The granular origins of inflation*

Santiago Alvarez-Blaser
University of Basel

Raphael Auer
Bank for International Settlements
and CEPR

Sarah M. Lein
University of Basel
CEPR and KOF ETHZ

Andrei A. Levchenko
University of Michigan
NBER and CEPR

March, 2024

Latest version

Abstract

This paper uses barcode-level price data for 14 advanced and emerging market countries over the period 2008-2020 to investigate the role of individual firms and product categories in aggregate inflation. We develop a decomposition of inflation into the components due to aggregate shocks, and the granular residuals capturing the impact of individual firms and product categories, respectively. In advanced economies, the firm granular residual accounts for nearly 40% of the variance of overall inflation, while the category granular residual accounts for another 20%. Most of the variation in the firm granular residual is due to idiosyncratic shocks, rather than to higher sensitivity of larger firms to the aggregate shocks. In emerging markets, the granular components together account for 13% of the total inflation variance, indicating that granular forces are weaker in higher-inflation environments where aggregate shocks matter more.

Keywords: Inflation, aggregate fluctuations, price setting, firm-level shocks, granular economies, large firms, globalization.

JEL Codes: E31, E32, F44

*The views presented in this paper are those of the authors and not necessarily those of the Bank for International Settlements. We thank AiMark and AiMark's data partners GFK, Kantar Worldpanel and IRI for providing us with the data for this paper and in particular Alfred Dijs and Elizabeth Ramaker from Aimark for their support throughout this project. We thank Saroj Bhattarai, Zeno Enders, Xavier Gabaix, Erwan Gautier, Gernot Mueller, Emi Nakamura, Raphael Schoenle, Jon Steinsson, Georg Strasser, and seminar participants at the ASSA, BIS, Banque de France, BdF-TSE Sectoral Inflation Dynamics Workshop, CEBRA, CEPR IMF Annual Meeting, ECB, EMMC Conference, Harvard University Macro Workshop, Inflation Drivers and Dynamics Conference, JME-SNB-SCG Conference, PSE-CEPR Policy Forum, VfS, VfS standing field committee on Monetary Economics, UC Berkeley, U Heidelberg, and U Tuebingen. Email: santiago.alvarez@unibas.ch, Raphael.Auer@bis.org, sarah.lein@unibas.ch, and alev@umich.edu.

1. Introduction

Textbook monetary economics views inflation as fundamentally driven by aggregate shocks, such as money supply or policy rates (Woodford, 2003; Galí, 2015).[1] While the modern literature models rich micro-level price adjustment heterogeneities, idiosyncratic firm behavior is typically integrated out, leaving no role for individual firms in aggregate inflation. At the same time, following Gabaix (2011)'s seminal contribution, an influential strand of the macro literature has modeled theoretically and documented empirically that shocks to individual (large) firms can generate fluctuations in real GDP, a phenomenon that goes by the name of "granularity."

However, not much is currently known about the role of such granularity for aggregate inflation. This paper uses detailed barcode-level price data for 14 advanced and emerging market countries over the period 2008-2020 covering 2.5 billion transactions to provide a forensic account of the contributions of individual firms and product categories to overall inflation. For each barcode-level price, we can identify the firm that produced the item, its product category, and sometimes the retail chain through which it is sold. This information enables us to decompose inflation into several components.

By definition, aggregate inflation is an expenditure-share-weighted change in individual prices. We posit that each micro price can be written as a sum of the aggregate (country-time) component, a firm-specific component, and a product-specific component. Aggregating up the barcode-level prices produces an additive decomposition of the inflation rate into (i) the aggregate component, (ii) the firm granular residual, and (iii) the product category granular residual. The firm (resp. product category) granular residual captures the contribution of firm (resp. product category) idiosyncratic components to the overall inflation.

We generalize the conventional granular residual decomposition (Gabaix, 2011; di Giovanni et al., 2014; Gabaix and Koijen, 2024) in two ways. First, we allow for multiple non-nested dimensions of granularity (in the baseline, firms and categories). Second, it has been understood since Gabaix (2011) that a granular residual can arise either from idiosyncratic shocks to large firms, or from differential responses of large firms to the same aggregate shocks. Our notion of granular residual explicitly allows for both of these driving forces. We also document which one is more powerful in our context.

Our results can be summarized as follows. The firm and category granular components are responsible for nearly 60% of the variance of inflation in advanced economies over the 2008-2020 period. The firm component is relatively more important, explaining some 38% of the inflation variance, whereas the category component accounts for 19%. The period we study is an era of low inflation in the advanced economies. To understand how results might differ in higher-inflation

[1] This view is most famously encapsulated by Milton Friedman's emblematic quote that "inflation is always and everywhere a monetary phenomenon" (Friedman, 1963).

environments, it might be informative to look at emerging markets. These economies had a higher overall inflation rate in our sample. Correspondingly, the granular components combined account for only 13% of the overall inflation variance in those countries. This comparison suggests that the role of granularity varies importantly with the overall inflation rate.

We next decompose the granular residuals into the components due to the differential responsiveness to aggregate shocks, and the true idiosyncratic shocks. The firm granular residual is almost entirely driven by the true idiosyncratic shocks. By contrast, more than half of the variability in the category granular residual is due to the categories' differential responsiveness to aggregate shocks.

We also investigate the role of a third dimension - the retailer. This dimension can also have an important granular component, as the retail sector is often dominated by a small number of large chains. Unfortunately, working with the retailer dimension constrains us to a significantly smaller sample as the identity of the retailer is not always recorded in our data and not all products are sold in multiple retailers. With that caveat, the retailer dimension appears less important. The retailer granular residual only explains 14% of the aggregate inflation variance in the advanced economies.

This paper draws from, and contributes to two strands of the literature. The first one studies the micro origins of aggregate fluctuations (Long and Plosser, 1983; Jovanovic, 1987; Acemoglu et al., 2012; Carvalho and Gabaix, 2013). Gabaix (2011) argued that when the firm size distribution is fat-tailed, firm-specific idiosyncratic shocks do not average out and thus produce fluctuations in aggregate output, introducing the concept of granular fluctuations. Subsequent contributions have theoretically modeled granular fluctuations (e.g. Carvalho, 2014), shown empirically that firm idiosyncratic shocks are important for aggregate fluctuations (e.g. di Giovanni et al., 2014), and studied this phenomenon in the context of international trade (di Giovanni and Levchenko, 2012; di Giovanni et al., 2018; Gaubert and Itskhoki, 2021), government policy (Gaubert et al., 2021), government spending (Cox et al., 2020), and banking (Amiti and Weinstein, 2018; Bremus et al., 2018), among others. The literature has for the most part neglected the implications of granularity for prices. Our paper generalizes the standard granular residual decomposition to allow for multiple dimensions of granularity, and uses micro-price data to document granularity in inflation.

Second, our analysis relates to the recent work on price-setting in multiproduct firms. This literature argues that strategic complementarities for pricing decisions can amplify real effects of nominal shocks (Carvalho, 2006; Alvarez and Lippi, 2014; Pastén et al., 2023), and also provides empirical evidence that price adjustments are strongly synchronized within firms (e.g. Midrigan, 2011; Bhattarai and Schoenle, 2014) or retailers (e.g. DellaVigna and Gentzkow, 2019; Bonomo et al., 2020; García-Lembergman, 2022). These contributions provide possible theoretical and empirical underpinnings for our findings.

The rest of the paper is organized as follows. Section 2 presents the data that will be used together with some summary statistics. Section 3 describes the methodology and the empirical results. Section 4 concludes. Details of the data construction and additional empirical results are collected in the appendix.

2. Data and summary statistics

2.1 Data assembly

Data source. The analysis employs a homescan dataset of retail prices and expenditures from AiMark for 14 countries (Argentina, Austria, Belgium, Brazil, China, Germany, Spain, France, Mexico, the Netherlands, Russia, Sweden, United Kingdom and United States). We observe most of these countries for the years 2008-2020, with Germany observed for the longest period (2005-2020), while data for Argentina, Brazil, China, Mexico, and Russia start only in 2011.

In each country, a participating representative sample of households logs its supermarket and drugstore purchases. Our raw data contain 2.5 billion transactions. Each entry in the dataset records a purchase of a product by a household. The entry records the household identifier, product barcode (a unique item identifier), price paid, date of purchase, and retailer name. For each barcode, data include information on the associated brand and firm (producer), and barcodes are further classified into product categories and subcategories. The data also record a set of socioeconomic characteristics of the households purchasing the items, including geographic location of the households' residence. To fix notation, product (barcode) i belongs to product category g, is sold in country c by firm f and in retail shop s.

Data preparation. For the main analysis, we compute the modal price (following Eichenbaum et al., 2014; Auer et al., 2021) and the total expenditures within country, quarter, and barcode combinations. We then take the year-on-year log difference in price as the measure of the inflation of a given barcode and country. Below we refer to each of these year-on-year barcode-level inflation rates as one observation.

We standardize and in some instances concord categories, firms, brands, and products across countries. First, we ensure cross-country comparability of categories, such as "Fruit Juice" or "Breakfast Cereals." For this, we establish a standardized set of 110 categories as follows. We start with the English category variable that is included in the raw data. This variable – "category name English" – is included in all datasets and is also consistent across countries, but it covers only 35% of the unique barcodes in our dataset. To complete the coverage of the standardized categories, we rely on the more comprehensive "category" variable in the native language as well as the finer "subcategory" variable that exists for most countries (also in the native language). We use manual matching of the "category" and "subcategory" information to our 110 standardized categories and in addition we utilize product barcodes that are present in multiple countries. For example, if a given barcode is categorized as a category-subcategory combination "Fruit Juice-Apple Juice" in 90% or more transactions in all other countries, that product is assigned "Fruit Juice-Apple Juice" also in countries in which the category-subcategory information is initially missing.

The names of firms and brands also differ across countries. We adopt a five-step procedure to harmonize firm names across countries, as described in Appendix A. This appendix also provides details on the outcomes of this matching process. To show robustness, Appendix Table A4 reproduces the empirical analysis with a simpler matching procedure.

Retailer sample. We also create a second data set that allows us to investigate the importance of large retailers in inflation. Given the small number of retailers, only two adjustments were needed for this dimension. First, if one retailer has a subsidiary chain, such as "Carrefour Express" we assigned this subsidiary to the parent chain, "Carrefour."

Second, for some purchases the retailer is not identified, with the retailer field coded as "other." Relatedly, for some countries in the data some small retailers were lumped together depending on the type of store, for example "Bakery" or "Pet store." We replaced the retailer entry with "other" in these cases. Appendix Figure A1 plots the cumulative share of aggregate expenditure on firms and retailers that could not be identified in the data. For most countries, the vast majority of total expenditure can be attributed to retailers and firms. However, the unidentified retailer share is non-negligible in some important emerging markets.

To deal with the unnamed "other" retailers, we adopt two approaches. The first is to drop these observations. As an alternative approach, we treat the "other" retailer as a separate regional retailer using the region or postal code information of the household. Thus, purchases made from "other" retailers in different cities in the same country are assumed to come from distinct retailers. Appendix Table A5 reports the results obtained following this alternative approach.

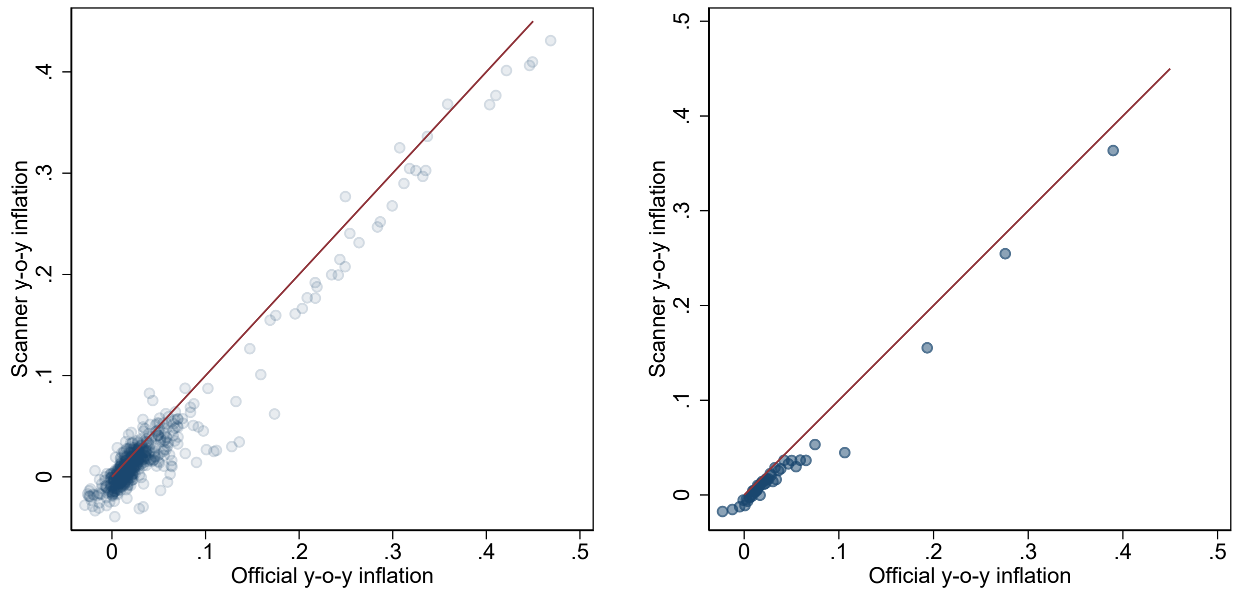
When investigating the retailer dimension, the sample of products is smaller as we need to observe products in two consecutive years in the same retailer-country-quarter in order to compute the underlying product inflation rates.

2.2 Basic patterns

Inflation measures from official sources compared to scanner data measures. We start by showing that inflation rates in our data closely correspond to official inflation rates for the same set of product categories.[2] Figure 1 plots the inflation computed from our scanner data against inflation for the same

[2] The finding that inflation rates from household scanner data closely co-move with official CPI inflation rates has been documented for other countries and time periods by, for example, Kaplan and Schulhofer-Wohl (2017), Redding and

Figure 1: Official vs. scanner data aggregated inflation



(a) Scatter

(b) Binscatter

Notes: This figure plots the inflation computed using the scanner data on the y-axis against the inflation for the same set of categories from official sources on the x-axis. Left panel shows a scatter including all countries and the right panel shows a binscatter of the same observations. Both panels include a 45-degree line as reference.

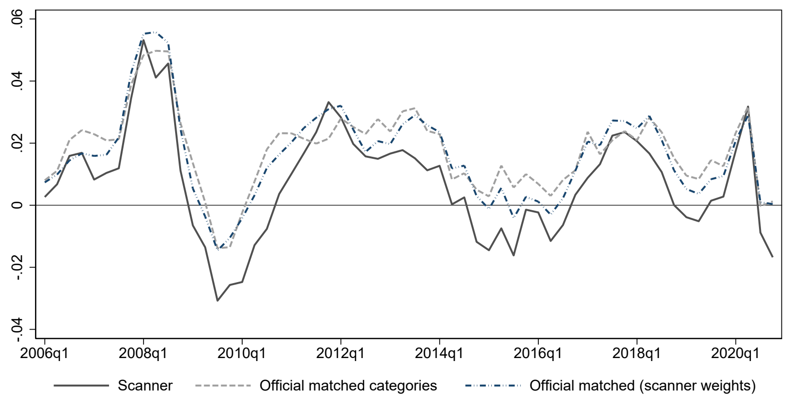
set of categories from official sources, for all countries and time periods, along with a 45-degree line. The overall correlation when pooling all countries is 0.97, while the average correlation of scanner and official inflation across countries is 0.83 (Mexico is an outlier at 0.46). Figure 2 shows inflation computed from our scanner data alongside the official indices for underlying matched CPI categories for Germany, the US, and Argentina over time. We calculate the price indices from the official data by utilizing only CPI categories that align with the categories available in the scanner data. Since some categories might be over- or under-represented in our scanner data compared to official CPI weights, we compute an official index using both scanner data weights and official weights.[3] The disparities between them are minimal. The country-specific correlations over time and figures for the rest of the countries can be found in Appendix Figure A2.

Summary statistics. Column 1 of Table 1 reports the numbers of the raw entries in the data, by country. Column 2 displays the number of observations for the product-level inflation rates in the sample that incorporates the retailer dimension, thus the number of inflation observations by country, retailer, quarter, and barcode. Each observation is indexed by $ifgsct$, thus Δp_{ifgsct} is the inflation of

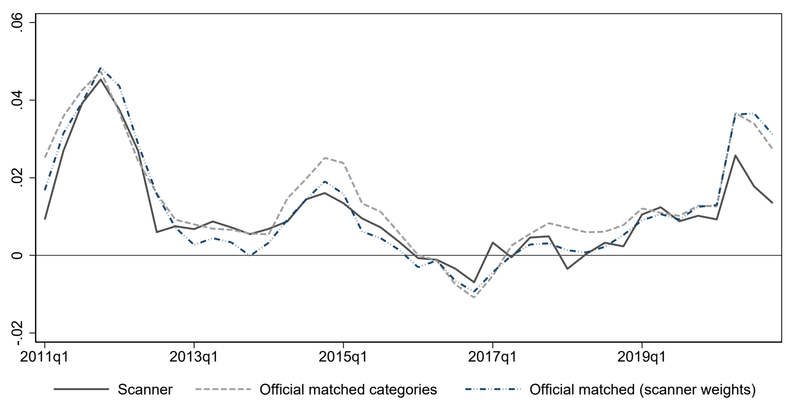
Weinstein (2019),Braun and Lein (2021), Beck and Jaravel (2021), or Beck et al. (2023).

[3] In the case of Argentina, only the scanner data weights are used because we could not find official category weights at the disaggregated level. The quality of the matched categories depends on the available disaggregated data. For China no official index was constructed given that no disaggregated CPI indices were available for the period covered in this paper.

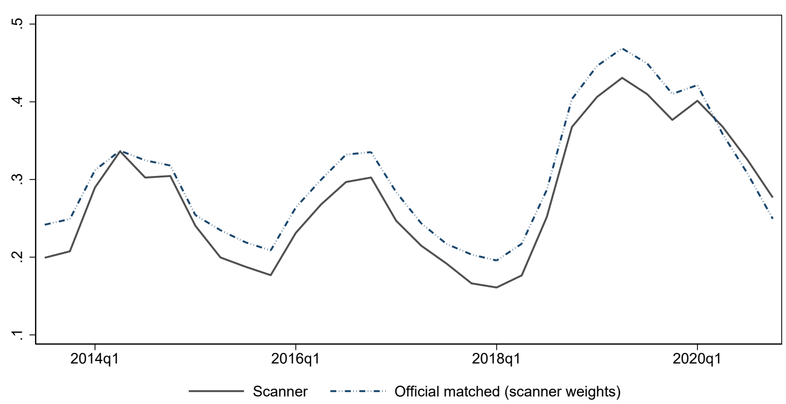
Figure 2: Official vs scanner data aggregate inflation



(a) Germany



(b) US



(c) Argentina

Notes: All figures show the year-on-year inflation rates. "Official matched categories" use official inflation rates and weights while "Official matched (scanner weights)" weights the official inflation rate of each category with the weight observed in the scanner data for the same category. Only three out of 14 countries shown. Sources of official indices are Eurostat, Bureau of Labor Statistics (BLS), and the Dirección General de Estadística y Censos. The rest of the countries can be found in Appendix Figure A2.

