	0.244*	(0.127)	0.290**	(0.129)	0.295**	(0.123)
60-69	0.227*	(0.127)	0.260**	(0.130)	0.279**	(0.124)
above 70	0.523***	(0.134)	0.501***	(0.137)	0.475***	(0.130)
Household size	-0.014	(0.021)	0.010	(0.021)	0.013	(0.018)
Region (rel. to North-East)						
South	0.060	(0.052)	0.064	(0.052)	0.030	(0.045)
West	0.157**	(0.062)	0.106*	(0.062)	0.075	(0.052)
No. of shopping trips			-0.462***	(0.058)	-0.461***	(0.050)
No. of shopping trips(-4)			0.319***	(0.060)	0.319***	(0.052)
Share of private label products			-0.319*	(0.183)	-0.458***	(0.154)
Share of purchases in (rel. to disc.)						
supermarkets			0.813***	(0.125)	0.617***	(0.107)
hypermarkets			0.544***	(0.134)	0.345***	(0.113)
markets			0.181	(0.383)	0.065	(0.305)
corner shops			1.709***	(0.644)	1.854***	(0.569)
other outlets			0.985	(0.629)	1.464***	(0.505)
Share of purchases of (rel. to food)						
non-alcoholic beverages			-0.334	(0.302)	-0.569**	(0.256)
alcoholic beverages			2.838***	(0.580)	2.386***	(0.493)
goods for household maintenance			-1.671**	(0.671)	-1.856***	(0.573)
medical products			11.617	(19.963)	-10.852	(15.070)
other recreational items			-0.989***	(0.262)	-0.829***	(0.226)
personal care			-2.092***	(0.405)	-2.294***	(0.344)
Constant	-1.111***	(0.200)	-0.790***	(0.291)	-0.239	(0.251)
Observations	91,314		91,314		91,314	
Adjusted R^2	0.005		0.009		0.012	

Notes: Dependent variable is the household inflation rate (with household-level prices) in a specific quarter relative to the aggregate inflation rate in that quarter. OLS estimation with time fixed-effects. Robust standard errors in parentheses. The reference group in the baseline regression is the lowest income group, with age below 30, living in the North-East. No. of shopping trips is the (log) number of the household's visits in any supermarket over the whole quarter (with -4 denoting the fourth lag). Share of private label denotes the (time-varying) share of private label purchases in overall purchases by each household. Share of store type is the share of each household's purchases in discounters (reference group), supermarkets, hypermarkets, markets, corner shops and other outlets. Share of product groups is the share of each household's purchases of products belonging to specified COICOP-3 groups (with food as the reference group). In the weighted regression, the (relative) number of matched barcodes of each household in each period is used as weight. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

pay lower prices, but also face lower inflation rates.[13] The mechanism responsible for this is most likely product substitution which will be investigated in the next section.

We furthermore find that households who consume more private label products tend to face relatively lower inflation rates. In a similar vein, those who primarily shop at discounters have significantly lower inflation rates than those shopping more frequently in regular supermarkets, hypermarkets or corner shops. The composition of the goods purchased also matters: Households whose consumption basket contains a larger share of food and alcoholic beverages experience higher inflation rates than those who have a higher consumption share of non-food products, such as personal care, household maintenance or recreational items. This finding is in line with average CPI-inflation rates of COICOP-3 sub-components in Austria, which are somewhat higher for food and beverages than for non-food components over our sample period.[14]

In the third column of Table 3 we present results for the extended specification where households are weighted by their relative number of matched products in each period. This approach assigns more weight to households with a larger consumption basket, aiming to address concerns that household-level inflation rates that are based on very few products may vary strongly across households and over time. Results of the weighted and the unweighted regression are generally very similar, with the only difference that in the former, households shopping more in "other outlets" also have significantly higher inflation rates than those shopping in discounters, while households with a larger consumption share of non-alcoholic beverages have significantly lower inflation rates than households shopping relatively more food products.

Another way of mitigating the problem of small and variable consumption baskets affecting our estimation results is trying to improve the stability of the consumption basket by increasing the minimum number of matched barcodes (e.g. from 5 to 10) or by including only households for which we can match at least several (e.g. 5) barcodes in each single period. These restrictions, however, come at the cost of reducing the sample size quite substantially. Finally, to check whether the results are driven by the composition of the sample, we repeat the regression without any minimum requirement on matched barcodes, including all available households.

The results for these specifications (shown in Table 4) indicate that the choice of the

[13] As alternative measures of search intensity, we also used the number of different products purchased in a quarter and the average amount spent per shopping trip by the household, but they did not turn out significant in the regressions.

[14] Food: 1.8%, non-alcoholic beverages: 2.5%, alcoholic beverages: 2.5%, goods for household maintenance: 1.2%, medical products: 1.6%, other recreational items: 1.6%, personal care: 1.4%.