Table 1: Descriptive statistics

	(1) Transactions (in M)	(2) N of Δp_{ifgsct}	(3) N of Δp_{ifgct}	(4) N_s	(5) N_f	(6) N_i	(7) Mean N_i^f	(8) Median N_i^f	(9) Int. exp.	(10) Years
AR	21.65	668,779	625,734	23	3,925	65,663	17	3	0.72	2011-2020
AT	27.32	2,178,897	1,150,140	133	4,258	130,897	24	3	0.96	2008-2020
BE	55.30	2,810,545	2,070,112	136	11,287	246,166	13	2	0.97	2008-2020
BR	68.22	2,174,693	1,133,168	420	14,225	116,837	6	2	0.60	2011-2020
CN	93.76	4,127,005	3,705,962	413	72,327	545,900	6	2	0.38	2011-2020
DE	405.92	10,735,553	5,931,749	21	10,194	490,473	10	3	0.97	2005-2020
ES	107.63	5,198,536	2,946,189	194	13,361	279,181	11	3	0.85	2007-2020
FR	181.20	9,399,914	4,446,998	302	6,006	363,437	17	2	0.95	2008-2020
MX	95.25	2,048,790	751,152	185	4,127	67,987	9	2	0.74	2011-2020
NL	167.60	5,600,508	2,723,866	133	9,855	315,661	8	2	0.96	2008-2020
RU	70.93	2,823,035	2,014,630	437	12,949	267,679	15	4	0.62	2011-2020
SE	20.56	1,859,183	777,350	104	3,064	70,650	10	2	0.85	2006-2020
UK	610.29	7,581,905	4,393,272	60	6,297	362,562	23	3	0.86	2005-2020
US	643.13	36,153,192	12,609,763	630	35,624	1,181,127	22	3	0.69	2010-2020
Total	2,568.74	93,360,535	45,280,085	3,010	174,553	4,183,232	24	2	-	2005-2020

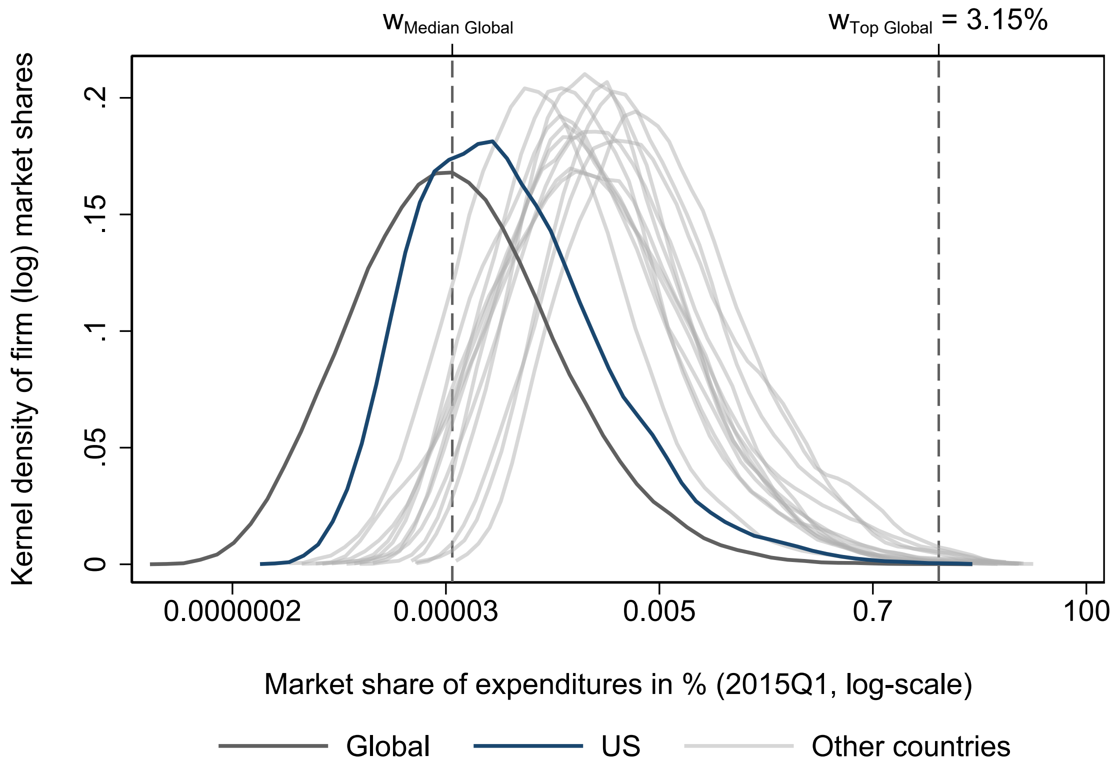
Notes: "Transactions" refers to the number of entries in the raw data. N of Δp_{ifgsct} and N of Δp_{ifgct} indicate the number of available year-on-year inflation rates using the product-retailer-quarter and product-quarter aggregation, respectively. N_s, N_f, N_i are the number of unique retailers, firms and products that appear in the data. Mean N_i^f and median N_i^f indicate the average and median number of products produced by a firm. "Int. exp." indicates the share of expenditures in each country in products of firms which appear in at least one other country of the data.

barcode i, which belongs to category g and is produced by firm f, in country c and sold by retailer s. The third column displays the number of observations for the product-level inflation rates in the sample that does not incorporate the retailer dimension, thus the number of inflation observations by country, quarter, and barcode. Our baseline decomposition is based on these approximately 45 million observations.

Columns 3 and 4 report the numbers of retailers and firms in each country. Overall, there are about 3,000 distinct retailers, and 175,000 distinct firms. Column 6 shows that we observe around 4.2 million unique products. Table 1 also reports some statistics on the mean and median number of products a firm produces (columns 7 and 8, respectively). In total The average (median) number of unique products one firm sells is 24 (2). In column 9, we report the share of expenditures in firms that operate in at least two countries within the dataset. In most countries, a significant portion of expenditures is allocated to international firms. Column 9 reports the years covered for each country.

Figure 3 shows the kernel densities of firm log expenditure weights when pooling all countries in our sample in the first quarter of 2015. Expenditure shares are strongly right-skewed across firms and the distribution of market shares shows "fat tails," an important indication that underlying granularities might affect aggregate inflation. The fat tails are visible even on log scale and do not disappear when averaging over countries (i.e. giving each country the same importance). The largest firm concentrates over 3% of global expenditures while the median firm has an expenditure share below 0.0001%.

Figure 3: Kernel densities for $\log(w_{fct})$ and $\log(w_{ft})$



Notes: Kernel densities computed from firm log market shares in the first quarter of 2015. The black line represents the distribution of world expenditure shares, the blue line represents the US, and each light gray line represents the kernel density of one of the other 13 countries. Global average firm weights computed giving each country the same weight and aggregating by firm across countries. Vertical lines indicate the position in the distribution of the weight of the median global firm and of the largest firm that appears in all countries in our sample in 2015Q1.

Table 2 reports the combined expenditure share of the top 10 and top 1% largest firms, categories and retailers. The market share of the top 10 firms reported in the first column is on average around 40%, with the highest in Mexico (59%) and lowest in Russia (17%). When looking at the weight of the top 10, concentration seems similar across the firm (column 1), category (column 3) and retailer (column 5) dimensions. However, the concentration when looking at the top percentile is much higher at the firm dimension, compared to the category or retailer dimensions (column 2, compared to columns 4 and 6). The underlying reason is that the number of firms in the sample is significantly larger than the number of retailers or categories. As a result, even one percent of the firms constitutes more than 10 firms. This reinforces the argument for the presence of fat tails, especially at the firm level. Despite the extensive number of firms represented, expenditures remain concentrated within a small proportion of them.

Table 2: Expenditure shares of top firms, retailers and categories

	(1)	(2)	(3)	(4)	(5)	(6)
	Firms		Categories		Retailers	
Weight of:	Top 10 f	Top percentile f	Top 10 g	Top percentile g	Top 10 s	Top percentile s
AR	0.31	0.66	0.37	0.06	0.35	0.08
AT	0.48	0.84	0.43	0.13	0.84	0.32
BE	0.53	0.86	0.38	0.11	0.85	0.41
BR	0.24	0.64	0.50	0.10	0.14	0.09
CN	0.19	0.76	0.52	0.07	0.19	0.13
DE	0.43	0.77	0.39	0.13	0.80	0.16
ES	0.46	0.78	0.38	0.11	0.75	0.53
FR	0.42	0.83	0.35	0.12	0.82	0.61
MX	0.59	0.89	0.55	0.11	0.33	0.18
NL	0.51	0.82	0.40	0.10	0.76	0.36
RU	0.17	0.47	0.46	0.15	0.44	0.35
SE	0.42	0.76	0.49	0.08	0.91	0.36
UK	0.53	0.83	0.44	0.09	0.90	0.29
US	0.31	0.89	0.38	0.12	0.51	0.46

Notes: Top 10 weight based on total expenditure share of the largest 10 firms, categories or retailers in each country across periods. Top percentile indicates the weight of the top one percentile of firms, categories or retailers sorted by expenditure share. Expenditure shares based on all expenditures, also including expenditures in not identified firms and retailers.

3. Granularity and the evolution of inflation

This section presents our main empirical results. We start with the standard granular residual and then develop our main decomposition that features multiple dimensions of granularity.

3.1 Warmup: simple granular residual

Denote by Δp_{ifgct} the year-on-year growth rate of the price of the barcode i belonging to product category g, produced by firm f, observed in country c and quarter t. To a first order, the growth rate in the aggregate price index in country c is:

$$\Delta p_{ct} = \sum_i w_{ifgct-4} \Delta p_{ifgct}, \tag{3.1}$$

where $w_{ifgct-4}$ is the share of barcode i in total expenditure in country c in the same quarter of the previous year.

Inflation can be decomposed as follows (Gabaix, 2011; Gabaix and Koijen, 2024):

$$\Delta p_{ct} = \underbrace{\frac{1}{N_{i\in c,t}}\sum_i \Delta p_{ifgct}}_{U_{ct}} + \underbrace{\sum_i w_{ifgct-4}\left(\Delta p_{ifgct} - \frac{1}{N_{i\in c,t}}\sum_i \Delta p_{ifgct}\right)}_{\Gamma_{ct}}, \tag{3.2}$$

where $N_{i\in c,t}$ is the number goods in country c and period t. The first term, U_{ct}, is the simple average price change across all barcodes in the economy. The second term, Γ_{ct}, is the granular residual. The granular residual is the expenditure-share weighted deviation of the price change in barcode i from the simple average price change across all barcodes in the economy. A non-zero granular residual will arise if barcodes with larger expenditure shares have systematically higher or lower relative price changes. Indeed, it can be rewritten as a covariance between price changes and expenditure shares (di Giovanni et al., 2024). By contrast, Γ_{ct} would equal 0 if either all products had the same expenditure weight or price changes were the same for all barcodes.

Though equation (3.2) can be implemented regardless of the data generating process for the prices and expenditure weights, to build intuition for this decomposition it is helpful to posit that each barcode-level price change is the sum of an aggregate shock and an idiosyncratic shock with mean zero:

$$\Delta p_{ifgct} = \delta_{ct} + \delta_{ifgct},$$

where $\frac{1}{N_{i\in c,t}}\sum_i \delta_{ifgct} = 0$. Then, it is immediate that in this economy, U_{ct} would be capturing the aggregate shock while Γ_{ct} would capture the weighted sum of firm idiosyncratic shocks:

$$\begin{aligned}\Delta p_{ct} &= \frac{1}{N_{i\in c,t}}\sum_i \left(\delta_{ct} + \delta_{ifgct}\right) + \sum_i \left(w_{ifgct-4} - \frac{1}{N_{i\in c,t}}\right)\left(\delta_{ct} + \delta_{ifgct}\right) \\ &= \delta_{ct} + \sum_i w_{ifgct-4}\delta_{ifgct}.\end{aligned}$$

Thus, U_{ct} reflects the relative importance of aggregate shocks, while Γ_{ct} is the contribution of idiosyncratic shocks to the aggregate inflation.

Figure 4 shows the time path of aggregate retail inflation and the simple granular component Γ_{ct} over the available periods for Germany, the US, and Argentina. We focus on the dynamics of these three countries in the main text given their size and heterogeneous inflation experience. Appendix D displays the inflation and the granular components for all other economies included in our sample. The granular residual is significant in magnitude for Germany and the US. In contrast, in Argentina Γ_{ct} has about the same absolute magnitude as it does in the US and Germany, but is a much less significant component of the overall inflation, which appears driven by aggregate shocks in that country.

The simple granular residual exercise reveals the presence of granularities in the inflation data but is not informative on the underlying sources. In particular, each barcode i has multiple overlapping characteristics. For example, it belongs to a firm that produced it, and it belongs to a broader category. (Below, we will also add the retailer dimension.) Thus, there are multiple distinct reasons Γ_{ct} can arise: multi-product firms adjust prices of different products simultaneously; and price changes are synchronized within categories, due to either common supply shocks or complementarities in pricing. These forces could coexist, and thus must be analyzed jointly. This is what we turn to next.

3.2 Granular layers methodology

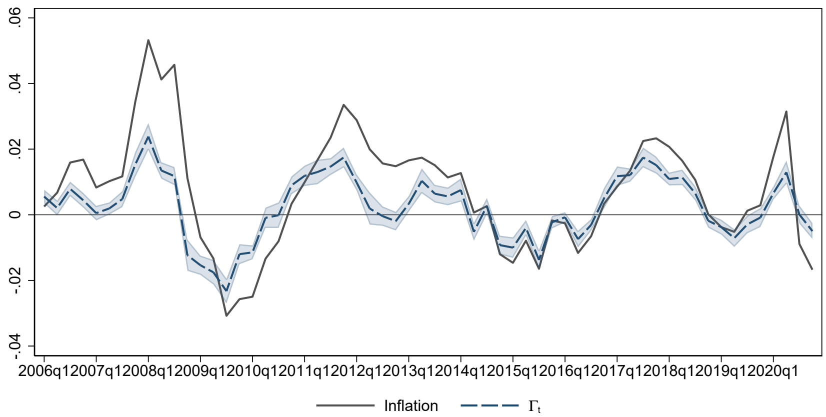
We now develop a decomposition of aggregate inflation into the aggregate component and granular residuals capturing the firm and category dimensions. We then describe the estimation procedure to extract all of these components from the micro price data. Assume that the growth rate in the price of barcode i in country c, approximated by a log difference, is given by:

$$\Delta p_{ifgct} = \delta_{ct} + \boldsymbol{\lambda}_{fc}\boldsymbol{\eta}_{ct}^{F} + \boldsymbol{\lambda}_{gc}\boldsymbol{\eta}_{ct}^{G} + \delta_{fct} + \delta_{gct} + \varepsilon_{ifgct}. \tag{3.3}$$

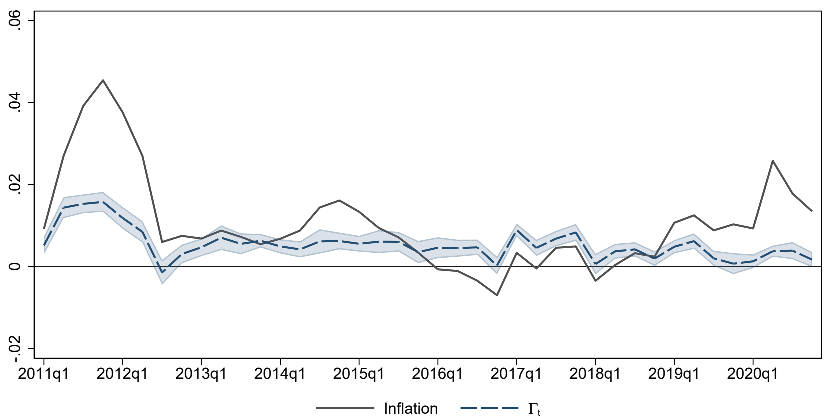
That is, the price change is a function of the aggregate shock δ_{ct}, firm(-country) shock δ_{fct}, category(-country) shock δ_{gct}, the response of firm f's prices to a vector of common shocks $\boldsymbol{\eta}_{ct}^{F}$, the response of category g's prices to a vector of common shocks $\boldsymbol{\eta}_{ct}^{G}$, and an idiosyncratic shock to the barcode ε_{ifgct}. The responses to common shocks are governed by firm- and category-specific loadings $\boldsymbol{\lambda}_{fc}$ and $\boldsymbol{\lambda}_{gc}$. A firm- or category-specific loading on latent aggregate factors may be important in order to absorb heterogeneous firm/category reactions to latent aggregate time-varying variables. For example, the $\boldsymbol{\lambda}$'s might vary because firms f/categories g have different import intensities. Alternatively, variation in $\boldsymbol{\lambda}_{fc}$ could also capture the possibility that large firms adjust prices by less following an aggregate shock. Since this heterogeneous adjustment can ultimately be related to an aggregate source, it is potentially important to keep this separate from the firm-specific shock δ_{fct}.[4] In practice, the baseline analysis will use one common factor per dimension, so the η's and λ's are scalars, but in Appendix Table A4 we repeat the analysis using up to three common factors.

[4]Such a pricing equation could arise, for example, in a market with oligopolistic competition. See Appendix B for a theoretical motivation of our approach following Amiti et al. (2019).

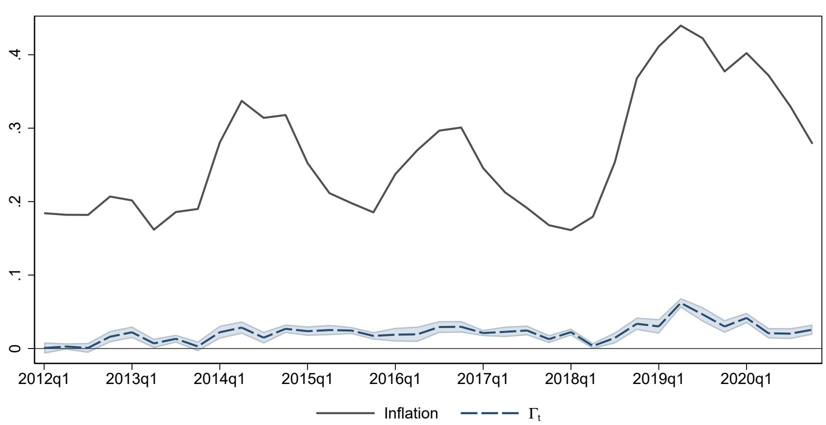
Figure 4: Aggregated retail inflation and simple granular residual



(a) Germany



(b) US



(c) Argentina

Notes: The figure displays the year-on-year overall inflation and the contribution of the simple granular residual. Only three out of 14 countries shown. Rest of the countries can be found in Appendix Figure A3.

Plugging (3.3) into (3.1) delivers the following decomposition:

$$\Delta p_{ct} = \underbrace{\delta_{ct}}_{U_{ct}} + \underbrace{\sum_f w_{fct-4}(\delta_{fct} + \boldsymbol{\lambda}_{fc}\boldsymbol{\eta}^{F}_{ct})}_{\Gamma^f_{ct}} + \underbrace{\sum_g w_{gct-4}(\delta_{gct} + \boldsymbol{\lambda}_{gc}\boldsymbol{\eta}^{G}_{ct})}_{\Gamma^g_{ct}} \qquad (3.4)$$

$$= U_{ct} + \Gamma^f_{ct} + \Gamma^g_{ct},$$

under the assumption that the idiosyncratic deviations from the firm- and category-components are mean-zero in expenditure-weighted terms, $\sum_i w_{ifgct-4}\varepsilon_{ifgct} = 0$.[5] As above, U_{ct} captures the aggregate sources of inflation: the component common to all prices. The firm granular residual Γ^f_{ct} reflects the contributions of firm-specific idiosyncratic components to aggregate inflation, while Γ^g_{ct} reflects the contribution of category-specific components.

The decomposition (3.4) echoes the "classic" one in (3.2), but is richer in two respects. First, it allows for contributions of idiosyncratic shocks in two distinct dimensions: at the firm level Γ^f_{ct}, and at the category level Γ^g_{ct}. Second, it explicitly allows for 2 ways in which large firms can contribute to aggregate price fluctuations. It has been understood since Gabaix (2011) that the granular residual can arise from idiosyncratic shocks to large firms, or from a differential response of the large firms to common shocks. Our granular components encompass both possibilities. The idiosyncratic firm shocks are picked up by the $\sum_f w_{fct-4}\delta_{fct}$ term. The differential response to common shocks is captured by $\sum_f w_{fct-4}\boldsymbol{\lambda}_{fc}\boldsymbol{\eta}^{F}_{ct}$. To understand this term better, suppose for the moment that there is only one common factor, and note that we can write:

$$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct} = \left[\mathrm{Cov}\left(\frac{w_{fct-4}}{\overline{w}_{fct}}, \lambda_{fc}\right) + \overline{\lambda}_{fc}\right]\eta^F_{ct},$$

where $\overline{w}_{fct}$ is the average expenditure share across firms (equalling $1/N_{f\in c,t}$ by construction), and $\overline{\lambda}_{fc}$ is the average firm loading on the common shock. The first term is the covariance between firm size and the loading. It shows that a positive aggregate shock will induce a granular residual if larger firms are on average more reactive to aggregate shocks. The second term is simply the unweighted average firm loading on the aggregate shock. In practice, because we will fit the factor model on demeaned data, this term is negligible. This discussion applies equally to the category granular residual. In the empirical analysis below we will further decompose Γ^f_{ct} and Γ^g_{ct} into these subcomponents, to establish which form of granularity matters most quantitatively.

[5]The idiosyncratic shocks ε_{ifgct} could be extracted from the residuals of the fixed effects regression. However, as we use weighted regressions, the weighted sum of the residuals is zero by construction. We could still subtract the unweighted average of the product idiosyncratic shock (and compute the weighted deviation) but the interpretation of this term would be difficult and not relevant for our analysis.

Shock estimation. In order to decompose aggregate inflation into these components, we must first estimate all of the objects in (3.3). As the first step to estimating δ_{fct} and δ_{gct}, we regress, separately for each period and country, p_{ifgct} on $N_{g\in c,t}$ category fixed effects and $N_{f\in c,t} - 1$ firm fixed effects, setting the average of the fixed effects to zero. That is, if $\hat{\delta}_{dct}$ is the estimated fixed effect for dimension $d = f, g$, country c and quarter t, the unweighted fixed effect will be defined as:

$$\tilde{\delta}_{dct} = \hat{\delta}_{dct} - \frac{1}{N_{d\in c,t}} \sum_{d\in c,t} \hat{\delta}_{dct}.$$

Using the $\tilde{\delta}_{dct}$ directly would imply that there is only one aggregate shock which affects all prices equally. In order relax this assumption and allow for different loadings on a latent aggregate shock, we estimate up to three latent aggregated factors η^{d}_{ct} for the two dimensions $d = f, g$ using Principal Component Analysis on the demeaned fixed effects:

$$\tilde{\delta}_{fct} \quad = \quad \boldsymbol{\lambda_{fc}}\boldsymbol{\eta^{F}_{ct}} + \delta_{fct} \text{ and } \tilde{\delta}_{gct} = \boldsymbol{\lambda_{gc}}\boldsymbol{\eta^{G}_{ct}} + \delta_{gct}.$$

Since the panel is unbalanced, we use the iterative Expectation Maximization algorithm as in Galaasen et al. (2021) and Gabaix and Koijen (2024). This algorithm starts by estimating the principal components based on a balanced panel. It then repeatedly regresses $\tilde{\delta}_{fct}$ on $\boldsymbol{\eta^{F}_{ct}}$ and then $\tilde{\delta}_{fct}$ on $\boldsymbol{\lambda_{fc}}$ until convergence.[6] We perform the same approach for categories, however there the panel is almost balanced. We use the residuals δ_{fct} and δ_{gct} as our firm and category specific idiosyncratic shocks.