Table 4: Explaining household inflation rates with household characteristics and shopping behavior – different samples

	At least 10 matched barcodes		At least 5 matched barcodes in every period		No minimum requirement on matched barcodes	
Income (rel. to below 20,000)						
20,000-39,999	-0.034	(0.096)	-0.068	(0.107)	-0.066	(0.108)
40,000-59,999	0.011	(0.094)	-0.063	(0.105)	-0.036	(0.106)
60,000-89,999	0.018	(0.096)	0.009	(0.107)	-0.059	(0.108)
above 90,000	0.003	(0.101)	-0.110	(0.112)	-0.040	(0.114)
Age (rel. to under 30)						
30-39	0.285**	(0.137)	-0.291	(0.168)	-0.076	(0.153)
40-49	0.401***	(0.134)	-0.060	(0.164)	0.088	(0.149)
50-59	0.497***	(0.135)	0.090	(0.165)	0.116	(0.150)
60-69	0.455***	(0.136)	0.029	(0.166)	0.162	(0.151)
above 70	0.711***	(0.142)	0.283	(0.172)	0.312**	(0.157)
Household size	0.017	(0.021)	0.041	(0.024)	0.005	(0.024)
Region (rel. to North-East)						
South	0.066	(0.052)	0.018	(0.058)	0.102*	(0.058)
West	0.098	(0.062)	0.052	(0.069)	0.068	(0.071)
No. of shopping trips	-0.450***	(0.060)	-0.490***	(0.068)	-0.503***	(0.068)
No. of shopping trips(-4)	0.284***	(0.063)	0.347***	(0.072)	0.267***	(0.070)
Share of private label products	-0.388**	(0.181)	-0.299	(0.196)	-0.449**	(0.214)
Share of purchases in (rel. to disc.)						
supermarkets	0.741***	(0.126)	0.700***	(0.136)	0.861***	(0.144)
hypermarkets	0.374***	(0.133)	0.497***	(0.143)	0.550***	(0.156)
markets	0.022	(0.399)	0.401	(0.347)	0.374	(0.440)
corner shops	2.717***	(0.658)	1.783***	(0.678)	1.774**	(0.748)
other outlets	1.346**	(0.595)	2.693***	(0.851)	1.157	(0.889)
Share of purchases of (rel. to food)						
non-alcoholic beverages	-2.218***	(0.587)	-3.085***	(0.691)	-2.569***	(0.792)
alcoholic beverages	-4.121***	(0.889)	-4.598***	(1.031)	-3.611***	(1.223)
goods for household maintenance	-13.122	(10.596)	-28.471	(17.647)	-1.107	(16.080)
medical products	-2.715***	(0.683)	-3.926***	(0.795)	-2.399***	(0.897)
other recreational items	-3.088***	(0.649)	-3.882***	(0.748)	-3.912***	(0.839)
personal care	-4.675***	(0.721)	-5.515***	(0.843)	-4.970***	(0.913)
Constant	1.617**	(0.652)	2.667***	(0.765)	2.181**	(0.861)
Observations	77,470		59,897		99,920	
Adjusted R^2	0.011		0.012		0.008	

Notes: Dependent variable is the household inflation rate (with household-level prices) in a specific quarter relative to the aggregate inflation rate in that quarter. OLS estimation with time fixed-effects. Robust standard errors in parentheses. For the definition of the variables, see the footnote of Table 3. The first specification includes only households for which at least 10 matched barcodes are observed in a given period, while the second requires at least 5 matched barcodes in every period, resulting in more stable consumption baskets over time. The last specification has no minimum requirement on the number of matched barcodes and thus includes all households for which at least 1 barcode could be matched in a given period. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

sample has only a limited effect on our findings. In the specification with a minimum threshold of 10 matched barcodes, the progression of inflation with the age of the household head comes out even more significant, while the geographic location appears no longer relevant for inflation differences across households. Similarly, in the specification where only households are included that consume at least 5 matched products in every period, our results regarding shopping behavior remain unchanged, but the sociodemographic factors no longer significantly explain inflation heterogeneity across households. Finally, the specification without any minimum requirement on the number of matched products delivers quite similar results to the one with at least 5 matched barcodes (baseline).[15] Therefore, we conclude that the sample composition does not drive any differences in our estimation results.

Overall, we document that the shopping behavior of households is a more relevant determinant of inflation heterogeneity than their sociodemographic characteristics. However, our observables only account for a tiny fraction of inflation heterogeneity across households in Austria. One reason for this might very well be related to the nature of our data, comprising purchases at irregular intervals and of (potentially) varying products. Consequently, the resulting inflation rates are based on varying consumption baskets (across households and over time) which can be explained by little else than the composition of each household's consumption itself.

Another potential source of unexplained heterogeneity among household inflation rates could be product substitution which has the potential to render consumption baskets at the barcode-level – and thereby also inflation rates – rather unstable. We will explore this issue in the following section.