The baseline results use one common factor, so $\boldsymbol{\eta^{F}_{ct}}$ and $\boldsymbol{\eta^{G}_{ct}}$ are scalars. In robustness, we report statistics using two and three factors in $\boldsymbol{\eta^{F}_{ct}}$ and $\boldsymbol{\eta^{G}_{ct}}$, even if this leaves some correlation among large firms' residuals. Appendix B shows that under oligopolistic competition large firms (i.e. those with non-negligible market share) react to both their own marginal costs and also to competitors' price adjustments. We hence follow Gabaix and Koijen (2024) and allow for correlation in the large firms' residuals, since the pass-through of an idiosyncratic shock to a large firm could move competitors' prices and therefore the competitors' estimated idiosyncratic firm components.[7] The uncovered

[6] We define convergence and stop the iterations for a specific country c when $0.01 > \frac{1}{N_f}\sum_f |\frac{(\hat{\lambda}^{N}_{fc}\hat{\eta}^{F,N}_{ct} - \hat{\lambda}^{N-1}_{fc}\hat{\eta}^{F,N-1}_{ct})}{\hat{\lambda}^{N-1}_{fc}\hat{\eta}^{F,N-1}_{ct}}|$. That is, when the average percentage change in the contribution of the factor $\hat{\lambda}^{N}_{fc}\hat{\eta}^{F,N}_{ct}$ across firms has changed by less than one percent between the current iteration and the previous one.

[7] See also Amiti et al. (2019). By adding factors to the principal component analysis, we risk muting the most important channels of idiosyncratic shocks' pass-through. We are aware that the estimated residuals then contain a mixture of pass-through of own idiosyncratic shocks and price complementarities arising from other large firms' idiosyncratic shocks that we cannot separate. However, for the focus of this paper this is not problematic since we want to know the aggregate effects of firm idiosyncratic shocks (including how they are multiplied via price complementarities), while we do not focus on the drivers of firm-specific prices.

aggregated shock can then be computed as:

$$\delta_{ct} = \frac{1}{N_g^{ct}} \sum_{g \in c,t} \hat{\delta}_{gct} + \frac{1}{N_f^{ct}} \sum_{d \in c,t} \hat{\delta}_{fct}.$$

All singletons or observations without a defined firm are dropped from the analysis. Categories that in one specific period and country covered less than ten products or five firms were replaced with the category “other retail products."

3.3 Main results

Micro level. We first document how important the firm and category components are for explaining variation in prices at the micro level. We report partial R^2's of the firm and category fixed effects, as well as the total R^2 that would give a sense of how much cross-sectional variation in price changes is due to idiosyncratic factors. The partial R^2 associated with dimension $d = f, g$ and country c is as follows:

$$Partial\ R_d^2 \quad = \quad 1 - \frac{RSS^F}{RSS_d^P},$$

where RSS^F is the sum of squared residuals from the full model (including all fixed effects), and RSS_d^P is the sum of squared residuals from the partial model that include the other (non-d) fixed effects only. We estimate this statistic for each country separately and also pooling across countries. When doing this for each country c separately we use the definition of $RSS^{M,c} = \sum_t \sum_i w_{ifgct-4}(p_{ifgct} - \hat{p}^M_{ifgct})^2$ where $M = \{F, P\}$ is the model (that is, either the full model or the partial model excluding one dimension). When pooling countries, we also sum the squared residuals across countries $RSS_M = \sum_c \sum_t \sum_i w_{ifgct-4}(p_{ifgct} - \hat{p}^M_{ifgct})^2$. We estimate the FEs and the resulting partial R^2 from a weighted regression in which each observation is weighted by its respective expenditure share in the previous year, and from an unweighted regression in which all goods in a given country-period have the same weight.[8] The R^2 for each country is computed with the usual formula.

Table 3 reports the resulting weighted and unweighted R^2 and partial R^2 for each country separately and for all countries together. Overall, the partial R^2's are low, with the firm components responsible for about 10% of the variation in prices when expenditure weights are used, and 7% without expenditure weights. The product component is even less important, with weighted and unweighted partial R^2's of 4% and 1%, respectively. Thus, at the micro level the large majority of

[8]Note that the “unweighted” regressions also contain an implicit weight $w_{ifgct-4} = 1/N_{i \in t,c}$ because we give every period the same weight and the weight of each observation is defined by the number of products observed in a given country-period $N_{i \in t,c}$. For this reason, in both cases there are weights $w_{ifgct-4}$ involved in the computation of the partial R^2.

Table 3: Explanatory power at the disaggregated level

	Unweighted			Weighted		
	Partial R^2			Partial R^2		
Country	Firm	Category	R^2	Firm	Category	R^2
AR	0.096	0.016	0.210	0.113	0.047	0.344
AT	0.057	0.009	0.069	0.077	0.031	0.120
BE	0.067	0.010	0.083	0.091	0.052	0.151
BR	0.104	0.004	0.112	0.145	0.020	0.179
CN	0.117	0.001	0.120	0.159	0.005	0.174
DE	0.060	0.017	0.089	0.101	0.104	0.220
ES	0.082	0.008	0.095	0.118	0.050	0.188
FR	0.045	0.005	0.054	0.050	0.019	0.083
MX	0.051	0.006	0.064	0.120	0.033	0.185
NL	0.057	0.007	0.068	0.088	0.033	0.130
RU	0.085	0.005	0.122	0.116	0.016	0.178
SE	0.061	0.011	0.080	0.091	0.037	0.152
UK	0.040	0.010	0.057	0.058	0.030	0.106
US	0.047	0.004	0.054	0.052	0.018	0.079
All	0.071	0.008	0.148	0.100	0.033	0.244

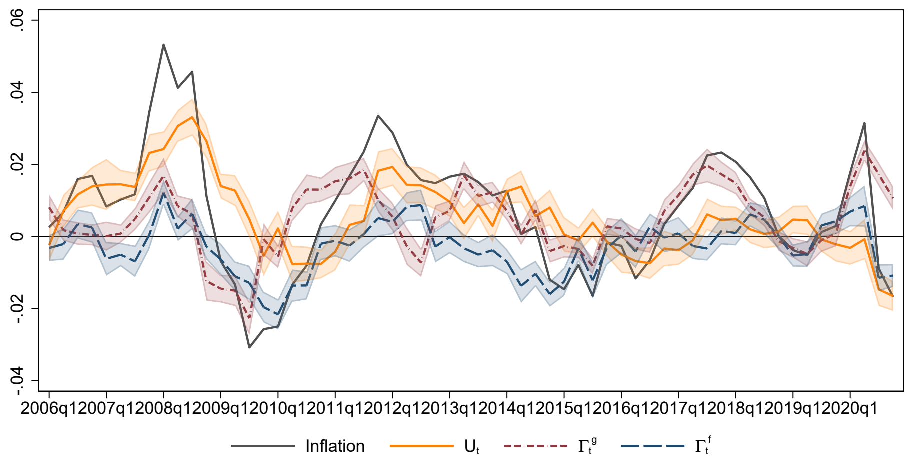
Notes: R^2s and partial R^2s calculated from the the sum of RSS and TSS across periods for each country. Last row shows the measures computed aggregating RSS and TSS also across countries. Unweighted columns display the R^2s resulting from an unweighted regression and weighted columns the R^2s resulting from a weighted regression using the barcode expenditure weights in the same quarter of the previous year.

the variation is idiosyncratic. This echoes the common finding in micro datasets (Haltiwanger, 1997; di Giovanni et al., 2014; Castro et al., 2015). To further explore the firm-level component in price setting, Appendix C uses a multinomial logit specification in the spirit of Bhattarai and Schoenle (2014) to document significant synchronization of price changes within firms.

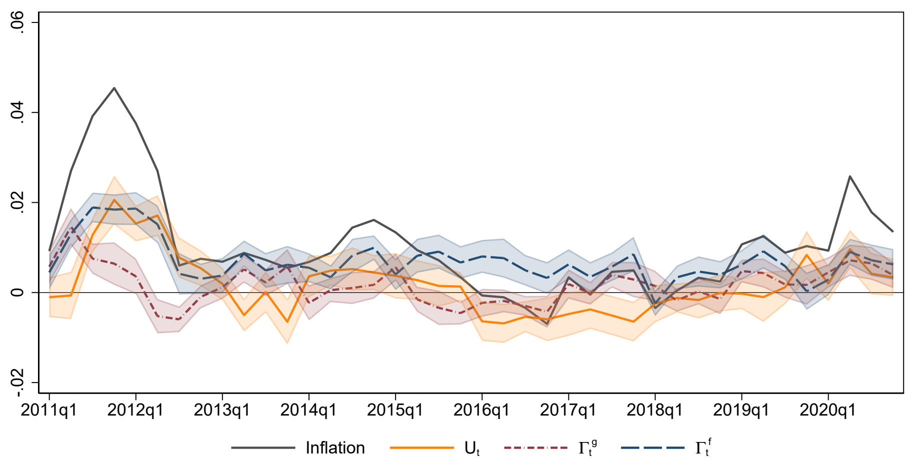
Macro level. We now present our central result: the contribution of individual firms and categories to overall inflation. Figure 5 shows the dynamics of inflation and its components in (3.4) for Germany, the US, and Argentina (for other countries see Appendix Figure A4). We also display 95% confidence intervals estimated using bootstrapping.[9] The firm granular component Γ^f_{ct} appears to contribute significantly to aggregate retail inflation in the advanced economies. The category granular component Γ^g_{ct} is also notable. In Argentina, where inflation is on average around 10 times higher than in the US or Germany, both granular components are relatively less important.

[9] For this we first estimate the components on 30 additional period-country-specific and randomly selected (with replacement) sub-samples of the observations (Δp_{ifgct}) available within each period-country. This guarantees that we estimate the components on the same number of observations in each random sample as in the original data. We then estimate the standard deviation of the components in each period using the bootstrapped samples.

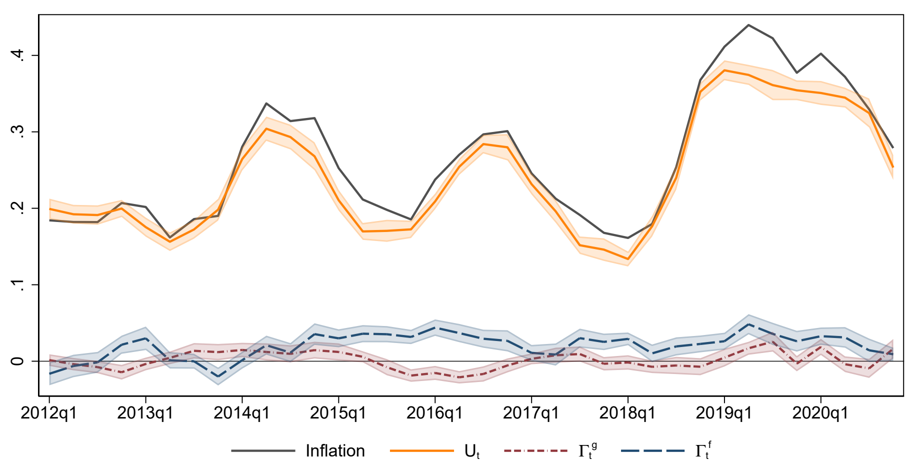
Figure 5: Aggregated retail inflation and granular components



(a) Germany



(b) US



(c) Argentina

Notes: Dynamics of aggregated year-on-year sample inflation and contribution each component displayed. Only three out of 14 countries shown. The rest of the countries can be found in Appendix Figure A4.

Table 4 presents the summary statistics for the aggregate and granular components. The table reports the results for the advanced economies (upper panel, including Austria, Belgium, Germany, Spain, France, the Netherlands, Sweden, the UK and the US) and for the emerging markets (lower panel, including Argentina, Brazil, China, Mexico and Russia) in our sample. Overall inflation has averaged 0.84% in the advanced economies over this period. Of this, the aggregate component contributes 0.38 percentage points, the firm granular component 0.17 percentage points, and the category component 0.30 percentage points. Large firms in our sample, therefore, experienced on average higher price increases than small firms (when controlling for category shocks). This finding introduces a nuance to the literature that examines the general rise in markups, such as De Loecker et al. (2020) and Döpper et al. (2023), and specifically relates to the result that large firms gained market share and increased their profitability in the face of post-Covid global supply chain shortages (e.g. Franzoni et al., 2023). It also relates to the observation that the pass-through of cost shocks into prices is increasing in industry concentration (e.g. Brauning et al., 2022).[10] The standard deviation of the aggregate component is the highest at 1.16 percentage points, followed by Γ^f_{ct} at 0.95 and Γ^g_{ct} at 0.70 percentage points.

All three terms contribute notably to the variability of actual inflation. The correlations between actual inflation and U_{ct}, Γ^f_{ct}, and Γ^g_{ct} are 0.61, 0.66, and 0.43, respectively. The last column of the table reports the Shapley (1953) value of each component in the total variance of inflation. When the components are mutually correlated, the Shapley value is a way of representing the contribution of each component to the total. Essentially, it averages the contribution to the total variance of each component across all permutations of the other components. Conveniently, the sum of Shapley values across all components is 1.[11] The aggregate component U_{ct} contributes 43% to the variance of the overall inflation, followed by 38% for the granular firm component, and 19% for the granular category component. Thus, in the advanced economy sample, granular components account for more than half of the total variance of inflation over this period.

The results are quite different for the emerging markets. Here, overall inflation is much higher (7.37% on average), and the aggregate component is much more important,contributing 6.35 percentage points on average. While all three components have a substantial correlation with the overall inflation, the Shapley values for the variance shares of the firm and category granular components are 11% and 2%, respectively, suggesting that in higher-inflation environments granular effects are much less important.

[10]The contributions of the granular components to mean inflation are lower bounds, as due to the intercept issue in the fixed effects regressions we renormalize the averages of firm and category fixed effects to 0. Thus, the positive averages Γ^f_{ct} and Γ^g_{ct} are entirely due to prices of larger firms/categories growing faster on average.

[11]In practice, the correlations between U_{ct}, Γ^f_{ct}, and Γ^g_{ct} are limited, and simply computing the variance share leads to substantively similar results as the Shapley values.

Table 4: Summary statistics and correlations of factor components

	Mean	St. Dev.	Corr.	Var(Δp_{ct}) share
Advanced Economies (N. Obs. = 457)				
Δp_{ct}	0.84	1.63	1.00	1.00
U_{ct}	0.38	1.16	0.61	0.43
Γ^f_{ct}	0.17	0.95	0.66	0.38
$\sum_f w_{fct-4}\delta_{fct}$	0.16	0.91	0.64	0.35
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	0.00	0.27	0.15	0.03
$\Gamma^{f\in top10f}_{ct}$	*0.10*	*0.61*	*0.64*	*0.25*
$\Gamma^{f\notin top10f}_{ct}$	*0.07*	*0.50*	*0.47*	*0.13*
Γ^g_{ct}	0.30	0.70	0.43	0.19
$\sum_g w_{gct-4}\delta_{gct}$	0.24	0.51	0.22	0.07
$\sum_g w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.06	0.46	0.41	0.12
Emerging Markets (N. Obs. = 180)				
Δp_{ct}	7.37	10.60	1.00	1.00
U_{ct}	6.35	9.97	0.61	0.87
Γ^f_{ct}	0.99	1.49	0.66	0.11
$\sum_f w_{fct-4}\delta_{fct}$	1.00	1.43	0.64	0.11
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	-0.01	0.41	0.15	-0.00
$\Gamma^{f\in top10f}_{ct}$	*0.52*	*0.93*	*0.64*	*0.03*
$\Gamma^{f\notin top10f}_{ct}$	*0.48*	*0.80*	*0.47*	*0.08*
Γ^g_{ct}	0.02	1.05	0.43	0.02
$\sum_g w_{gct-4}\delta_{gct}$	0.00	0.99	0.22	0.02
$\sum_g w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.02	0.33	0.41	-0.01

Notes: "Mean" denotes the average inflation rate, "St. Dev." the standard deviation, "Corr" the correlation between the component in the row and aggregated sample inflation Δp_{ct}, and "Var(Δp_{ct}) share" denotes the share of the variance of actual inflation accounted for by each component, as measured by the Shapley values. The top panel reports the results computed pooling advanced economies (Austria, Belgium, Germany, Spain, France, the Netherlands, Sweden, UK, US) and the bottom panel emerging markets (Argentina, Brazil, China, Mexico and Russia).

We next undertake two further decompositions to highlight the nature of inflation granularity. First, as discussed above, Γ^f_{ct} can arise either because of idiosyncratic shocks to large firms (the $\sum_f w_{fct-4}\delta_{fct}$ subcomponent), or from higher responsiveness of large firms to common shocks (the $\sum_f w_{fct-4}\boldsymbol{\lambda_{fc}\eta^F_{ct}}$ subcomponent), and similarly for the category component Γ^g_{ct}. Table 4 decomposes Γ^f_{ct} and Γ^g_{ct} into the subcomponents, see equation (3.4). For the firm granular component, there is a clear winner: idiosyncratic shocks. This component is responsible for virtually all of the average growth in Γ^f_{ct} (0.16 percentage points of the total of 0.17), and contains nearly all of the variability of Γ^f_{ct} (standard deviation of 0.91 out of the total of 0.95). Of the total 38% contribution of Γ^f_{ct} to the aggregate inflation variance, the firm idiosyncratic component accounts for 0.35 percentage points.

For the category granular component it is more split. The idiosyncratic component accounts for 0.24% out of the total 0.30% average growth, and the standard deviations of the two subcomponents are similar at 0.5%. The contribution of the differential sensitivity to aggregate shocks to the variability of aggregate inflation is actually larger, at 12% out of the total of 19%.

Next, we highlight the role of large firms by separating the Γ_{ct}^{f} additively into the components accounted for by the top 10 firms vs. the rest. The top 10 firms are an important source of granular fluctuations. The top 10 firms alone are responsible for 0.1 out of the 0.17 percentage points average growth, and for 0.25 out of the total 0.38 contribution of Γ_{ct}^{f} to the aggregate inflation variance.

Appendix Table A4 reports as robustness the above-discussed statistics for a sample using a simplified approach for identifying missing firms and up to three factors. Using a simpler methodology for matching firms does not change our estimates significantly.[12] Adding more factors decreases the importance of the idiosyncratic shocks for both the firm and category dimensions. However, the firm idiosyncratic shocks still account for the majority of the variability of Γ_{ct}^{f}: with three factors the firm idiosyncratic component still accounts for 0.29 of the 0.38 contribution of Γ_{ct}^{f} to the aggregate inflation variance. On the other hand, with more factors the product category idiosyncratic shocks account for a lower share of the overall Shapley value contribution of Γ_{ct}^{g}.

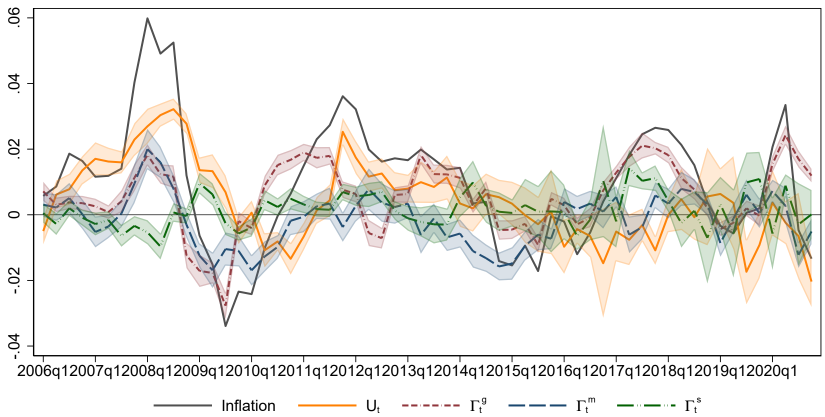
3.4 The retailer dimension

We next present the results also taking into account the retailer dimension. As noted above, we do not adopt this decomposition as the baseline because the retailer information in these data is imperfect. Many of the transactions are coded as "other" retailer, and we have to make a decision on how to assign a retailer component to those. We perform two versions. In the main text, we report the results when dropping observations with an undefined retailer. In Appendix Table A5, we report results when assigning these observations to an "other" retailer specific to the region of the household.

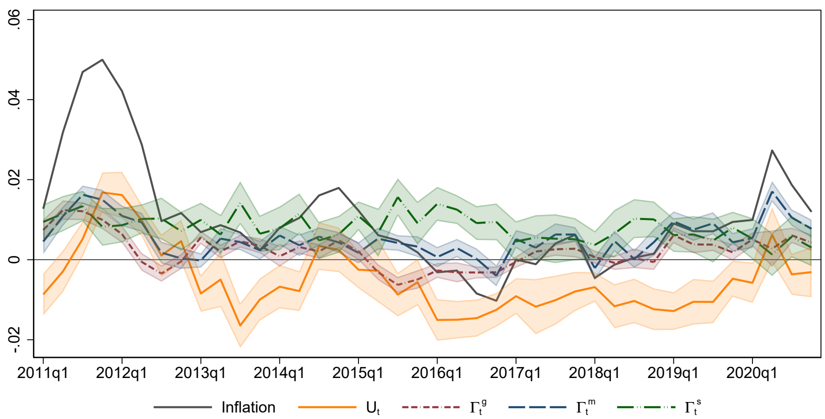
Figure 6 and Table 5 reproduce the main results with the retailer dimension. Adding the retailer component leaves the firm and category granular component quite similar to the baseline, but reduces the importance of the aggregate components. The contribution to the variance of the aggregate component falls from 43% to 32%, and the difference is largely picked up by the retailer component, which accounts for 14% of the inflation variance. In order to assess if this is caused by a change in the underlying sample of observations, in panel B of Appendix Table A5 we only estimate Γ_{ct}^{f} and Γ_{ct}^{g} in the retailer sample. One can see that the contribution of U_{ct}, Γ_{ct}^{f} and Γ_{ct}^{g} to the variance of aggregate retail inflation remains roughly unchanged when using the retailer sample, indicating that this result is not driven by changes in the underlying sample.

[12] See Appendix A.1 for a detailed discussion of the alternative firm matching procedures.

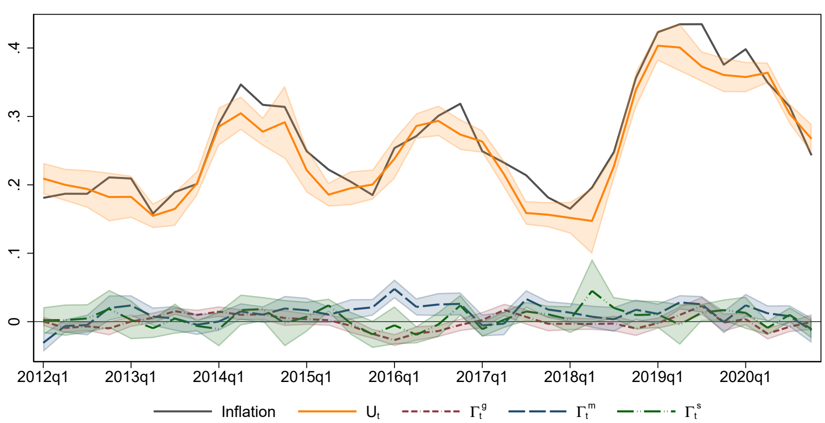
Figure 6: Aggregated retail inflation and granular components



(a) Germany



(b) US



(c) Argentina

Notes: Dynamics of aggregated year-on-year sample inflation and contribution each component displayed. Only three out of 14 countries shown. Rest of the countries can be found in Appendix Figure A5.

Table 5: Retailer sample - Summary statistics and correlations of factor components

	Mean	St. Dev	Corr	Var(Δp^r_{ct}) share
Advanced Economies (N. Obs = 457)				
Δp^r_{ct}	1.10	1.78	1.00	1.00
U_{ct}	0.19	1.46	0.35	0.32
Γ^f_{ct}	0.11	0.93	0.64	0.34
$\sum_f w_{fct-4}\delta_{fct}$	0.12	0.86	0.61	0.29
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	-0.00	0.33	0.20	0.05
Γ^g_{ct}	0.36	0.72	0.48	0.20
$\sum_g w_{gct-4}\delta_{gct}$	0.31	0.53	0.32	0.09
$\sum_g w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.05	0.47	0.38	0.11
Γ^s_{ct}	0.44	1.17	0.28	0.14
$\sum_s w_{sct-4}\delta_{sct}$	0.47	1.12	0.27	0.13
$\sum_s w_{sct-4}\lambda_{sc}\eta^S_{ct}$	-0.04	0.35	0.07	0.01
Emerging Markets (N. Obs = 180)				
Δp^r_{ct}	7.33	10.74	1.00	1.00
U_{ct}	6.42	10.37	0.35	0.92
Γ^f_{ct}	0.46	1.20	0.64	0.05
$\sum_f w_{fct-4}\delta_{fct}$	0.47	1.15	0.61	0.05
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	-0.01	0.28	0.20	0.00
Γ^g_{ct}	0.02	0.98	0.48	0.00
$\sum_g w_{gct-4}\delta_{gct}$	0.00	0.84	0.32	0.02
$\sum_g w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.01	0.54	0.38	-0.02
Γ^s_{ct}	0.43	1.19	0.28	0.03
$\sum_s w_{sct-4}\delta_{sct}$	0.39	1.06	0.27	0.02
$\sum_s w_{sct-4}\lambda_{sc}\eta^S_{ct}$	0.03	0.60	0.07	0.01

Notes: "Mean" denotes the average inflation rate, "St. Dev." the standard deviation, "Corr" the correlation between the component in the row and aggregated sample inflation Δp^r_{ct} using the product-retailer level dataset, and "Var(Δp_{ct}) share" denotes the share of the variance of actual inflation accounted for by each component, as measured by the Shapley values. The top panel reports the results computed pooling all advanced economies and the bottom panel all emerging markets. Δp^r_{ct} refers to aggregated inflation computed using the retailer-country-quarter level sample, which slightly differs from the aggregated inflation in the baseline sample Δp_{ct}. The top panel reports the results computed pooling advanced economies (Austria, Belgium, Germany, Spain, France, the Netherlands, Sweden, UK, US) and the bottom panel emerging markets (Argentina, Brazil, China, Mexico and Russia).

3.5 Granularity and international inflation synchronization

Many of the firms present in our dataset span multiple countries. This raises the question whether synchronized pricing by these multi-country firms contributes to the well-known substantial comovement in inflation across countries. For each country c, we compute the average inflation of all the countries other than c:

$$\Delta p_{-ct} = \frac{1}{N_c - 1} \sum_{d \neq c} \Delta p_{dt}.$$

Then the correlation between one country and the rest of the world is given by:

$$\rho(\Delta p_{ct}, \Delta p_{-ct}) = \frac{\text{Cov}(\Delta p_{ct}, \Delta p_{-ct})}{\sigma_c \sigma_{-c}}, \tag{3.5}$$

Table 6: Inflation correlations and contributions

Shapley value	All pairs	AE-AE pairs	EM-EM pairs
U_{ct}	0.31	0.42	1.28
Γ^f_{ct}	0.38	0.38	0.26
Γ^g_{ct}	0.31	0.20	-0.54
Mean Corr Δp_{ct}	0.12	0.47	-0.10

Notes: This table reports the shares of each subcomponent in explaining the average inflation correlation in each subsample of countries: all pairs, advances country pairs, emerging market pairs. The share of inflation correlation due to each component is measured by its Shapley value. The bottom row displays the average correlation of inflation.

where σ_c and σ_{-c} are the standard deviations of inflation in c and in the rest of the sample, respectively.

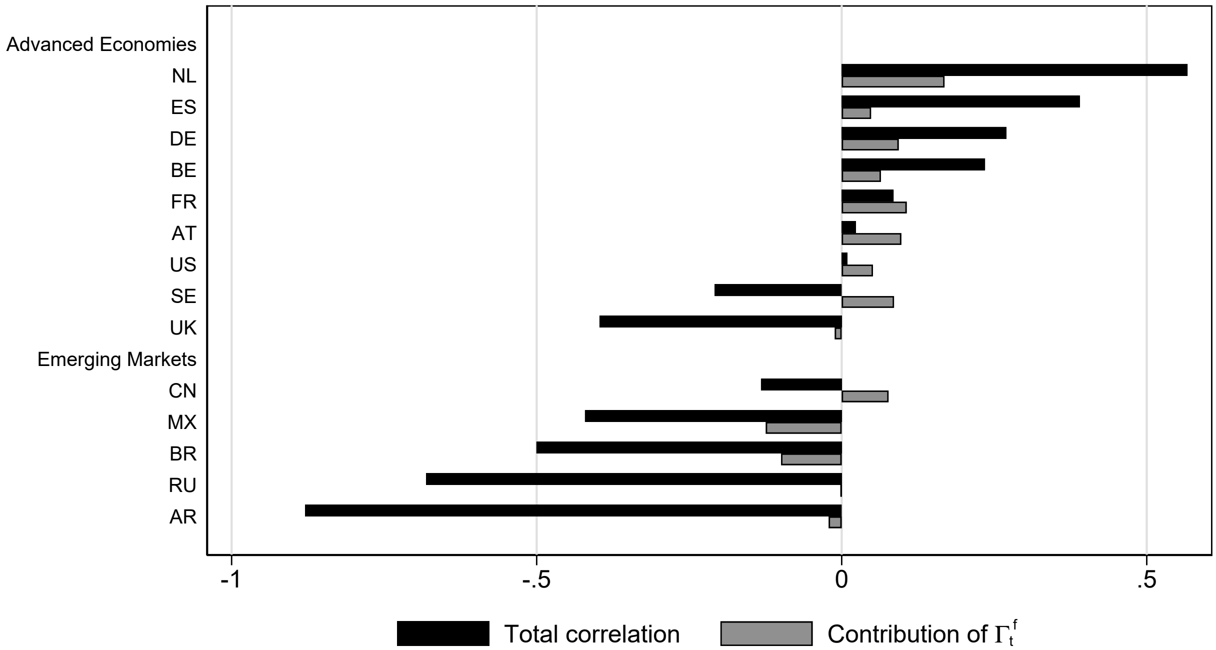
To focus on the contribution of firms, we isolate Γ^f_{ct}: $\Delta p_{ct} = \Delta\widetilde{p}_{ct} + \Gamma^f_{ct}$, where $\Delta\widetilde{p}_{ct} = U_{ct} + \Gamma^g_{ct}$ collects the other components of inflation. The global average is similarly decomposed into the firm granular and the other components. Then, inflation correlation can be written as:

$$\begin{aligned}\rho(\Delta p_{ct}, \Delta p_{-ct}) &= \frac{\mathrm{Cov}(\Delta\widetilde{p}_{ct} + \Gamma^f_{ct}, \Delta\widetilde{p}_{-ct} + \Gamma^f_{-ct})}{\sigma_c \sigma_{-c}} \\ &= \underbrace{\rho(\Gamma^f_{ct}, \Gamma^f_{-ct}) \frac{\sigma_{\Gamma^f_c} \sigma_{\Gamma^f_{-c}}}{\sigma_c \sigma_{-c}}}_{\text{Contribution of } \Gamma^f} + \underbrace{\rho(\Delta\widetilde{p}_{At}, \Delta\widetilde{p}_{Gt}) \frac{\sigma_{\Delta\widetilde{p}_c} \sigma_{\Delta\widetilde{p}_{-c}}}{\sigma_c \sigma_{-c}}}_{\text{Contribution of other}} \\ &\quad \underbrace{+Z + Y.}_{\text{Cross-term correlations}}\end{aligned} \tag{3.6}$$

The black bars in Figure 7 plot the correlation of each country's inflation with the global average, (3.5). The grey bars depict the contribution of Γ^f, as in (3.6). In most advanced economies the firm granular component adds around 0.1 to the correlation of inflation with the rest of the world, indicating that there is a contribution of this component to the comovement of inflation at least in these countries. Surprisingly, this is not the case in emerging markets, but this could be caused by the small share of expenditures in common firms.

We can also use the Shapley value decomposition to compute the contributions of each of the components U_{ct}, Γ^f_{ct}, and Γ^g_{ct} to average overall inflation comovement. Table 6 displays the results. Inflation comovement is highest among the advanced country pairs, at 0.47. Of this, 38% is accounted for by the firm granular component, and 20% by the category granular component. The results are quite different in the emerging market subsample, as the average correlation is negative at −0.1. Here, the firm granular component still contributes increasing comovement, but the product component exerts a substantial negative effect, acting to reduce inflation comovement.

Figure 7: Correlation with global inflation and the contribution of the firm granular residual (only periods after 2008 Q1 kept)



Notes: The black bars depict, for each country, the correlation between its inflation and the average inflation in the rest of the sample, $\rho(\Delta p_{ct}, \Delta p_{-ct})$. The gray bars depict the contribution of the firm granular component, $\rho(\Gamma^f_{ct}, \Gamma^f_{-ct})\frac{\sigma_{\Gamma^f_c}\sigma_{\Gamma^f_{-c}}}{\sigma_c\sigma_{-c}}$.

4. Conclusion

A sizeable and growing literature has established that large firms play an important role in the economy, and that idiosyncratic shocks to these firms contribute substantially to macroeconomic fluctuations. However, there has been no empirical evidence on how inflation is affected by this phenomenon.

This paper uses barcode-level data for 14 countries and an extension of the granular residual methodology of Gabaix (2011) to study the role of individual firms and categories in the overall inflation. Indeed, we find that in a low-inflation environment that characterizes the advanced economies over our sample period, idiosyncratic firm components explain a substantial share – nearly 40% – of the variance of inflation. Shocks to categories explain an additional 20%, implying that most of the variability of inflation prior to 2021 in advanced economies was due to granular sources. The picture is quite different in emerging markets, where the overall inflation is higher, and the all the granular components combined contribute less than 15% to the variation in inflation. We also examine the role of large retailers for fluctuations in overall inflation, finding that is has a moderate role.

Our methodology allows us to decompose the overall granular residuals into the parts due to truly idiosyncratic shocks, and due to the greater responsiveness of large firms to aggregate shocks. We find that the former is by far more important.

Last, we examine the importance of granularities and the presence of large firms in multiple countries for inflation co-movement across countries, finding a moderate synchronizing effect.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80, 1977–2016.

Alvarez, F. and F. Lippi (2014): "Price setting with menu cost for multiproduct firms," *Econometrica*, 82, 89–135.

Amiti, M., O. Itskhoki, and J. Konings (2019): "International shocks, variable markups, and domestic prices," *The Review of Economic Studies*, 86, 2356–2402.

Amiti, M. and D. E. Weinstein (2018): "How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data," *Journal of Political Economy*, 126, 525–587.

Auer, R., A. Burstein, and S. M. Lein (2021): "Exchange Rates and Prices: Evidence from the 2015 Swiss Franc Appreciation," *American Economic Review*, 111, 652–86.

Beck, G. W., K. Carstensen, J.-O. Menz, R. Schnorrenberger, and E. Wieland (2023): "Nowcasting consumer price inflation using high-frequency scanner data: Evidence from Germany," Tech. rep.

Beck, G. W. and X. Jaravel (2021): "Prices and Global Inequality: New Evidence from Worldwide Scanner Data," Tech. rep.

Bhattarai, S. and R. Schoenle (2014): "Multiproduct firms and price-setting: Theory and evidence from US producer prices," *Journal of Monetary Economics*, 66, 178–192.

Bonomo, M., C. Viana de Carvalho, O. Kryvtsov, S. Ribon, and R. Rigato (2020): "Multi-product pricing: Theory and evidence from large retailers in Israel," Bank of Canada Staff Working Paper, No. 2020-12.

Braun, R. and S. M. Lein (2021): "Sources of Bias in Inflation Rates and Implications for Inflation Dynamics," *Journal of Money, Credit and Banking*, 53, 1553–1572.

Brauning, F., J. L. Fillat, and G. Joaquim (2022): "Cost-Price Relationships in a Concentrated Economy," Working paper.

Bremus, F., C. M. Buch, K. N. Russ, and M. Schnitzer (2018): "Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity," *Journal of Money, Credit and Banking*, 50, 1785–1825.

Burya, A. and S. Mishra (2022): "Variable markups, demand elasticity and pass-through of marginal costs into prices," .

Carvalho, C. (2006): "Heterogeneity in price stickiness and the real effects of monetary shocks," *Frontiers in Macroeconomics*, 2.

Carvalho, V. and X. Gabaix (2013): "The great diversification and its undoing," *American Economic Review*, 103, 1697–1727.

Carvalho, V. M. (2014): "From micro to macro via production networks," *Journal of Economic Perspectives*, 28, 23–48.

Castro, R., G. Clementi, and Y. Lee (2015): "Cross-Sectoral Variation in The Volatility of Plant-Level Idiosyncratic Shocks," *Journal of Industrial Economics*, 63, 1–29.

Cox, L., G. Müller, E. Pasten, R. Schoenle, and M. Weber (2020): "Big g," Tech. rep., National Bureau of Economic Research.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 135, 561–644.

DellaVigna, S. and M. Gentzkow (2019): "Uniform Pricing in U.S. Retail Chains," *The Quarterly Journal of Economics*, 134, 2011–2084.

di Giovanni, J. and A. A. Levchenko (2012): "Country size, international trade, and aggregate fluctuations in granular economies," *Journal of Political Economy*, 120, 1083–1132.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014): "Firms, destinations, and aggregate fluctuations," *Econometrica*, 82, 1303–1340.

——— (2018): "The micro origins of international business-cycle comovement," *American Economic Review*, 108, 82–108.

——— (2024): "Foreign shocks as granular fluctuations," *Journal of Political Economy*, 132, 391–433.

Döpper, H., A. MacKay, N. Miller, and J. Stiebale (2023): "Rising Markups and the Role of Consumer Preferences," Working Paper No. 22-025, Harvard Business School Strategy Unit.

Eichenbaum, M., N. Jaimovich, S. Rebelo, and J. Smith (2014): "How Frequent Are Small Price Changes?" *American Economic Journal: Macroeconomics*, 6, 137–55.

Franzoni, F. A., M. Giannetti, and R. Tubaldi (2023): "Supply Chain Shortages, Large Firms' Market Power, and Inflation," Research Paper No. 23-105, Swiss Finance Institute.

Friedman, M. (1963): *Inflation: Causes and Consequences*, New York: Asia Publishing House.

GABAIX, X. (2011): "The granular origins of aggregate fluctuations," *Econometrica*, 79, 733–772.

GABAIX, X. AND R. S. KOIJEN (2024): "Granular Instrumental Variables," Forthcoming, *Journal of Political Economy*.

GALAASEN, S., R. JAMILOV, R. JUELSRUD, AND H. REY (2021): "Granular Credit Risk," Working Paper 27994, National Bureau of Economic Research.

GALÍ, J. (2015): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*, Princeton, NJ: Princeton University Press.

GARCÍA-LEMBERGMAN, E. (2022): "Multi-establishment Firms, Pricing and the Propagation of Local Shocks: Evidence from US Retail," Mimeo, Pontifical Catholic University of Chile.

GAUBERT, C. AND O. ITSKHOKI (2021): "Granular comparative advantage," *Journal of Political Economy*, 129, 871–939.

GAUBERT, C., O. ITSKHOKI, AND M. VOGLER (2021): "Government Policies in a Granular Global Economy," *Journal of Monetary Economics*.

HALTIWANGER, J. C. (1997): "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence," *Federal Reserve Bank of St.Louis Review*, 79, 55–77.

HOTTMAN, C. J., S. J. REDDING, AND D. E. WEINSTEIN (2016): "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 131, 1291–1364.

JOVANOVIC, B. (1987): "Micro shocks and aggregate risk," *The Quarterly Journal of Economics*, 102, 395–409.

KAPLAN, G. AND S. SCHULHOFER-WOHL (2017): "Inflation at the household level," *Journal of Monetary Economics*, 91, 19–38.

LONG, J. B. AND C. I. PLOSSER (1983): "Real business cycles," *Journal of political Economy*, 91, 39–69.

MIDRIGAN, V. (2011): "Menu costs, multiproduct firms, and aggregate fluctuations," *Econometrica*, 79, 1139–1180.

PASTÉN, E., R. SCHOENLE, AND M. WEBER (2023): "Sectoral Heterogeneity in Nominal Price Rigidity and the Origin of Aggregate Fluctuations," Forthcoming, *American Economic Journal: Macroeconomics*.

REDDING, S. J. AND D. E. WEINSTEIN (2019): "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences*," *The Quarterly Journal of Economics*, 135, 503–560.

Shapley, L. S. (1953): "A Value for N-Person Games," in *Contributions to the Theory of Games*, ed. by A. Tucker and R. D. Luce, Princeton, NJ: Princeton University Press, 31–40.

Woodford, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton, NJ: Princeton University Press.

Appendix

A. Data construction

A.1 Identifying firms

We adopt a five-step procedure to harmonize firm names across countries. First, for the largest firms, we manually match the brands to the firms that own them and harmonize variations in their names. This helps us to fill missing firm information in instances where we have brand information but no firm information and to replace instances where the brand is listed in the firm field with the firm name. For example, we use "Unilever" as the firm name whenever the brands are "Dove," "Knorr," or "Ben & Jerry's." We also harmonize firm variations such as "Company Unilever" and "Unilever International" to the unique firm name "Unilever". Second, we fill in missing information on the firms using barcodes with the same prefix (the first eight digits).[13] To do this, we sort our data by barcode. If a product without firm information shares the same barcode prefix with both the previous and the next product in this sorted list of barcodes, and both the previous and the next product have the same firm identifier, we use the firm name also for the middle product. Third, for the remaining products with missing firm identifier, we use the most common firm name within the eight-digit prefix and country. This is motivated by the methodology used in Hottman et al. (2016) and recently in Burya and Mishra (2022) and confirmed by manual checks of the allocated barcodes and their ownerships in GS1.[14]

In the fourth step, we append the data for all countries and use information from the overlap of barcodes across countries. If we observe that a given barcode is always associated with the same firm in some countries, we also use this firm name in countries in which the firm name was missing in the original data.

Specifically, if firm "X" from a country was matched in N barcodes from another country and it was always matched to the same firm "Y," we populate the firm name with "X" in this country for all barcodes identified to the firm "Y," and also all the barcodes identified with firm "Y" without a barcode match. We do this bilaterally for all countries and barcodes that had so far not been matched with a firm in the previous bilateral combination.[15]

In the final step, we implement the fuzzywuzzy string matching algorithm in order to match brands and firms that are similar but not exactly identical across countries. The version of the algorithm we employ not only uses the Levenshtein distance to measure the distance between different words but in addition it tokenizes the strings, gets rid of punctuation, takes out the common tokens and measures the standard Levenshtein distance similarity ratio out of pairwise combinations of the tokens.[16] We

[13]Typically, a barcode has eight to 13 digits. It is assigned to products by GS1, a global collaboration platform, that assigns unique barcodes to products. Firms have to apply for these barcodes with GS1 and are usually identified in the first seven to eleven digits of the barcode, which is what we refer to as "prefix", as described also in Hottman et al. (2016).