5 Product substitution at the household level

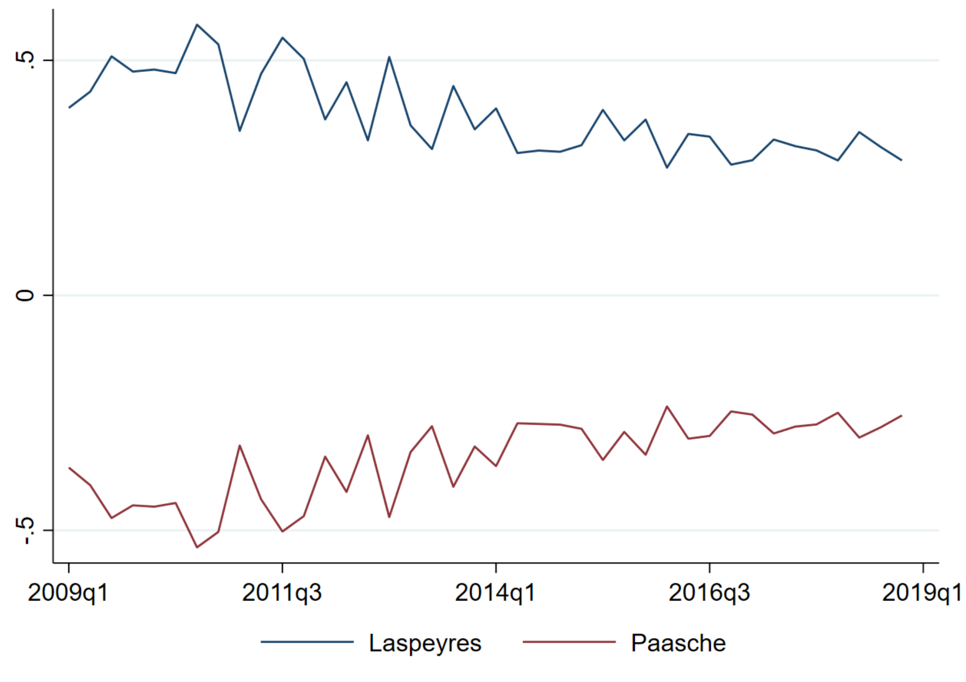
Product substitution is a well known source of measurement bias in inflation statistics, stemming from the tendency of households to substitute away from relatively more expensive products to compensate for the loss in purchasing power when relative prices change. In price indices with fixed baskets and initial-period consumption shares used as weights (as in the Laspeyres index used for the CPI) this substitution is not immediately accounted for, leading to an overestimation of inflation. The extent of overestimation

[15] We run an additional regression where the minimum number of matched barcodes per household was increased further to 20. This reduces the number of observations to about 50,000, but has virtually no effect on the results (available upon request).

depends on the frequency of updates of the weighting structure and the type of aggregation formula.

Our data allow us to measure product substitution at a very disaggregate level, i.e. at the household-barcode level. Our measure of substitution bias is, as such, likely to be more informative than those based on CPI sub-indices, as we are able to discriminate between very close substitutes.[16] The difference between household inflation rates according to the Laspeyres index, which weights products by their initial-period consumption shares and the Paasche index, which weights products by their final-period consumption shares is an upper-bound estimate of the substitution bias. An even better measure for substitution is the difference between the Laspeyres index and the geometric mean of the Laspeyres and the Paasche indices, i.e. the Fisher index which is considered an approximately correct measure of inflation in the light of substitution.

Figure 7: Mean differences of Laspeyres and Paasche inflation rates from Fisher index



Notes: Mean differences of Laspeyres- and Paasche- from Fisher-implied inflation rates calculated over all households available in the specified period.

Figure 7 plots the mean of the differences of inflation rates based on the biased Laspeyres and Paasche indices from the unbiased Fisher index over all households and over time. As

[16] According to the notation used in the literature of upper-level substitution bias (substitution between product categories) and lower-level substitution bias (substitution within product categories), our estimate maximizes the upper-level substitution bias and eliminates the lower-level substitution bias by considering the barcode as the lowest level of aggregation.