[14]For these manual checks, we relied on the GS1 search tool (https://gepir.gs1.org/index.php/search-by-gtin) to retrieve firm information for a subset of barcodes lacking these data and also on the adjacent barcodes in the sorted data with available firm information as explained in the main text. The website was accessed in March 2023.

[15]On the other hand, if firm "X" from a country was matched N times but to different firms, we do not replace it for the barcodes which did not have a match. This step helps to fill missing information and to match differently labelled firms especially in countries sharing European Article Number (EAN) barcodes, since those are unique across countries.

[16]We also added a penalty to the 1000 most-repeated words across all firms. In our algorithm, frequently repeated words such as "international" or "bio" are only taken into account if the matching score without these words is still better than score of the next best match. That is, we also estimate the distance between the two strings excluding often repeated words, and the resulting score needs to be above the threshold score used and above the score of the second best possible match. This would, for example, stop us from matching "International bio Unilever" to "International bio Mars" and instead give preference to "Unilever" for the match.

implemented the algorithm by looping across countries. That is, ordering the countries by the number of observations, we first match the brands of all countries with the brands of the country with the most observations, thereafter the brands which have not been matched yet with the brands of the country with the second-most observations, and so forth. Appendix A.1 provides additional details on the outcomes of this matching process. Appendix Table A4 also provides the estimates from the empirical analysis without implementing the last two steps.

Table A1 reports summary statistics before and after the matching process for firms. Panel A reports overview statistic from the original data.[17] Panel A shows that many of the national datasets have a large share of observations without identified firm, and most of the firms are national only (ie observed on one market only).

Table A1: Firms before vs after matching procedure in quarterly data

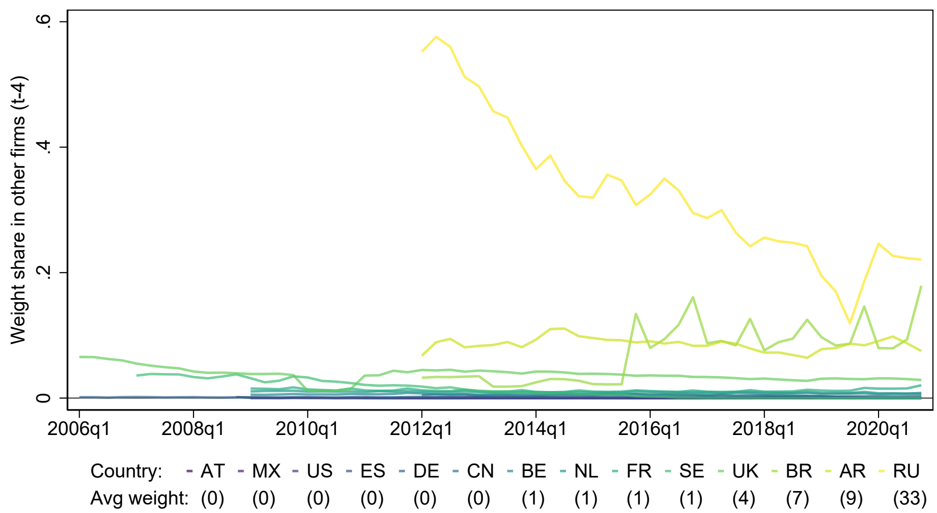
		A: Firms in original data			B: Firms after matching			
	Obs	Missing	Number	Int.	Missing	Number	Int.	Years
AR	656,574	0.18	2,896	0.10	0.05	3,926	0.24	2011-2020
AT	1,150,185	0.00	4,546	0.46	0.00	4,259	0.71	2008-2020
BE	2,091,574	0.02	14,117	0.47	0.01	11,288	0.65	2008-2020
BR	1,166,165	0.15	11,285	0.05	0.03	14,226	0.12	2011-2020
CN	3,724,740	0.46	61,733	0.03	0.01	72,328	0.07	2011-2020
DE	5,963,100	0.02	8,938	0.30	0.01	10,195	0.57	2005-2020
ES	2,947,746	0.01	13,030	0.11	0.00	13,362	0.29	2007-2020
FR	4,676,088	0.18	3,156	0.41	0.05	6,007	0.70	2008-2020
MX	752,202	0.00	4,085	0.11	0.00	4,128	0.22	2011-2020
NL	2,817,307	0.08	11,768	0.53	0.03	9,856	0.68	2008-2020
RU	2,063,858	0.03	13,533	0.09	0.02	12,950	0.19	2011-2020
SE	789,373	0.02	3,323	0.30	0.02	3,065	0.48	2006-2020
UK	4,656,687	0.06	6,448	0.20	0.06	6,298	0.35	2005-2020
US	12,638,612	0.01	36,523	0.05	0.00	35,625	0.12	2010-2020
Total	46,094,211	0.08	177,905	0.06	0.02	174,554	0.10	2005-2020

Notes: “Obs” are the number of product-country-YoY differences available using quarterly frequency. “Missing” is the share of these observations for which the manufacturer could not be found. “Number” is the number of different firms available and “Int.” (international) is the share of these different firms which is also observed in at least one other country.

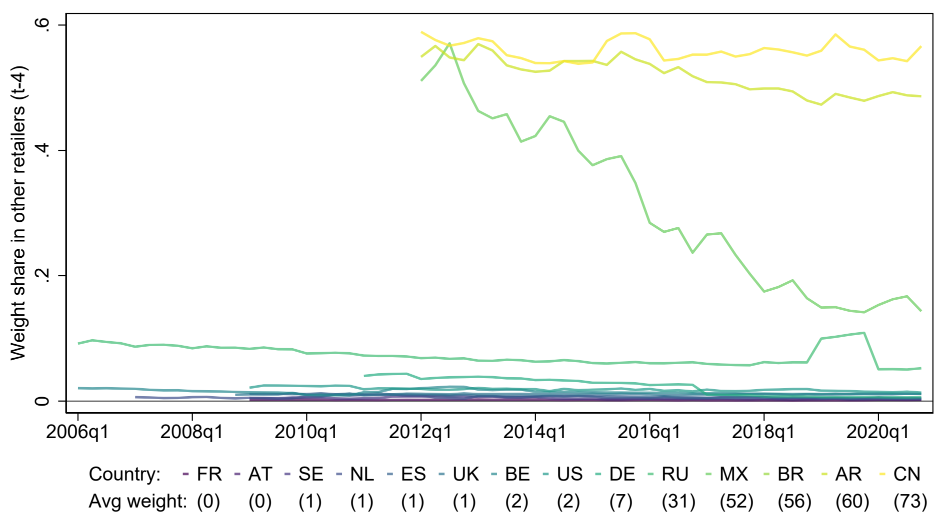
Panel B in Table A1, reports the same descriptive statistics as in Panel A after the matching procedure described in the main text. From comparison of the country-specific statistics in panel A with panel B of Table A1, it is evident that the number of observations with missing firms strongly declines. This is mainly because we found the information in another country using the same unique barcode or because we used available brand information instead. The later step results in a larger number of firms available in some countries after the matching procedure. Second, we can see

[17] The observations included in table A1 and throughout the analysis already contain some minor adjustments on the barcodes of some countries that had an extra digit or prefix. For example, in the French data, the barcodes had a prefix with either zeros or a digit denoting products from a specific shop. In addition, for finding missing firms we had to find all the country-specific labels for “other" firms and replace them with “other".

Figure A1: Share of expenditure weight in not identified retailers and firms



(a) Market share of unidentified firms

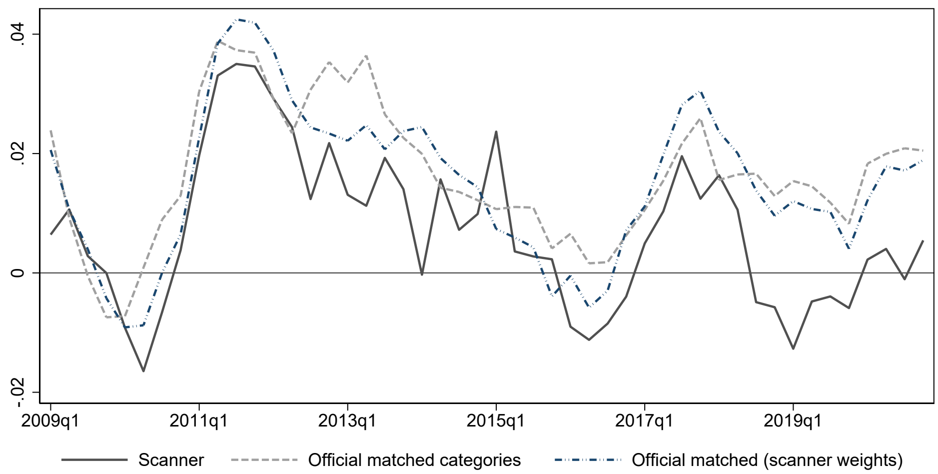


(b) Market share of unidentified retailers

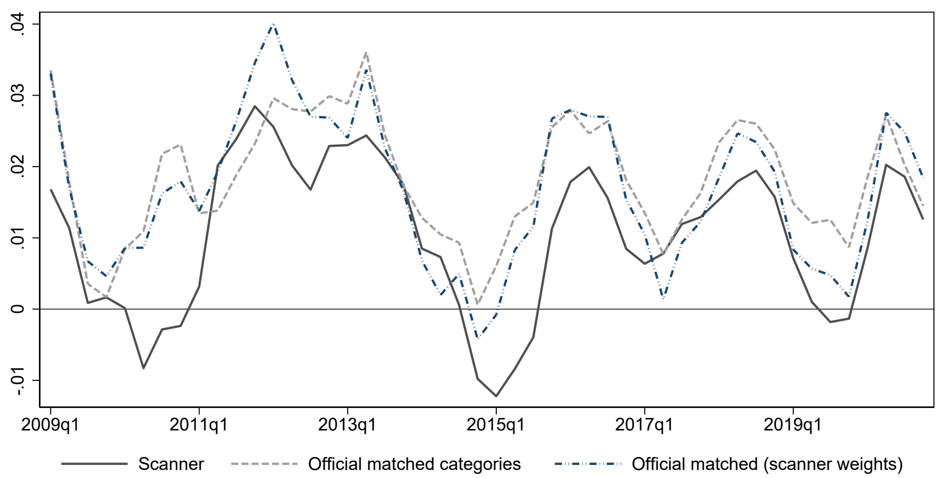
Notes: The figure depicts the total share of expenditure for which the firm (top panel) and retailer (bottom panel) cannot be identified.

that from the available firms in each country, the share of those that appear in at least a second country strongly increases. For most European countries this number is well above 50%. Finally, we observe a decline in Finally when looking at the pooled numbers, the total amount of unique firms across countries declines by around 10% and the share of observations with missing firm information declines from 10% to zero.

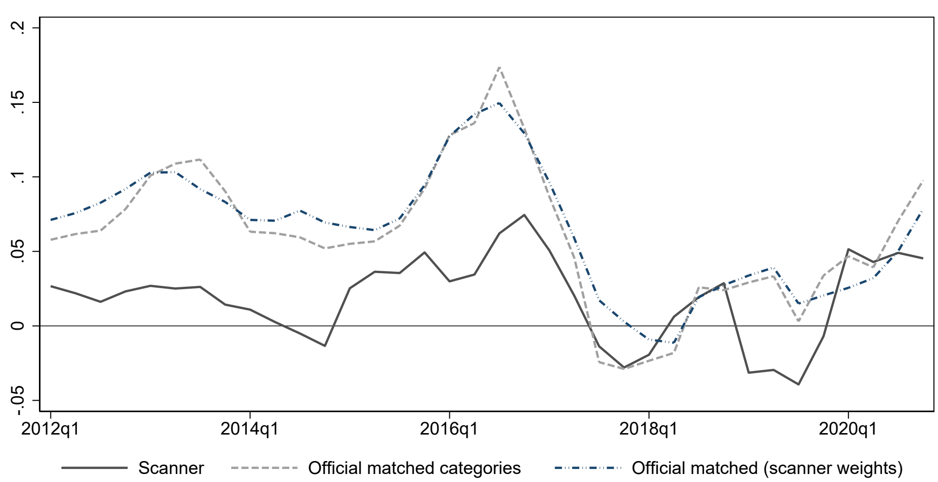
Figure A2: Official vs scanner data aggregate inflation



(a) AT

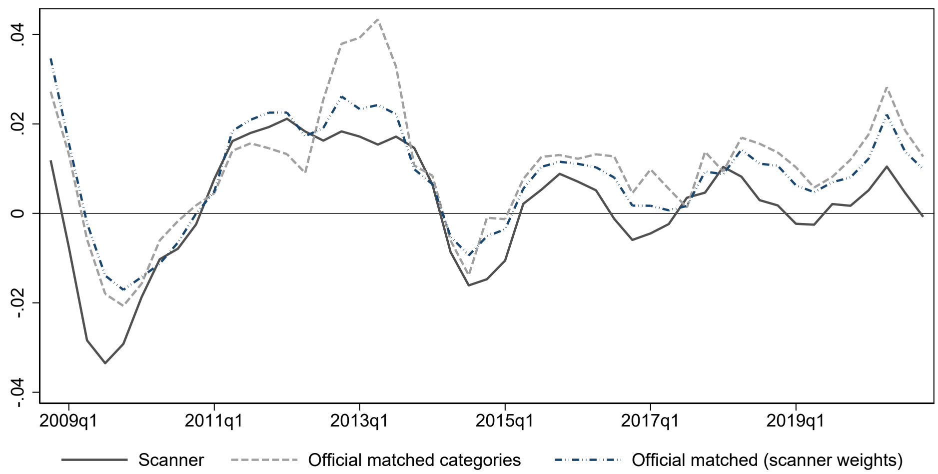


(b) BE

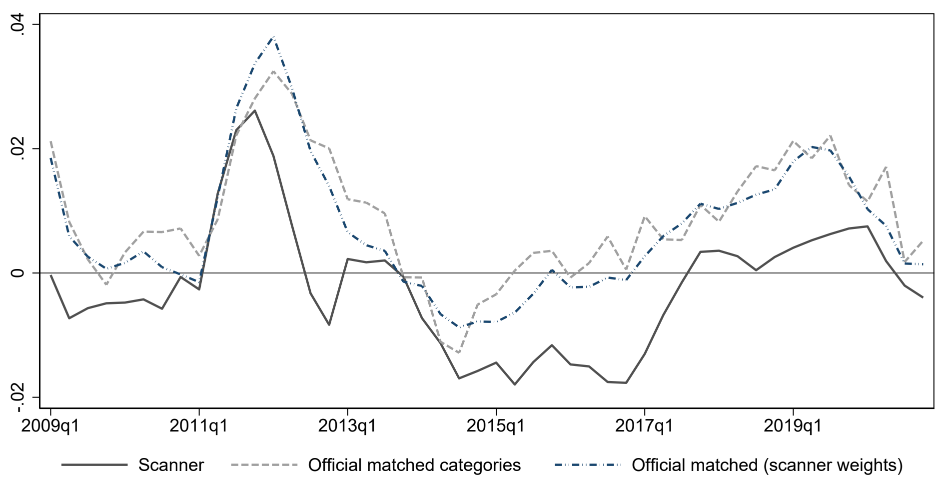


(c) BR

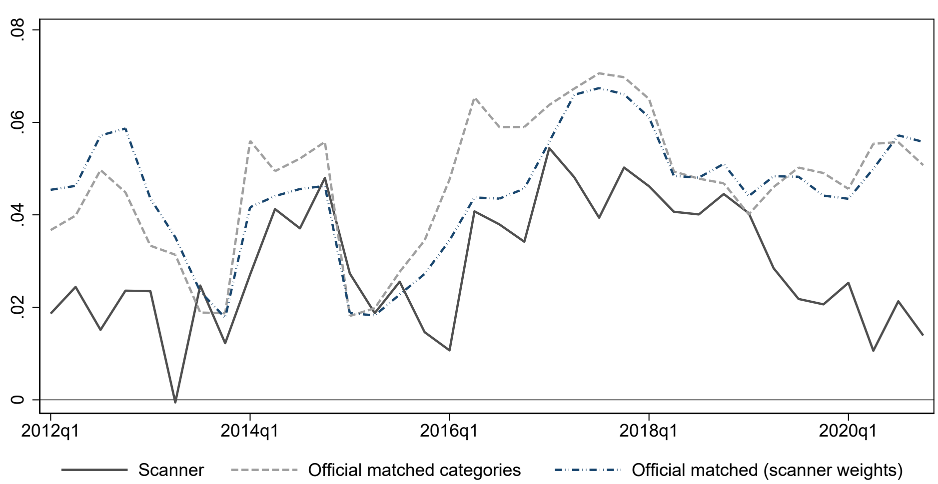
Figure A2: Official vs scanner data aggregate inflation



(a) ES

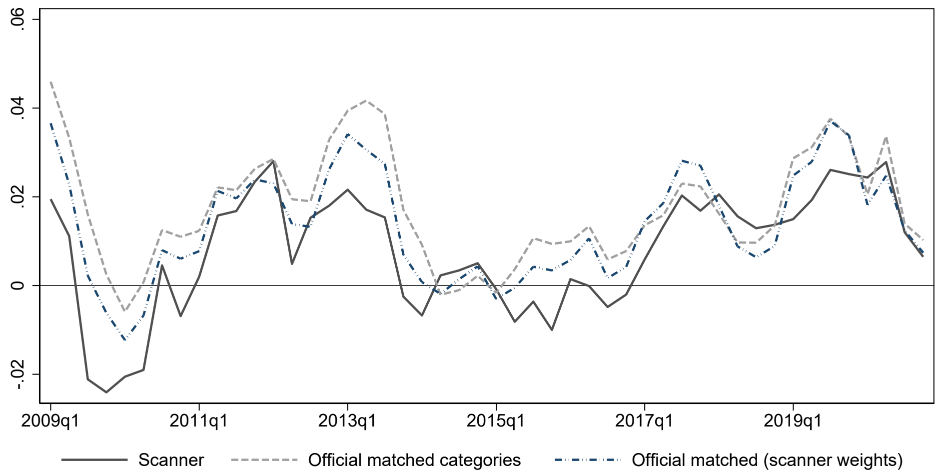


(b) FR

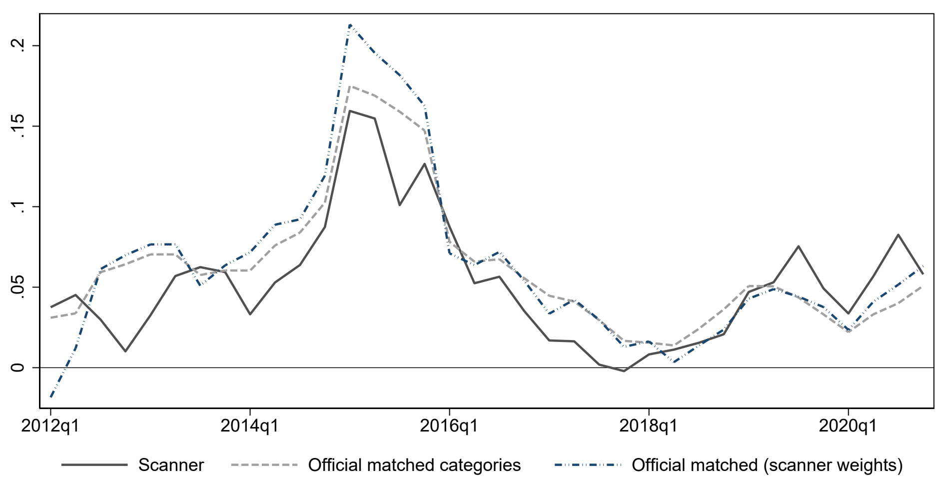


(c) MX

Figure A2: Official vs scanner data aggregate inflation

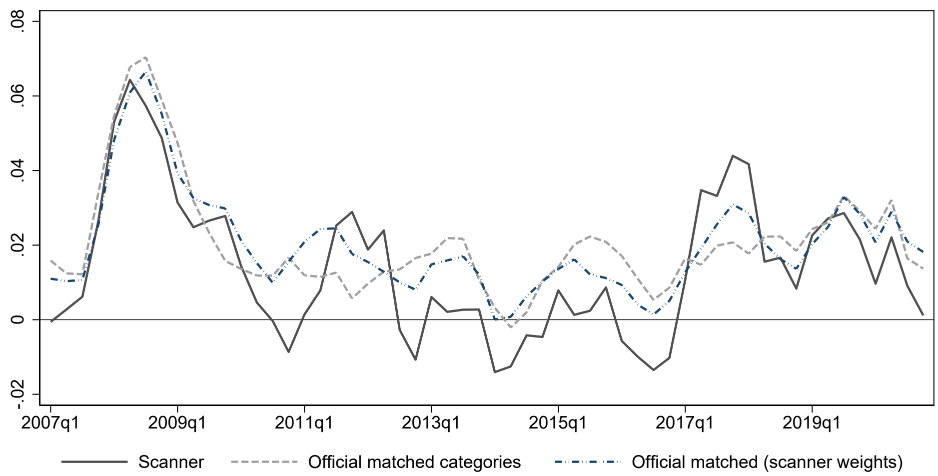


(a) NL

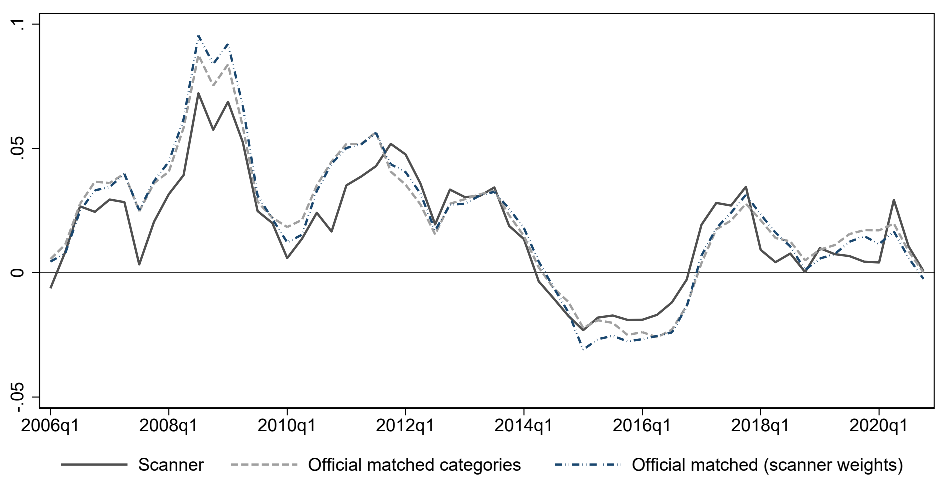


(b) RU

Figure A2: Official vs scanner data aggregate inflation



(a) SE



(b) UK

Table A2: Correlation with official inflation

	Correlation
AR	0.98
AT	0.83
BE	0.83
BR	0.60
DE	0.96
ES	0.87
FR	0.84
MX	0.46
NL	0.86
RU	0.85
SE	0.89
UK	0.94
US	0.93
Average correlation	0.83
Total correlation	0.97

Notes: This table reports the correlation of the aggregate inflation in the scanner data with the official inflation statistics for the same consumption categories. “Total correlation” is computed pooling all countries.

B. Derivation of pricing equation

Motivation for keeping the number of factors low. Following Amiti et al. (2019), we start from the pricing equation of a firm f:[18]

$$p_{ft} = mc_{ft} + M_f(p_{ft}, \boldsymbol{p}_{-ft}; \boldsymbol{\xi}_t)),$$

where p_{ft} is the log price, mc_{ft} are the log marginal costs, and M_f is the log markup function which depends on the own price p_{ft}, the vector of competitors prices $\boldsymbol{p}_{-ft}$ and the vector of demand shocks of all firms $\boldsymbol{\xi}_t$.

Taking the total derivative we get

$$\Delta p_{ft} = \frac{1}{1+\Gamma_{ft}} \Delta mc_{ft} + \frac{\Gamma_{-ft}}{1+\Gamma_{ft}} \Delta p_{-ft} + \underbrace{\frac{1}{1+\Gamma_{ft}} \sum_{j=1}^{N} \frac{\partial M_f(\boldsymbol{p}_{ft}; \boldsymbol{\xi}_t)}{\partial \xi_{jt}} \Delta \xi_{jt}}_{\varepsilon_{ft} \text{ effective demand shock}}, \tag{B.1}$$

where j indexes firm f's competitors, Δp_{-ft} is the Laspeyres price index of the competitors' price changes, $\Gamma_{ft} \equiv -\frac{\partial M_f(p_{ft}, \boldsymbol{p}_{-ft}; \boldsymbol{\xi}_t)}{\partial p_{ft}}$ and $\Gamma_{-ft} \equiv \sum_{j \neq f} \frac{\partial M_f(p_{ft}, \boldsymbol{p}_{-ft}; \boldsymbol{\xi}_t)}{\partial p_{jt}}$.

Using the assumptions in Amiti et al. (2019), $\Delta p_{-ft} = \sum_{j \neq f} \frac{S_{jt}}{1-S_{ft}} \Delta p_{jt}$ and $\Gamma_{ft} = \Gamma_{-ft}$, aggregating to Δp_t, replacing Δp_{-ft}, and solving yields

$$\Delta p_{ft} = \frac{1}{1+\tilde{\Gamma}_{ft}} \Delta mc_{ft} + \frac{\tilde{\Gamma}_{-ft}}{1+\tilde{\Gamma}_{ft}} \frac{1}{\sum_{j=1}^{N} \frac{S_{jt}}{1+\tilde{\Gamma}_{jt}}} \sum_{j=1}^{N} \left[\frac{S_{jt}}{1+\tilde{\Gamma}_{jt}} \Delta mc_{jt} + S_{jt} \tilde{\varepsilon}_{jt} \right] + \tilde{\varepsilon}_{ft} \tag{B.2}$$

with $\tilde{\Gamma}_{ft} \equiv \frac{\Gamma_{ft}}{1-S_{ft}}$ and $\tilde{\varepsilon}_{jt} \equiv \frac{1}{1+\frac{S_{jt}\tilde{\Gamma}_{jt}}{1+\tilde{\Gamma}_{jt}}} \varepsilon_{jt}$.

Note that under Cournot competition and nested CES demand, with between- and within-industry elasticities of substitution ρ and η, the elasticities are:

$$\Gamma_{ft} = \Gamma_{-ft} = \frac{(\rho-1)S_{ft}}{1 + \frac{\rho(\eta-1)}{(\rho-\eta)(1-S_{ft})}}. \tag{B.3}$$

Small firms ($S_{it} \to 0$) only react to own marginal costs while bigger firms also react strongly to competitors' shocks and less to own costs.

Assuming $\Delta mc_{ft} = \delta_t + \lambda_f \eta_t + \delta_{ft}$ (marginal costs have a common component δ_t, differential senstiivity to aggregate shocks $\lambda_f \eta_t$, and an own idiosyncratic shock δ_{it}), we can rewrite the pricing

[18] This derivation follows closely Appendix C in Amiti et al. (2019), which can be consulted for further details.

equation as:

$$
\begin{aligned}
\Delta p_{ft} &= \delta_t + \left\{ \frac{1}{1+\tilde{\Gamma}_{it}} \lambda_f + \frac{\tilde{\Gamma}_{-ft}}{1+\tilde{\Gamma}_{ft}} \frac{1}{\sum_{j=1}^{N} \frac{S_{jt}}{1+\tilde{\Gamma}_{jt}}} \sum_{j=1}^{N} \left[\frac{S_{jt}}{1+\tilde{\Gamma}_{jt}} \lambda_j \right] \right\} \eta_t \\
&+ \frac{\tilde{\Gamma}_{-ft}}{1+\tilde{\Gamma}_{ft}} \frac{1}{\sum_{j=1}^{N} \frac{S_{jt}}{1+\tilde{\Gamma}_{jt}}} \underbrace{\sum_{j=1}^{N} \left[\frac{S_{jt}}{1+\tilde{\Gamma}_{jt}} \delta_{jt} \right]}_{\eta_{2,t}} \\
&+ \frac{\tilde{\Gamma}_{-ft}}{1+\tilde{\Gamma}_{ft}} \frac{1}{\sum_{j=1}^{N} \frac{S_{jt}}{1+\tilde{\Gamma}_{jt}}} \underbrace{\sum_{j=1}^{N} \left[S_{jt} \tilde{\varepsilon}_{jt} \right]}_{\eta_{3,t}} + \frac{1}{1+\tilde{\Gamma}_{ft}} \delta_{ft} + \tilde{\varepsilon}_{ft}.
\end{aligned}
$$

In addition to the true latent factor η_t there are two additional "factors" $\eta_{2,t}$ and $\eta_{3,t}$. Then the observed correlation which we try to absorb with more factors could have a firm level idiosyncratic origin, as top firms have a high loading on $\eta_{2,t}$, $\eta_{3,t}$ and a high contribution on $\eta_{2,t}$, $\eta_{3,t}$ at the same time – e.g., the second and third factors will absorb the effect of a Unilever shock on the economy.

C. Price synchronization at the firm level

This section presents empirical results on microeconomic pricing decisions of firms and retailers, that motivate the focus on the firm dimension in the main analysis. More precisely, we document synchronization of price changes within firms and retailers, which is usually larger than the synchronization within categories.

We follow the literature on price-setting by multiproduct firms and estimate a multinomial logit model similar to the one used in Bhattarai and Schoenle (2014). The difference in our paper is that we analyze two competing synchronization forces, retailers and firms. For this reason, we use price changes aggregated at the product-retailer-country-quarter level (p_{ifgsct}). We estimate the following multinomial logit model for each country:

$$Pr(Y_{ifgst} = 1, 0, -1 | X_{ifgst} = \chi) \quad = \quad \phi(\beta X_{ifgst})$$

where Y_{ifgst} is an indicator variable for positive, no, or negative average price adjustment of product i, produced by firm f and sold by retailer s between quarter t and $t-1$. Product i belongs to category g.[19]

The main explanatory variables of interest is the share of same-signed price changes within the firm, the retailer, and the category, excluding the price change of the product i. As additional control variables we include quarter fixed effects, aggregate retail inflation and also add the average price change of products in the same firm, retailer and category as a measure of marginal costs.

Table A3 shows that synchronization of prices at the firm level is substantial and of comparable size if not larger than the synchronization driven by retailers and categories. The table reports the percentage point change in the probability of a positive or negative price change after a one-standard deviation change around the mean share of same signed price changes for each dimension.[20] For example in the US, a one standard deviation change in the fraction of positive price changes of products of the same firm is associated with a 3.88 percentage points higher probability of a positive price change.

[19] The base category of the model is no price change. We weight each product with expenditure weights.

[20] All other dimensions are left at their respective weighed averages with the exception of the quarter fixed effects which are all set equal to 0.25 in order to give each quarter the same importance.

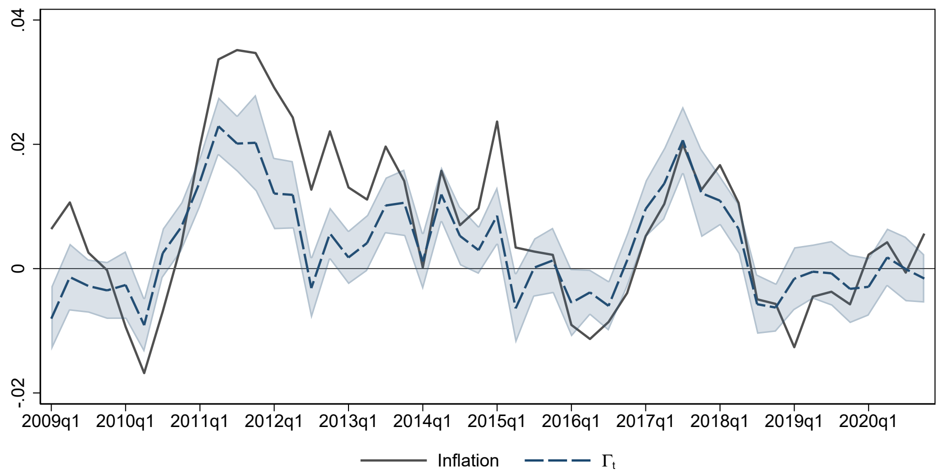
Table A3: Marginal effect of a 1 $Std.Dev.$ on the probability of a Q-o-Q price change

	Positive change			Negative change			
	g	f	s	g	f	s	Obs
AR	5.24	6.52	6.35	0.43	2.74	3.91	926,569
AT	4.42	4.33	3.05	3.84	3.16	3.03	2,685,373
BE	3.65	7.39	3.93	3.84	3.77	4.81	3,572,527
BR	3.19	2.69	2.82	2.55	2.18	3.38	3,345,732
CN	2.29	3.37	4.46	1.94	2.68	4.68	5,789,515
DE	5.46	2.84	0.91	5.94	4.67	0.40	13,003,922
ES	3.85	5.96	3.56	2.76	4.03	5.79	6,484,983
FR	3.28	4.40	4.05	0.37	6.58	4.77	11,510,012
MX	2.55	4.69	2.04	3.64	3.27	2.65	2,811,364
NL	4.20	6.38	0.24	2.64	5.84	1.92	7,433,293
RU	4.35	4.47	5.45	4.60	3.49	3.95	3,959,745
SE	4.79	4.03	1.70	4.85	2.95	1.12	2,285,503
UK	5.28	4.25	2.20	3.27	3.71	1.42	9,741,835
US	3.77	3.88	7.40	2.40	2.44	8.81	45,738,693

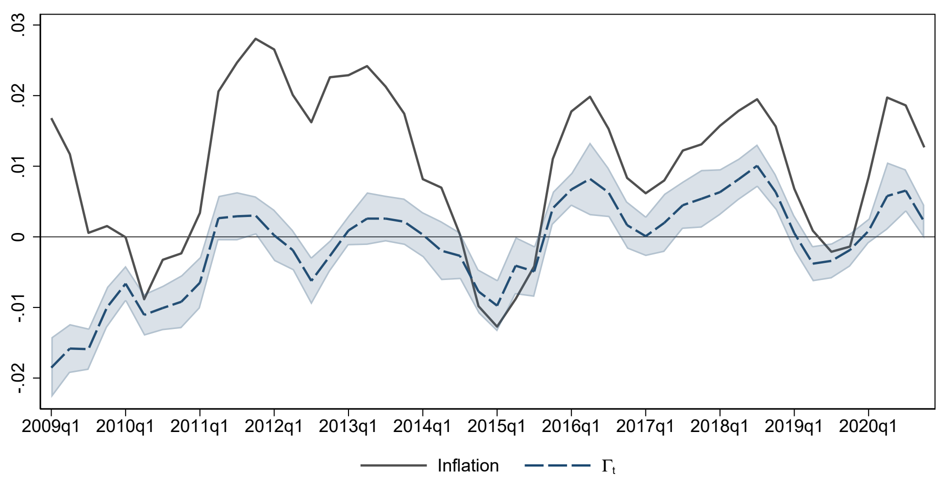
Notes: Columns g, f, s report the change in the probability (in percentage points) of a positive or negative price change after a one-standard deviation change of the share of same-sign price changes around the mean in each dimension. "Obs" reports the number of observations included in the model.

D. Additional figures and tables

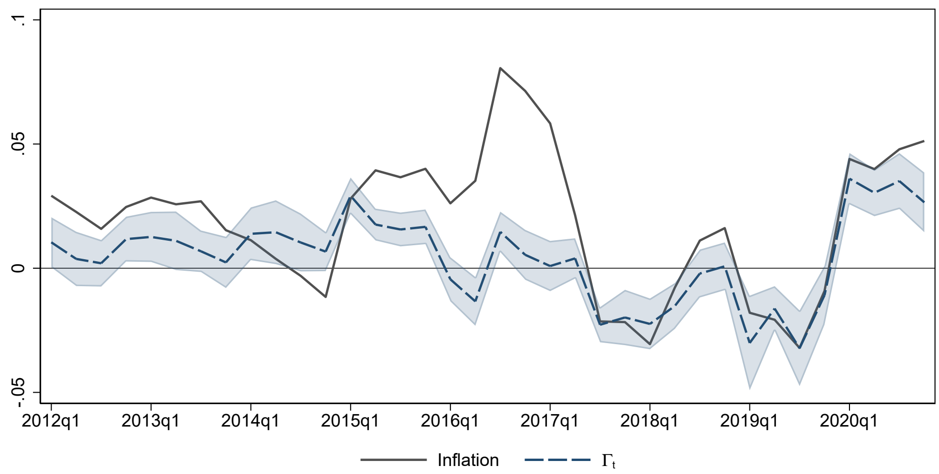
Figure A3: Aggregated retail inflation and the simple granular residual



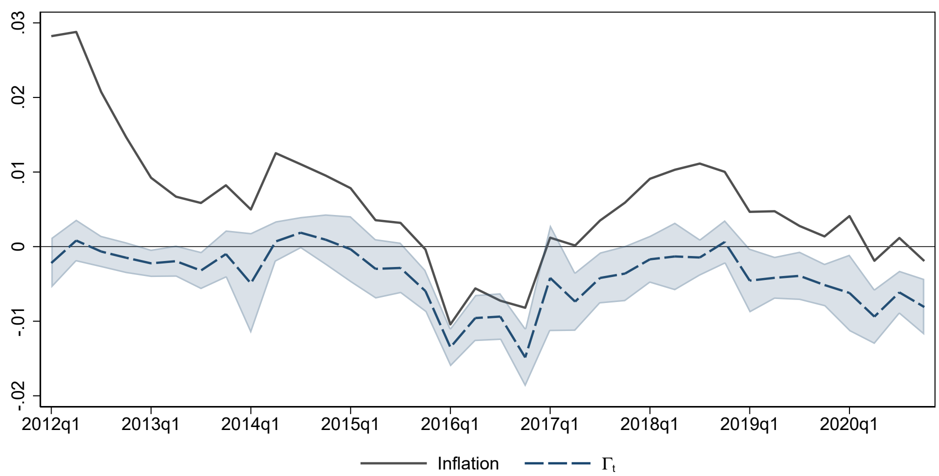
(a) AT



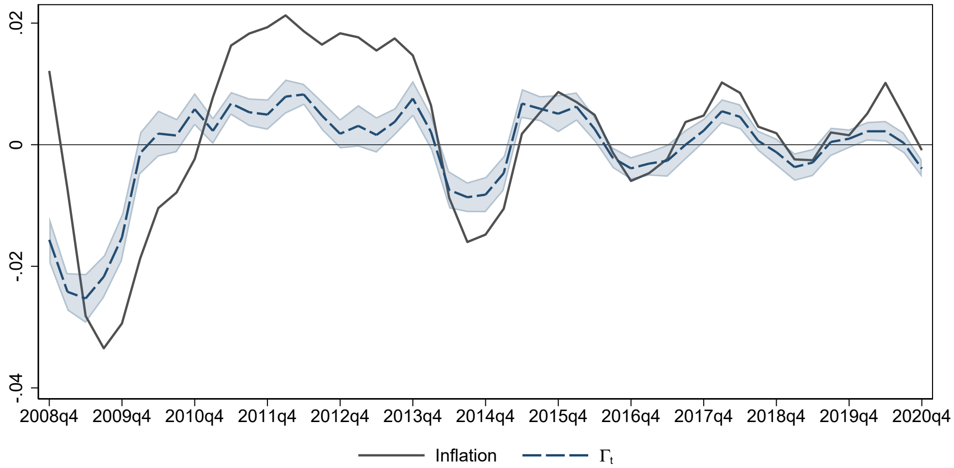
(b) BE


.1
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

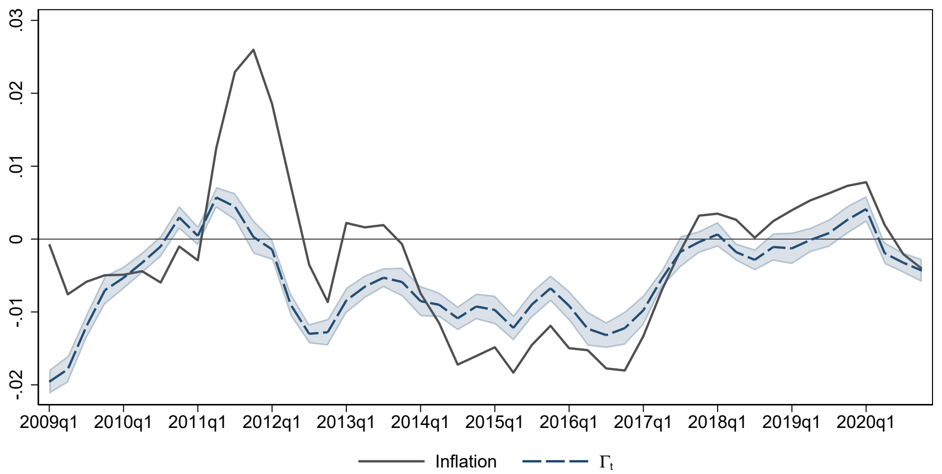
(c) BR


.03
.02
.01
0
-.01
-.02
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

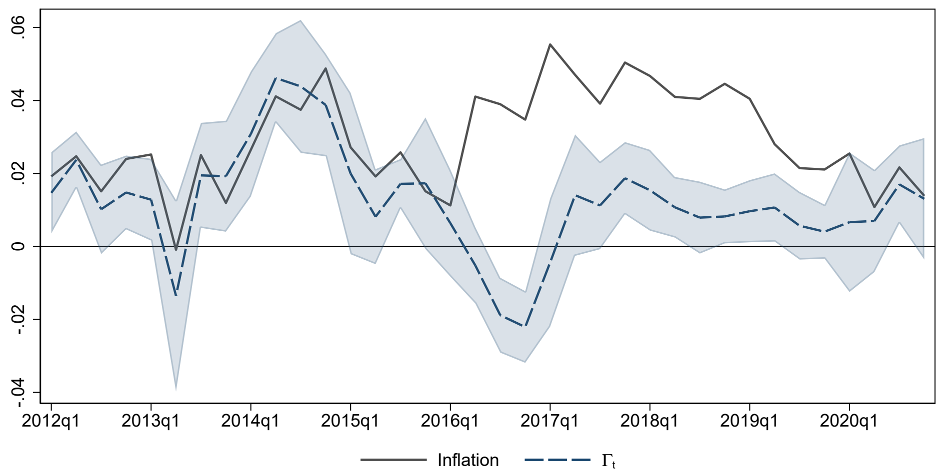
(d) CN


.02
0
-.02
-.04
2008q4
2009q4
2010q4
2011q4
2012q4
2013q4
2014q4
2015q4
2016q4
2017q4
2018q4
2019q4
2020q4
Inflation
Γ_t