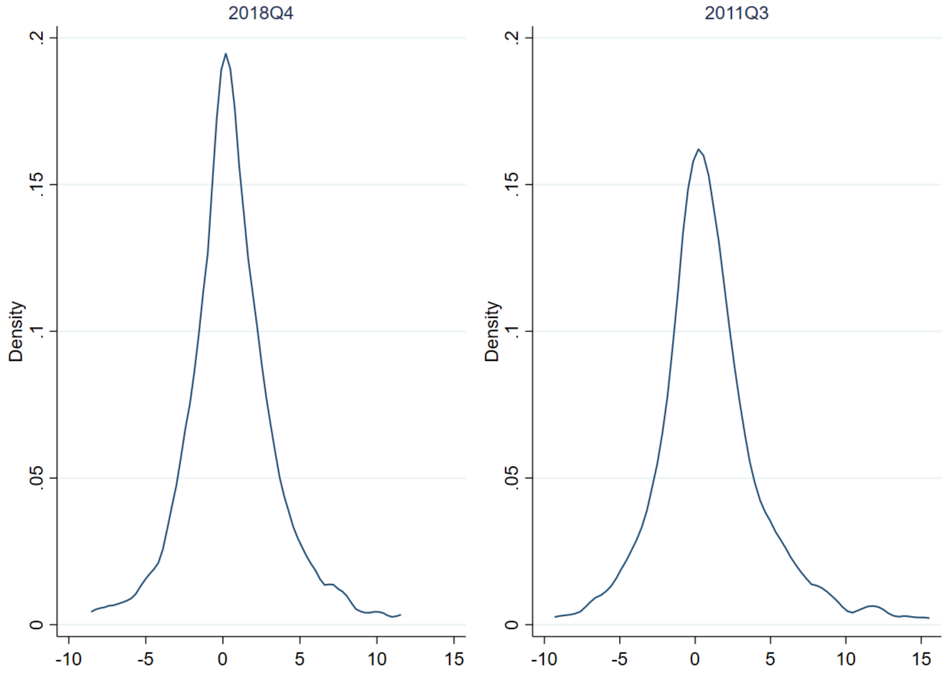
predicted by models of consumer demand, on average inflation based on the Laspeyres index is always higher than based on the Paasche index. Our estimate of the average substitution bias (difference between Laspeyres and Fisher inflation rates) is roughly 0.4 percentage points. This is broadly in line with other estimates of the substitution bias in the literature: 0.4 pp according to the Boskin Commission (Boskin et al., 1996), 0.33 pp based on FMCG (Kaplan and Schulhofer-Wohl, 2017) and 0.3 pp (Moulton, 2018), all for the US; 0.22 pp for Canada (Saborin, 2012); 0.2-0.25 pp based on FMCG for 10 euro area countries (Gábor-Tóth and Vermeulen, 2019); 0.38 pp based on FMCG for Switzerland (Braun and Lein, 2021).[17] Over time, the bias appears to decrease from almost 0.5 percentage points at the beginning of the sample period to about 0.3 percentage points at the end of the sample period. This tendency indicates that either substitution has become less important over time, due to competition or other factors, or substitution, which can go in either direction, has become more balanced.

To determine whether households always substitute in the expected direction – looking for a cheaper substitute when the relative price of the original product has increased, rather than the other way around – we plot the distribution of the difference between Laspeyres and Paasche inflation rates at the household level for the two quarters selected above, i.e. 2018Q4 and 2011Q3 (the last quarter with most households in the sample and the quarter with the highest inflation rate) in Figure 8. Both panels reveal that a sizeable part of the distribution falls on the negative side, indicating that a considerable amount of product substitution actually occurs in the "wrong" direction. In numbers, about 40% of all households have a lower inflation rate based on the Laspeyres index than on the Paasche index. Moreover, when comparing both panels, the distribution in the high inflation period (right panel) is wider. This indicates that there is more scope for substitution when inflation is relatively high, but there is still a considerable amount of substitution in the "wrong" direction (39% in the right panel vs. 42% in the left panel).

This suggests that the average positive substitution bias masks a great deal of heterogeneity at the household level. Substituting cheaper with more expensive goods can be rationally explained only if households frequently experience individual preference or income shocks (Kaplan and Schulhofer-Wohl, 2017). A more technical explanation for the considerable amount of substitution in the wrong direction could again be due to the nature of our

[17] Even though the magnitude of conventional estimates of the substitution bias is similar to our estimate, it is difficult to compare the estimates as (i) we are using a period-by-period chained index where weight changes are incorporated immediately. This would downsize the substitution bias compared to estimates from an index like the CPI where weights are updated at most once per year. On the other hand, (ii) our estimate is based on very narrowly defined barcode-level data which magnifies our estimate compared to estimates based on more broadly defined product groups.

Figure 8: Distribution of differences between Laspeyres and Paasche inflation rates



Notes: Household-level differences between inflation rates based on the Laspeyres and the Paasche index. Kernel density estimate using Epanechnikov kernel function with a bandwith of 0.4 on 200 points. Data include 2,328 households with at least 5 matched barcodes between 2017Q4 and 2018Q4 (left panel) and 2,003 households with matched barcodes between 2010Q3 and 2011Q3 (right panel). The plot is truncated at the 1st and 99th percentile of the distribution.

data, which are characterized by small and variable consumption baskets. In such an environment, the motive for substitution could also be related to product availability or idiosyncratic preferences rather than to cost considerations.

To investigate the drivers of substitution at the individual level, we run regressions equivalent to equation (11) with the individual substitution bias, i.e. the percent difference between Laspeyres and Fisher inflation rates at the household level, as the dependent variable. We regress this variable on the sociodemographic characteristics, household-specific inflation and shopping behavior variables. The results are shown in Table B1 in Appendix B. Of the sociodemographic variables (column 1), only income seems to be somewhat related with the individual substitution bias. Households with a yearly (upper-middle) income between 60,000 and 90,000 EUR substitute significantly less than low-income households.[18] Household-level inflation is found to be positively correlated with individual substitution (second column), implying that the higher a household's inflation rate the more it is inclined to substitute from higher to lower priced goods. However,

[18]See Strasser et al. (2023) for similar results for other countries.

this correlation could be partly mechanical as a relatively higher inflation rate offers the household more scope for substitution to cheaper products. Finally, the shopping behavior variables do not explain the individual substitution bias in any way (last column). Overall, these variables are again only able to explain a tiny fraction of the total variation in the individual substitution bias.

6 Conclusions

Aggregate inflation rates are composed of thousands of prices and should reflect the consumption of a "representative" household. Actual households, however, deviate from this representative household in the goods and services they consume and also in the prices they pay for these products. Using transaction-level data from a large household panel for Austria from 2008 to 2018, we document a substantial and persistent degree of heterogeneity in inflation rates at the household level. On average over our sample period, the interquartile range among household inflation rates amounts to about 6 percentage points. Even though heterogeneity at the household level is found to be substantial, the mean and median of household inflation rates are very close to aggregate inflation (for food). This supports the notion that, although an aggregate price index masks a great deal of heterogeneity, it accurately reflects the average of inflation rates of all households.

In contrast to earlier studies for Austria, the data used in this paper allow us to consider the differences in the price paid for the same product as additional source of heterogeneity. We find this to matter a lot, as most of the variation in the household inflation rates comes from differences in prices paid by consumers for the same product. If household inflation rates were calculated using average product prices instead of household specific prices, our measure of heterogeneity would decrease by about two thirds. Despite the limited product coverage of our sample (only FMCG), the heterogeneity documented in this study is substantially larger than in previous studies for Austria using price index data and expenditure shares at the product category level (see Fessler and Fritzer, 2013; Fessler et al., 2022; Koch et al., 2022). Our results are, however, qualitatively similar to the findings of Kaplan and Schulhofer-Wohl (2017), who also use scanner data for the US.

Investigating whether a specific household consistently faces above- or below-average inflation rates or whether the position of a household in the distribution of inflation rates changes over time, we clearly find that a household's inflation rate is almost uncorrelated over time. This indicates that households move rather strongly within the distribution of inflation rates from period to period.

Apart from the statistical explanation of the observed heterogeneity (with price and consumption differences), we find only limited evidence that sociodemographic characteristics and the shopping behavior of households explain inflation differences. Factors such as age of household members, the search intensity and preferences for private label products and certain store types account for a small fraction of the observed heterogeneity. In contrast, household income seems to be unrelated to individual inflation rates in our sample. In line with Argente and Lee (2021), we provide evidence that the level of inflation matters in this respect: In periods of very high food price inflation, the lowest income group indeed faces higher inflation rates than all other income groups. Accordingly, higher-income households are relatively less exposed to high levels of inflation in times of strongly rising prices, as they have more opportunities to substitute expensive for cheaper products than households with less income who already consume the cheapest products.

Households substitute quite actively and are thereby able to reduce their inflation rates on average by 0.4 percentage points. However, households do not always substitute from expensive to cheaper products, they may even substitute in the other direction. Such behaviour can only be explained rationally if households have non-homothetic preferences or if they are subject to individual preference or income shocks (Strasser et al., 2023). Comparing substitution across income groups, we find that the extent of substitution diminishes with higher income. This indicates that relatively stronger substitution of low-income groups could indeed be a reason for the muted inflation differences across income groups in our study.

Our findings have several important implications: First, inflation heterogeneity is generally a persistent phenomenon which may affect the transmission of monetary policy through its impact on real incomes. Household-specific inflation rates are, however, rather variable over time. Second, the ability of households to substitute is a central mechanism to cushion their inflation exposure. Safeguarding this self-corrective mechanism (e.g. by fostering competition and transparency, access to product varieties) is thus vital, especially in times of high inflation. Lastly, aggregate inflation indices mask underlying price dynamics and exhibit measurement (e.g. substitution) biases. Complementing aggregate price indices with information from micro price data can help to make informed policy decisions and can be especially valuable if we are interested in the distributional effects of inflation.

References

Argente, D. and Lee, M. (2021). Cost of Living Inequality During the Great Recession. *Journal of the European Economic Association*, 19(2):913–951.

Beck, G. W. and Jaravel, X. (2020). Prices and global inequality: New evidence from worldwide scanner data. *IO: Theory eJournal*.

Blanchflower, D. G. and MacCoille, C. (2009). The formation of inflation expectations: an empirical analysis for the UK. NBER Working Papers 15388, National Bureau of Economic Research, Inc.

Boskin, M. J., Dulberger, E., Gordon, R., Griliches, Z., and Jorgenson, D. (1996). Toward a More Accurate Measure of the Cost of Living. Final Report to the Senate Finance Committee.

Braun, R. and Lein, S. M. (2021). Sources of bias in inflation rates and implications for inflation dynamics. *Journal of Money, Credit and Banking*, 53(6):1553–1572.

Cravino, J., Lan, T., and Levchenko, A. A. (2020). Price stickiness along the income distribution and the effects of monetary policy. *Journal of Monetary Economics*, 110(C):19–32.

Diamond, P. A. (1993). Search, sticky prices, and inflation. *The Review of Economic Studies*, 60(1):53–68.