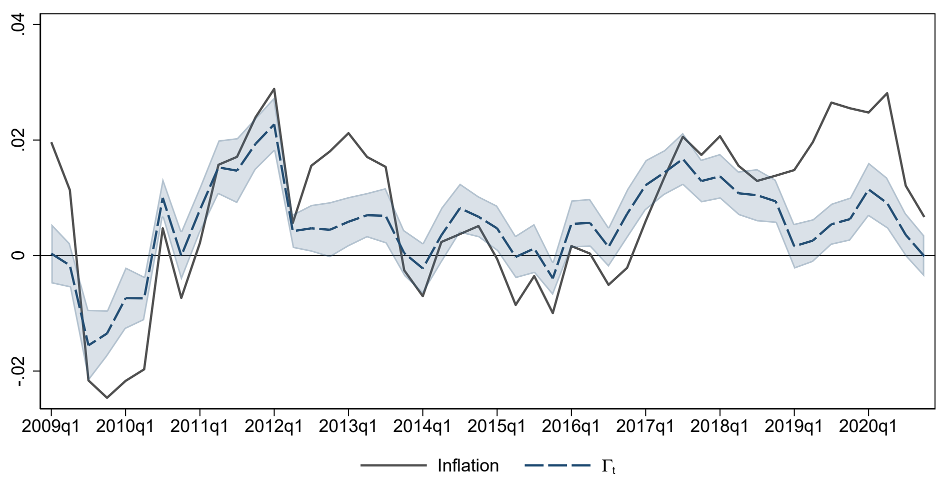
(e) ES


.03
.02
.01
0
-.01
-.02
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

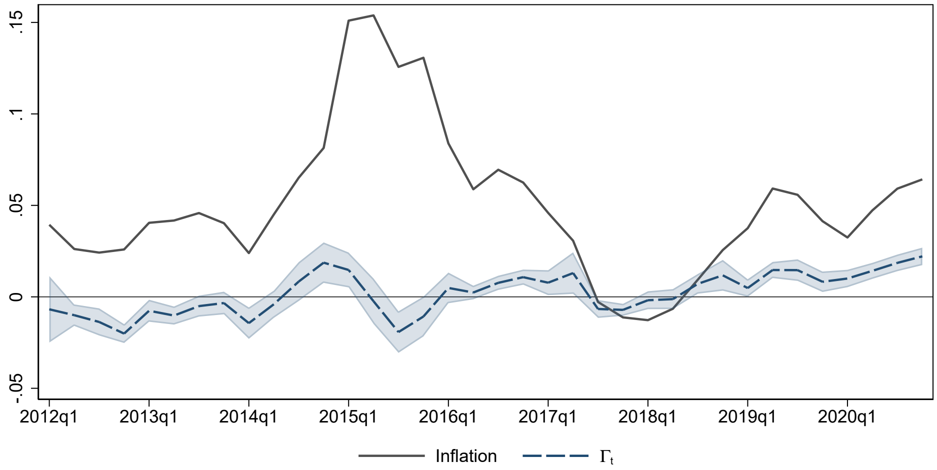
(f) FR


.06
.04
.02
0
-.02
-.04
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

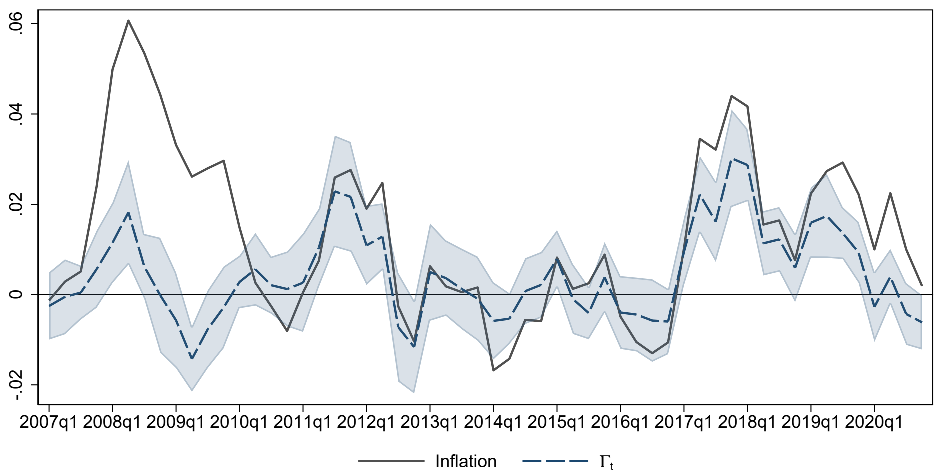
(g) MX


.04
.02
0
-.02
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

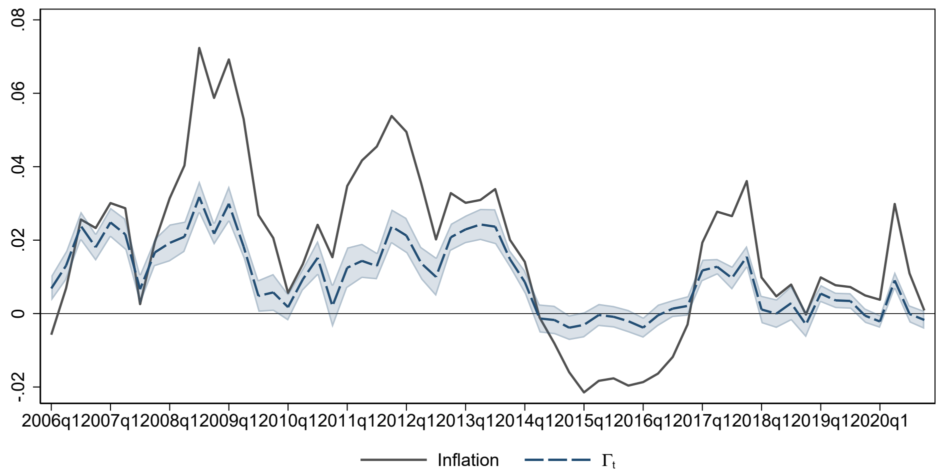
(h) NL


.15
.1
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

(i) RU

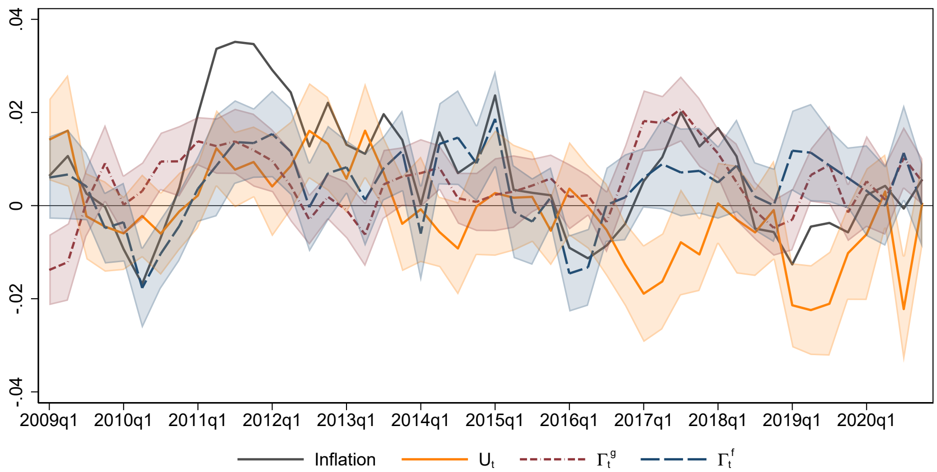

.06
.04
.02
0
-.02
2007q1
2008q1
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

(j) SE

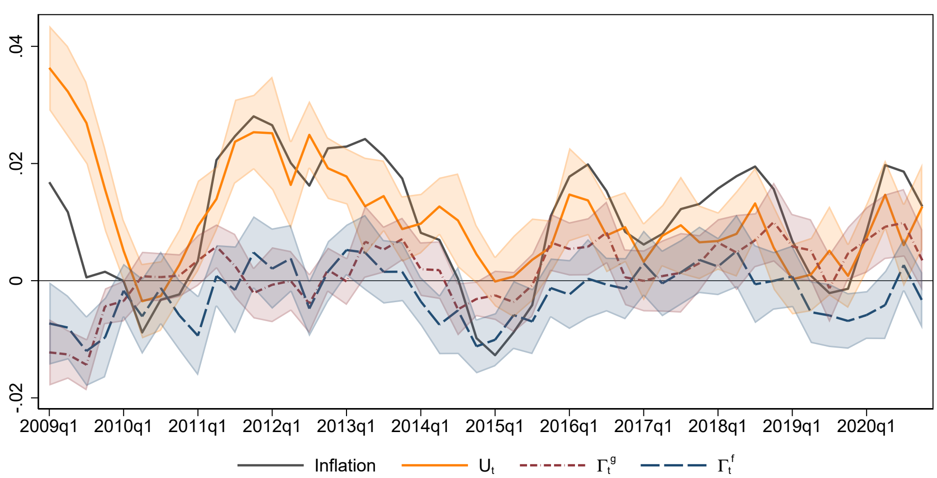

.08
.06
.04
.02
0
-.02
2006q1
2007q1
2008q1
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
Γ_t

(k) UK

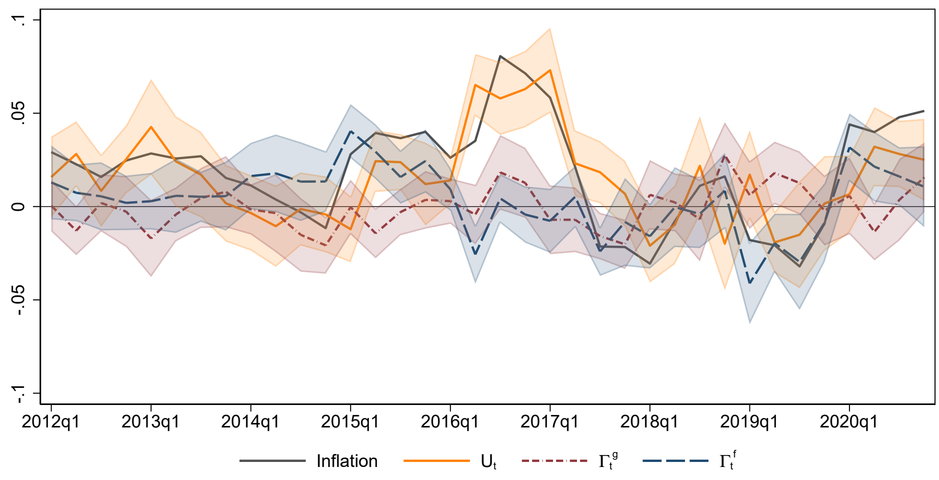
Figure A4: Aggregated retail inflation and granular components



(a) AT

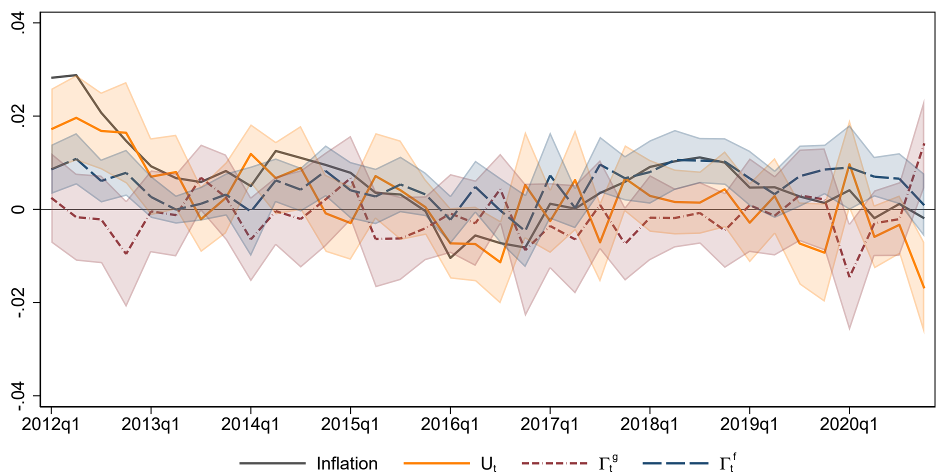


(b) BE



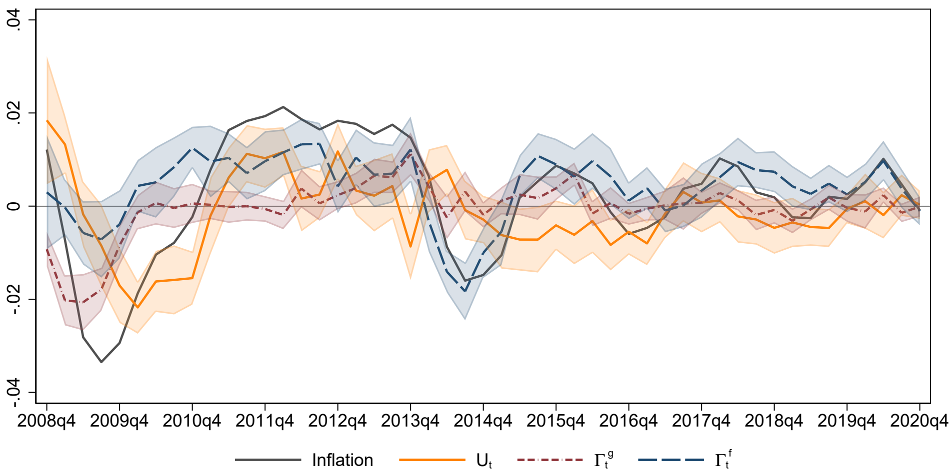
.1
.05
0
-.05
-.1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

(c) BR

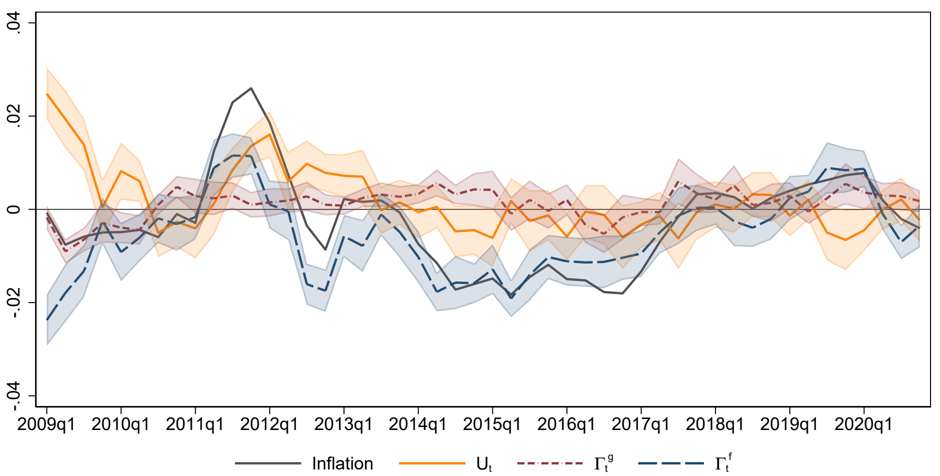


.04
.02
0
-.02
-.04
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

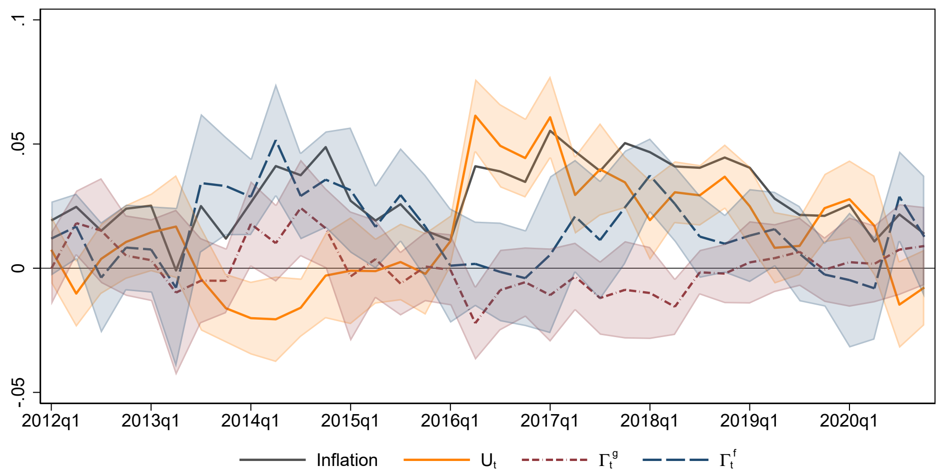
(d) CN


.04
.02
0
-.02
-.04
2008q4
2009q4
2010q4
2011q4
2012q4
2013q4
2014q4
2015q4
2016q4
2017q4
2018q4
2019q4
2020q4
Inflation
U_t
Γ_t^g
Γ_t^f

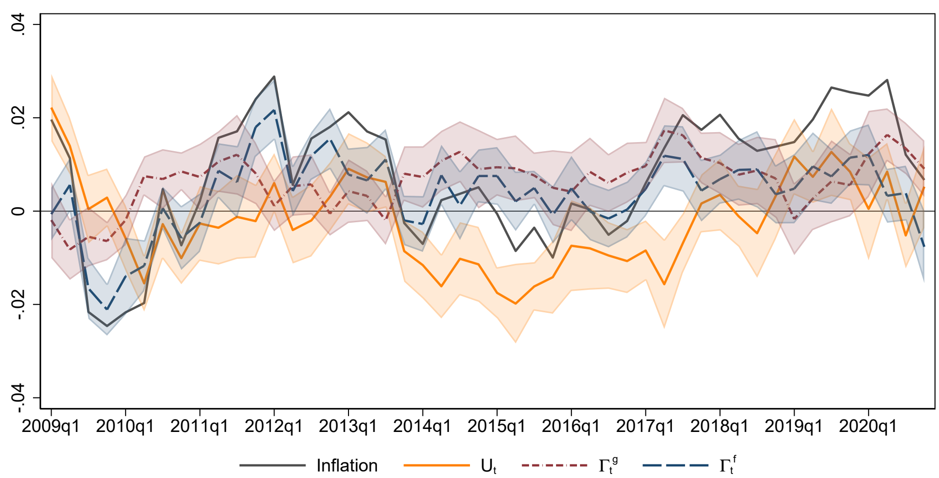
(e) ES


.04
.02
0
-.02
-.04
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

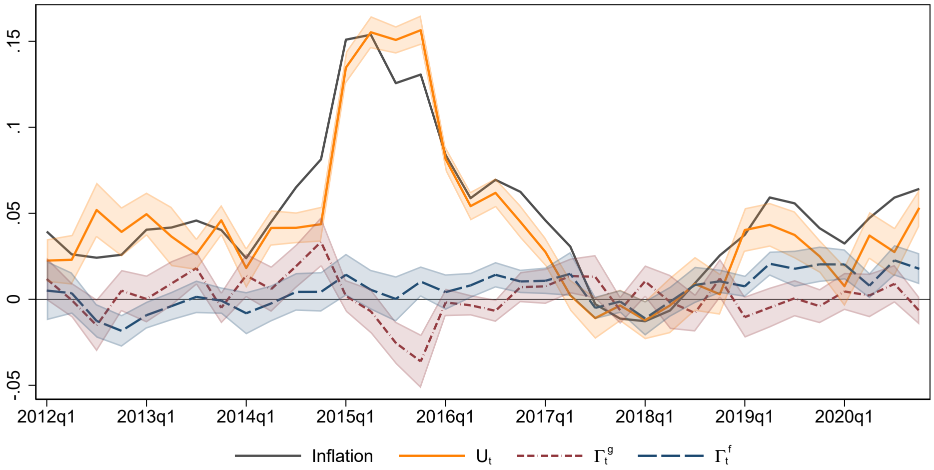
(f) FR


.1
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

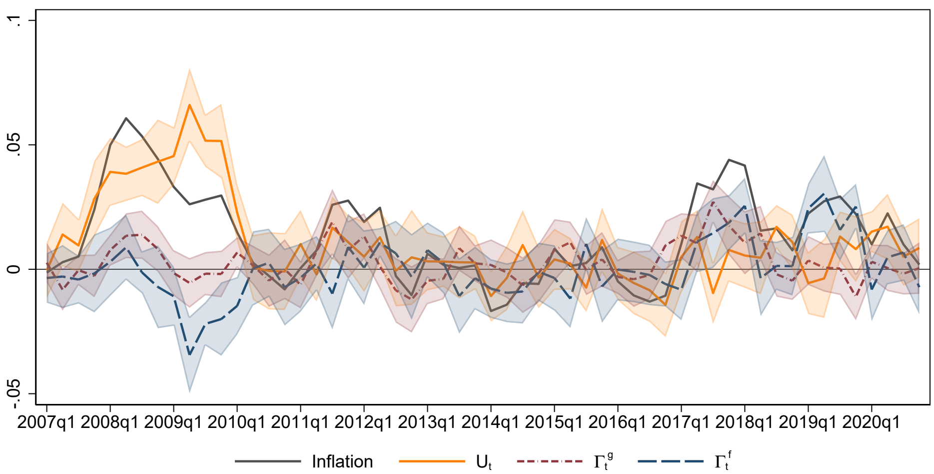
(g) MX


.04
.02
0
-.02
-.04
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

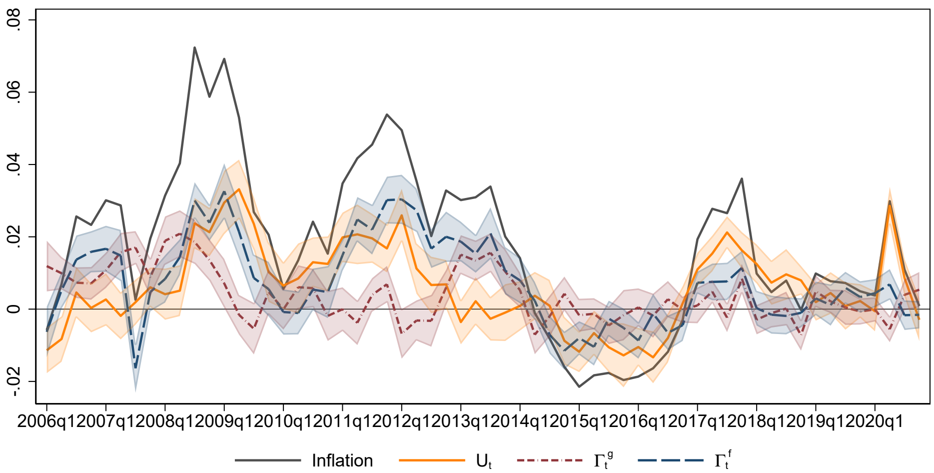
(h) NL


.15
.1
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

(i) RU

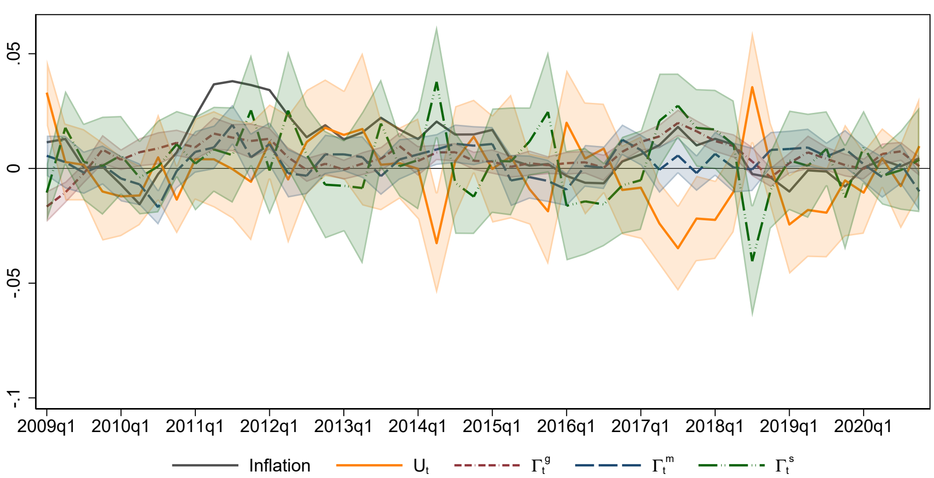

.1
.05
0
-.05
2007q1 2008q1 2009q1 2010q1 2011q1 2012q1 2013q1 2014q1 2015q1 2016q1 2017q1 2018q1 2019q1 2020q1
Inflation
U_t
Γ_t^g
Γ_t^f