Fessler, P. and Fritzer, F. (2013). The Distribution of Inflation among Austrian Households. *Monetary Policy & the Economy*, (3):12–28.

Fessler, P., Fritzer, F., and Salish, M. (2022). Who pays the price when prices rise. *Monetary Policy & the Economy*, (4):1–18.

Griffith, R., Leibtag, E., Leicester, A., and Nevo, A. (2009). Consumer shopping behavior: How much do consumers save? *Journal of Economic Perspectives*, 23(2):99–120.

Gábor-Tóth, E. and Vermeulen, P. (2019). Elementary index bias: Evidence for the Euro Area from a large scanner dataset. *German Economic Review*, 20(4):e618–e656.

Gürer, E. and Weichenrieder, A. (2020). Pro-rich inflation in Europe: Implications for the measurement of inequality. *German Economic Review*, 21(1):107 – 138.

Hagemann, R. P. (1982). The variability of inflation rates across household types. *Journal of Money, Credit and Banking*, 14(4):494–510.

Handbury, J., Watanabe, T., and Weinstein, D. E. (2013). How much do official price indexes tell us about inflation? Working Paper 19504, National Bureau of Economic Research.

Hobijn, B. and Lagakos, D. (2005). Inflation inequality in the United States. *Review of Income and Wealth*, 51(4):581–606.

Ivancic, L., Diewert, W. E., and Fox, K. J. (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1):24–35.

Jaravel, X. and O'Connell, M. (2020). High-frequency changes in shopping behaviours, promotions and the measurement of inflation: Evidence from the Great Lockdown. *Fiscal Studies*, 41(3):733–755.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, 56(4):1165–1206.

Kaplan, G. and Schulhofer-Wohl, S. (2017). Inflation at the household level. *Journal of Monetary Economics*, 91(C):19–38.

Koch, S. P., Neusser, K., and Haupt, S. M. (2022). Inflationsraten nach Haushaltstypen und soziodemographischen Merkmalen. Policy Brief Nr. 14/2022, Institute for Advanced Studies.

Lieu, P.-T., Chang, C., and Chang, J.-R. (2004). Inflation rate variations across household: Empirical evidence from Taiwan. *International Journal of Business*, 9(1):103–124.

Michael, R. T. (1979). Variation across households in the rate of inflation. *Journal of Money, Credit and Banking*, 11(1):32–46.

Moulton, B. R. (2018). The Measurement of Output, Prices, and Productivity: What's Changed Since the Boskin Commission? Hutchins Center on Fiscal & Monetary Policy, The Brookings Institution.

Nevo, A. and Wong, A. (2019). The elasticity of substitution between time and market goods: Evidence from the Great Recession. *International Economic Review*, 60(1):25–51.

Rumler, F. and Valderrama, M. T. (2020). Inflation literacy and inflation expectations: Evidence from Austrian household survey data. *Economic Modelling*, 87:8–23.

Saborin, P. (2012). Measurement bias in the Canadian consumer price index: An update. *Bank of Canada Review*, Summer 2012:1–12.

Strasser, G., Messner, T., Rumler, F., and Ampudia, M. (2023). Inflation heterogeneity at the household level. Occasional Paper Series No 325, European Central Bank.

Appendix A

Further information on the sample and data

The GfK household scanner data combines information on the quantities bought at a given price for a particular product at a given retailer by an individual household. The panelist records all purchases made during a shopping trip either with a hand-held scanner or a smartphone application. Each observation corresponds to a specific transaction by a household and includes information on the expenditure, the quantity bought (in unit or volume), the date of the purchase, a barcode[19], and a GfK-specific product category which we mapped to COICOP categories. It also contains detailed product information, such as the brand name and a flag for private-label products, the sub-brand, the manufacturer, as well as the name of the store chain. Additionally, it contains information on the household characteristics of the panelist, i.e. the age of the household head, the size of the household, income, and the location. In total, the dataset includes a sample of 3,500 to about 4,500 households per year carrying out around 2 million transactions per year corresponding to 65,000–83,000 distinct barcodes and totalling 4.5–6 million EUR, depending on the year (see Table A1). The panel grows over time as all variables show a rising trend over the years.

Table A1: Breakdown of the dataset over time

	2008	2009	2010	2011	2012	2013
No. of purchases (in 1,000)	1,939	1,850	1,912	1,728	2,191	2,325
Value of purchases (in 1,000 EUR)	4,650	4,490	4,670	4,360	5,380	5,840
No. of products (in 1,000 barcodes)	65.4	65.7	66.8	66.9	75.5	75.6
No. of households	3,477	3,485	3,372	3,194	4,264	4,329

	2014	2015	2016	2017	2018
No. of purchases (in 1,000)	2,325	2,221	2,120	2,221	2,061
Value of purchases (in 1,000 EUR)	6,070	5,940	5,690	6,020	5,630
No. of products (in 1,000 barcodes)	76.7	78.4	80.1	83.3	83.3
No. of households	4,274	4,142	3,847	4,639	4,675

Notes: No. of households is the number of unique panelists included in the dataset in a given year.

Generally, purchases in such a household panel display a high sampling variation in prices and quantities. This reflects the nature of household shopping: Shopping trips are

[19]The barcode field contains either the GTIN (Global Trade Item Number) or the SKU (Stock Keeping Unit) of the product. The GTIN is a universal 13-digit number that uniquely identifies a product, while the SKU is a shop-specific article number that does not allow an identification of products across stores. The latter is often used for unpackaged items, such as fresh meat or fruit.

infrequent and unevenly spaced over time. Therefore, aggregation over time is necessary. In order to observe the same product in two different periods – to calculate a price change – at least monthly or, better, quarterly aggregation is necessary.

Despite the richness of the data, they come with three unavoidable limitations: First, the panel distinguishes only households but contains no information on household members (apart from the number). Thus, household attributes, such as age, typically refer to the non-specified head of the household and there is no information on personal attributes such as gender or the age of other household members. Second, all transactions are self-reported and might therefore not include all purchases of all household members. There is no cross-validation with retailer databases such that incorrect entries in the form of un- or misclassified products, inconsistently reported quantities, and typos are unavoidable. Third, there is usually no directly comparable pricing history at the store level which would be necessary to identify sale offers or the use of coupons and discount cards.

Table A2 shows the composition of our sample in terms of COICOP groups compared to the Austrian CPI. Roughly 80% of the sample is composed of food and (alcoholic and non-alcoholic) beverages and 20% of non-food items. These proportions are comparable to the Austrian CPI. Considering only matched barcodes, which is necessary to calculate inflation rates and has the effect of reducing the sample from around 23 to 4.4. million observations, the composition shifts somewhat towards food and beverage products. Within the food category, processed food items such as milk products, oils and fats or drinks, are over-represented in our data compared to unprocessed food items such as fresh meat or fruit. This has probably to do with the fact that the latter often come with SKU numbers which do not always allow identification of products across shops and thus drop out from our sample.

In addition to the reduction of the sample by matching barcodes over time, our sample is further reduced by the requirement of a minimum number of matched barcodes (5 in our baseline calculation) for a household's inflation rate to be included in the displayed statistics. Furthermore, we remove price changes that are larger/smaller than a factor of 3 as outliers, i.e. if $p_{ij,t}/p_{ij,t-4} > 3$ or if $p_{ij,t}/p_{ij,t-4} < 0.33$.

Table A2: Distribution of spending across COICOP-4 groups in Austrian CPI and GfK data

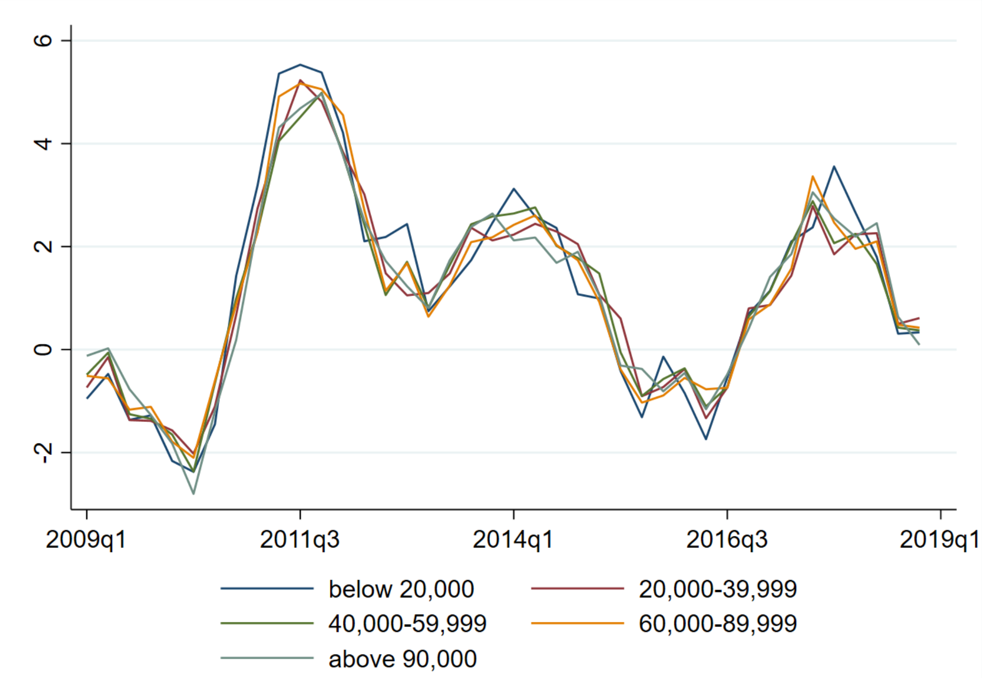
COICOP groups	CPI absolute	CPI relative	GfK all data	matched barcodes
Food and Non-Alcoholic Beverages				
Food				
Bread and cereals	2.4	15.4	7.4	6.6
Meat	2.4	15.5	1.4	1.1
Fish and seafood	0.4	2.6	3.3	2.4
Milk, cheese and eggs	1.6	10.6	18.5	24.2
Oils and fats	0.4	2.4	3.9	4.2
Fruit	0.1	1.0	2.1	1.8
Vegetables	0.5	3.1	4.1	3.8
Sugar, jam, honey, chocolate and confectionery	1.0	6.7	9.8	8.5
Food products n.e.c.	0.5	3.4	8.7	7.0
Non-alcoholic beverages				
Coffee, tea and cocoa	0.6	3.6	5.0	6.5
Mineral waters, soft drinks, fruit and vegetable juices	0.9	5.8	8.0	9.9
Alcoholic Beverages, Tobacco and Narcotics				
Alcoholic beverages				
Spirits	0.2	1.4	1.9	1.8
Wine	0.7	4.5	1.0	0.8
Beer	0.6	3.6	3.8	6.1
Furnishings, Household Equipment and Routine Household Maint.				
Goods and services for routine household maintenance				
Non-durable household goods	0.9	6.0	4.6	2.1
Health				
Medical products, appliances and equipment				
Other medical products	0.1	0.4	0.04	0.01
Recreation and Culture				
Other recreational items and equipment, gardens and pets				
Pets and related products	0.6	3.8	5.3	6.6
Miscellaneous Goods and Services				
Personal care				
Electric appliances for personal care	0.1	0.4	0.3	0.04
Other appliances, articles and products for personal care	1.5	9.9	11.0	6.7
Sum	15.4	100	100	100

Notes: Numbers are averages over 2008–2018 in %. Subcategories are not exhaustive and do not necessarily add up to higher-level categories.

Appendix B

Further results

Figure B1: Household inflation rates by income group over time



Notes: Income groups refer to gross yearly income of all household members from all income sources.

Table B1: Explaining the substitution bias with household characteristics and shopping behavior

	Baseline		Baseline + household inflation		+ shopping behavior	
Income (rel. to below 20,000)						
20,000-39,999	0.259	(0.299)	0.261	(0.299)	0.293	(0.301)
40,000-59,999	-0.398	(0.469)	-0.399	(0.470)	-0.322	(0.436)
60,000-89,999	-0.662*	(0.359)	-0.662*	(0.359)	-0.558	(0.344)
above 90,000	-0.253	(0.209)	-0.253	(0.209)	-0.110	(0.190)
Age (rel. to under 30)						
30-39	-0.890	(0.770)	-0.891	(0.770)	-0.889	(0.777)
40-49	-0.040	(0.312)	-0.045	(0.313)	-0.039	(0.311)
50-59	-0.361	(0.253)	-0.368	(0.253)	-0.350	(0.265)
60-69	-0.351	(0.304)	-0.358	(0.303)	-0.300	(0.305)
above 70	-0.491	(0.734)	-0.507	(0.736)	-0.374	(0.702)
Household size	0.106	(0.105)	0.107	(0.105)	0.075	(0.112)
Region (rel. to North-East)						
South	0.356	(0.256)	0.354	(0.256)	0.336	(0.257)
West	0.040	(0.250)	0.035	(0.248)	0.068	(0.262)
Household inflation rate			0.030*	(0.016)	0.032*	(0.017)
No. of shopping trips					0.240	(0.262)
No. of shopping trips(-4)					0.057	(0.234)
Share of private label products					0.421	(0.750)
Share of purchases in (rel. to disc.)						
supermarkets					-0.342	(0.569)
hypermarkets					-0.423	(0.687)
markets					0.169	(1.052)
corner shops					-1.180	(3.346)
other outlets					-0.504	(1.604)
Share of purchases of (rel. to food)						
non-alcoholic beverages					1.443	(1.917)
alcoholic beverages					-1.239	(1.699)
goods for household maintenance					-14.382*	(7.591)
medical products					-1.219	(5.890)
other recreational items					0.616	(1.227)
personal care					2.456	(2.385)
Constant	0.147	(0.335)	0.181	(0.338)	-0.783	(1.273)
Observations	91,118		91,118		91,118	
Adjusted R^2	0.001		0.001		0.001	

Notes: Dependent variable is the percent difference between household-specific Laspeyres- and Fisher-implied inflation rates in a given quarter. OLS estimation with time fixed-effects. Robust standard errors in parentheses. For the definition of the variables, see the footnote of Table 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Acknowledgements

We would like to thank the members of the PRISMA (Price-Setting Microdata Analysis) Network of the ESCB and the participants of an OeNB seminar for helpful comments and suggestions, and Anaëlle Touré as well as David Wittekopf for excellent research assistance. The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank, the Oesterreichische Nationalbank, or the Eurosystem.

Teresa Messner

Oesterreichische Nationalbank, Vienna, Austria; email: teresa.messner@oenb.at

Fabio Rumler

Oesterreichische Nationalbank, Vienna, Austria; email: fabio.rumler@oenb.at

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

This paper can be downloaded without charge from www.ecb.europa.eu, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-6374-9 ISSN 1725-2806 doi:10.2866/167630 QB-AR-24-011-EN-N