(j) SE

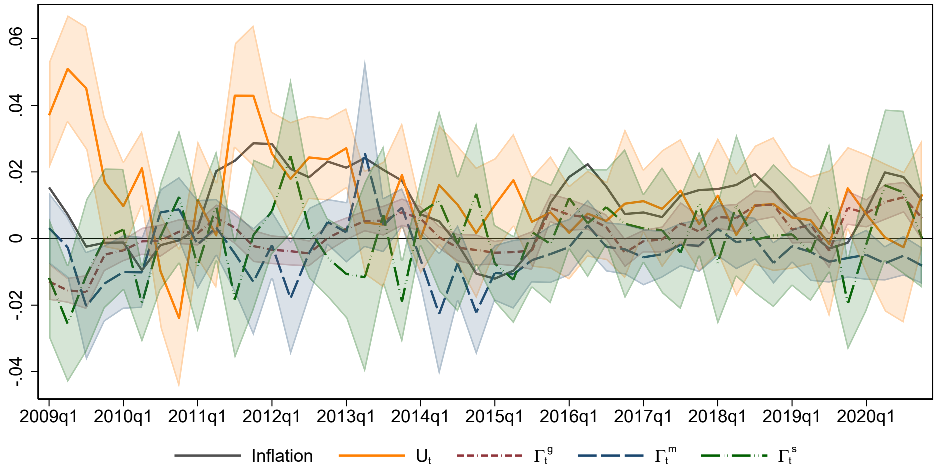


(k) UK

Figure A5: Aggregated retail inflation and granular components, with the retailer dimension

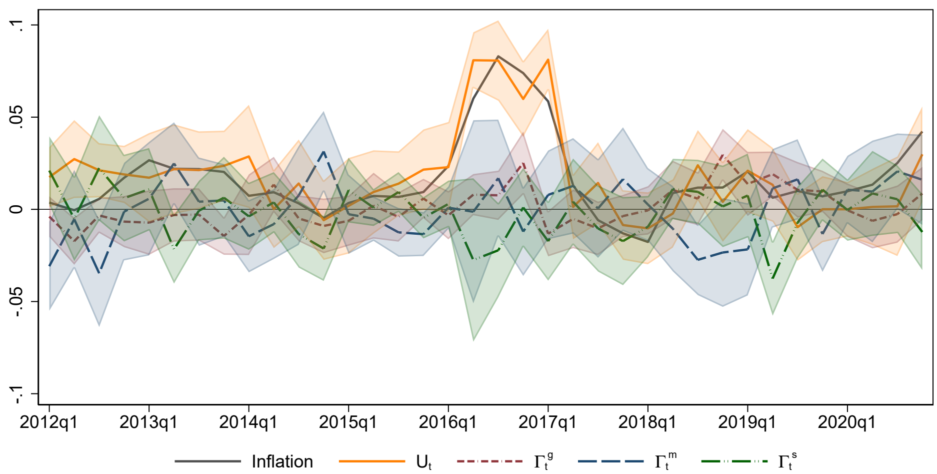


(a) AT



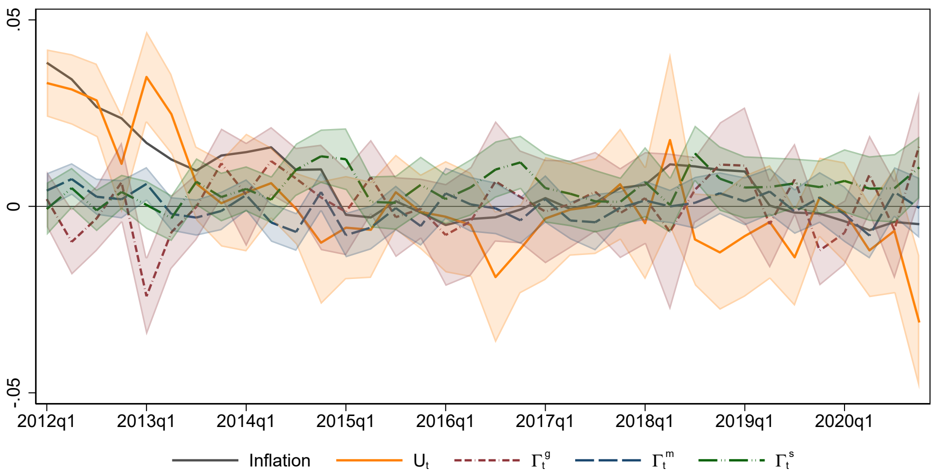
.06
.04
.02
0
-.02
-.04
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(b) BE



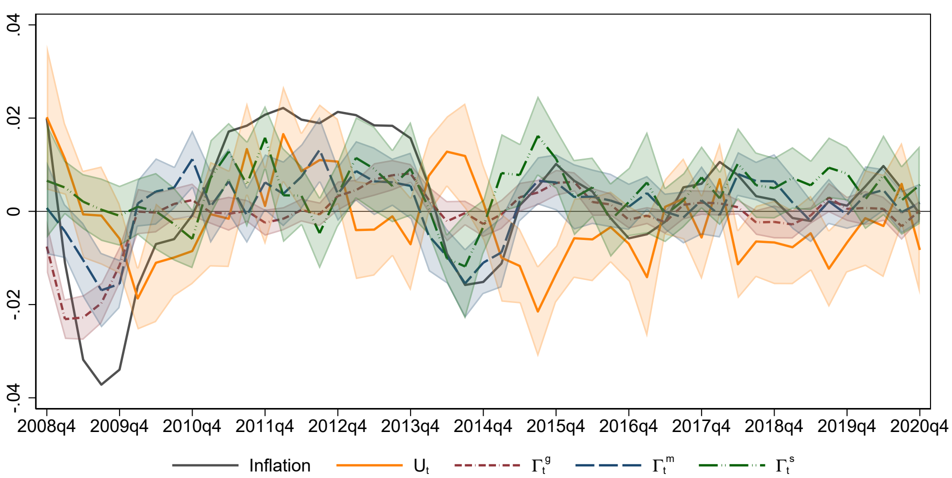
.1
.05
0
-.05
-.1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(c) BR



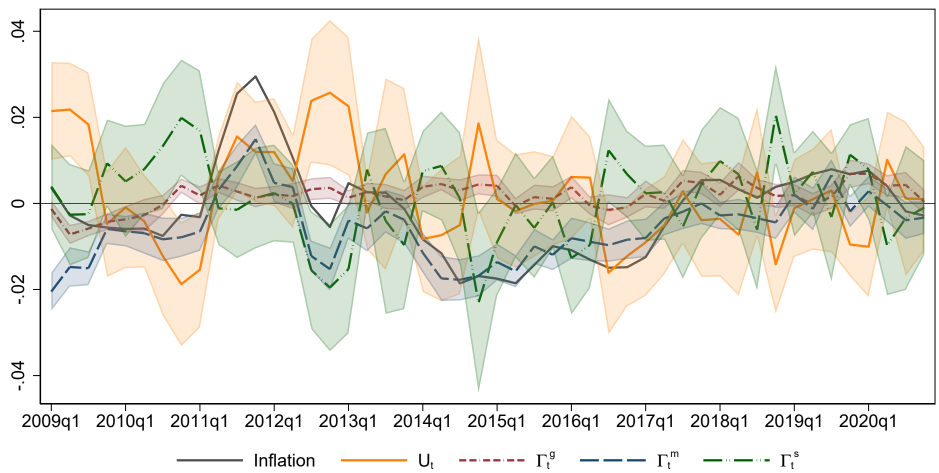
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(d) CN

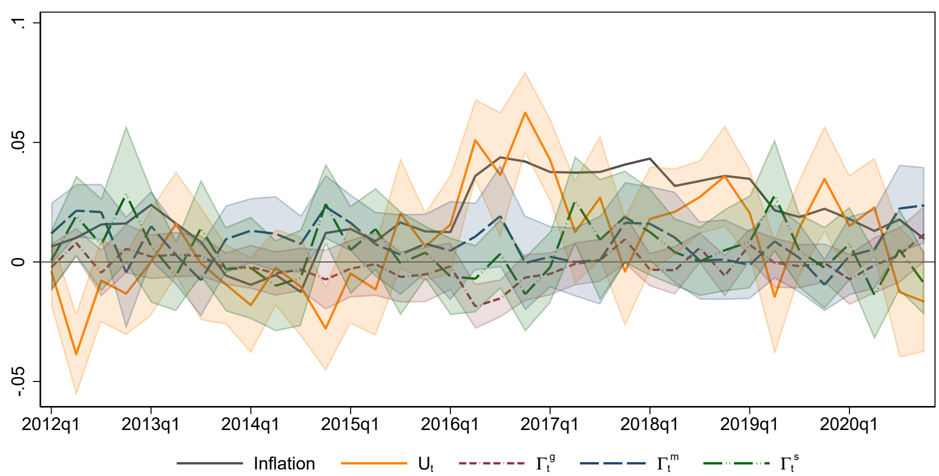


.04
.02
0
-.02
-.04
2008q4
2009q4
2010q4
2011q4
2012q4
2013q4
2014q4
2015q4
2016q4
2017q4
2018q4
2019q4
2020q4
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

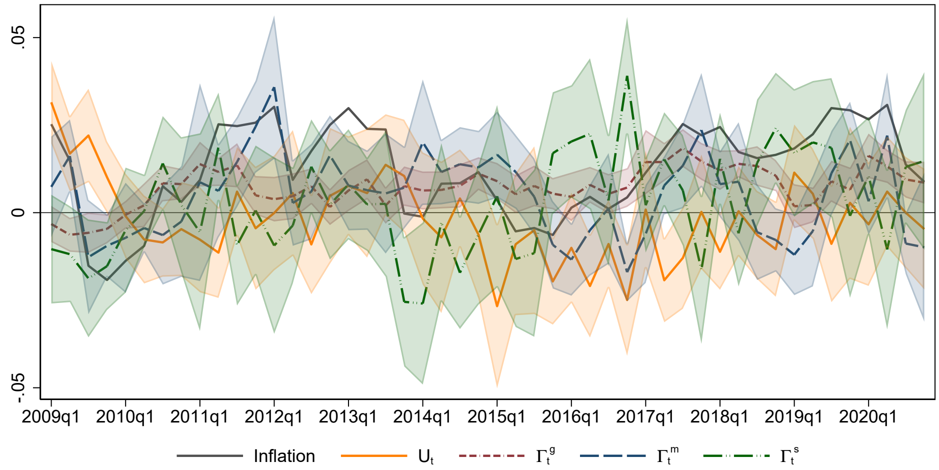
(e) ES


.04
.02
0
-.02
-.04
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

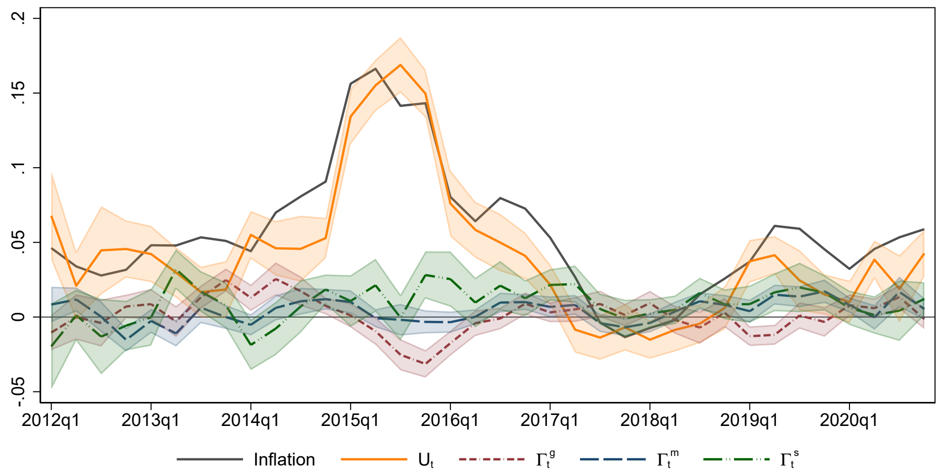
(f) FR


.1
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

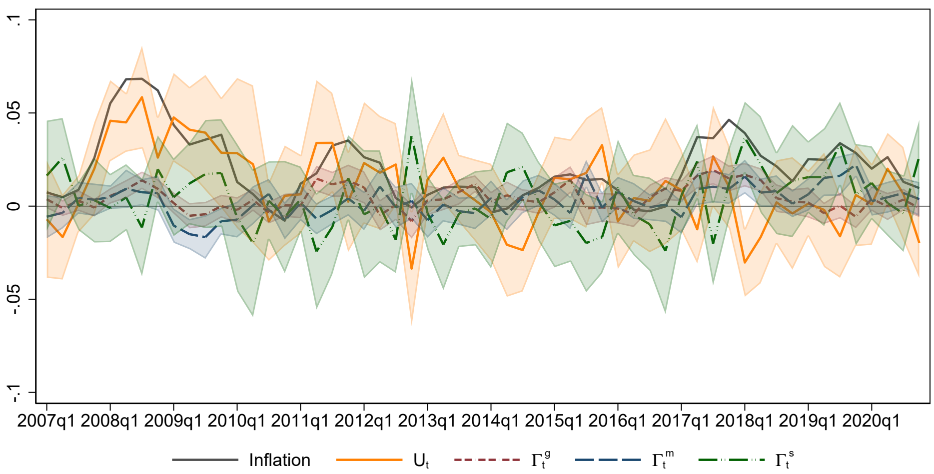
(g) MX


.05
0
-.05
2009q1
2010q1
2011q1
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(h) NL

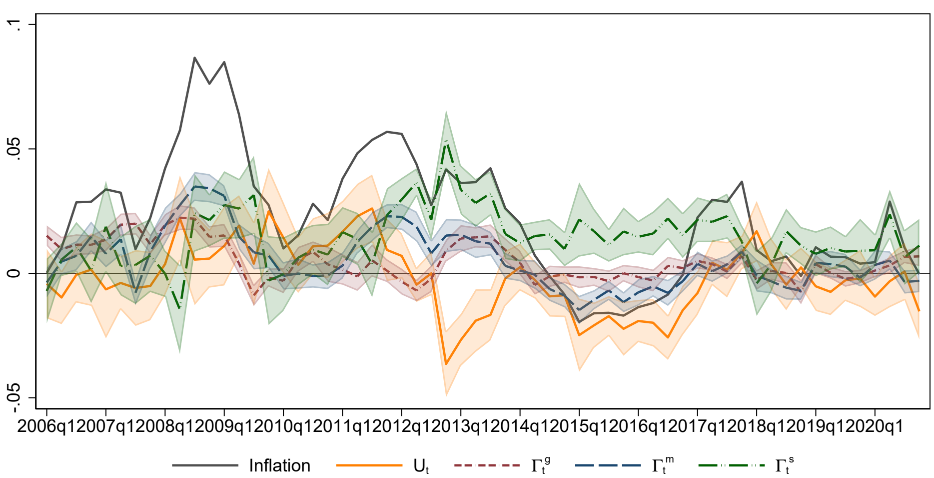

.2
.15
.1
.05
0
-.05
2012q1
2013q1
2014q1
2015q1
2016q1
2017q1
2018q1
2019q1
2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(i) RU



.1
.05
0
-.05
-.1
2007q1 2008q1 2009q1 2010q1 2011q1 2012q1 2013q1 2014q1 2015q1 2016q1 2017q1 2018q1 2019q1 2020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(j) SE



.1
.05
0
-.05
2006q12007q12008q12009q12010q12011q12012q12013q12014q12015q12016q12017q12018q12019q12020q1
Inflation
U_t
Γ_t^g
Γ_t^m
Γ_t^s

(k) UK

Table A4: Summary statistics and correlations of factor components

	A) Basic firm match				B) 2 Factors				C) 3 Factors			
	Mean	St. Dev	Corr	Var(Δp_{ct}) share	Mean	St. Dev	Corr	Var(Δp_{ct}) share	Mean	St. Dev	Corr	Var(Δp_{ct}) share
Advanced Economies (N. Obs = 457)												
Δp_{ct}	0.84	1.63	1.00	1.00	0.84	1.63	1.00	1.00	0.84	1.63	1.00	1.00
U_{ct}	0.38	1.16	0.61	0.43	0.38	1.16	0.61	0.43	0.38	1.16	0.61	0.43
Γ^f_{ct}	0.17	0.95	0.66	0.38	0.17	0.95	0.66	0.38	0.17	0.95	0.66	0.38
$\sum_f w_{fct-4}\delta_{fct}$	0.16	0.91	0.64	0.35	0.17	0.88	0.65	0.34	0.13	0.80	0.61	0.29
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	0.00	0.27	0.15	0.03	-0.00	0.32	0.16	0.04	0.04	0.44	0.31	0.09
$\Gamma^{f\in top10f}_{ct}$	*0.10*	*0.61*	*0.64*	*0.25*	*0.10*	*0.61*	*0.64*	*0.25*	*0.10*	*0.61*	*0.64*	*0.25*
$\Gamma^{f\notin top10f}_{ct}$	*0.07*	*0.50*	*0.47*	*0.13*	*0.07*	*0.50*	*0.47*	*0.13*	*0.07*	*0.50*	*0.47*	*0.13*
Γ^g_{ct}	0.30	0.70	0.43	0.19	0.30	0.70	0.43	0.19	0.30	0.70	0.43	0.19
$\sum_f w_{gct-4}\delta_{gct}$	0.24	0.51	0.22	0.07	0.23	0.48	0.15	0.04	0.22	0.44	0.12	0.03
$\sum_f w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.06	0.46	0.41	0.12	0.07	0.49	0.46	0.15	0.08	0.52	0.46	0.16
Emerging Markets (N. Obs = 180)												
Δp_{ct}	7.37	10.60	1.00	1.00	7.37	10.60	1.00	1.00	7.37	10.60	1.00	1.00
U_{ct}	6.35	9.97	0.99	0.87	6.35	9.97	0.99	0.87	6.35	9.97	0.99	0.87
Γ^f_{ct}	0.99	1.49	0.45	0.11	0.99	1.49	0.45	0.11	0.99	1.49	0.45	0.11
$\sum_f w_{fct-4}\delta_{fct}$	1.00	1.43	0.47	0.11	1.00	1.41	0.45	0.10	0.97	1.30	0.45	0.09
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	-0.01	0.41	0.02	-0.00	-0.00	0.42	0.08	0.01	0.02	0.58	0.14	0.02
$\Gamma^{f\in top10f}_{ct}$	*0.52*	*0.93*	*0.35*	*0.03*	*0.52*	*0.93*	*0.35*	*0.03*	*0.52*	*0.93*	*0.35*	*0.03*
$\Gamma^{f\notin top10f}_{ct}$	*0.48*	*0.80*	*0.43*	*0.08*	*0.48*	*0.80*	*0.43*	*0.08*	*0.48*	*0.80*	*0.43*	*0.08*
Γ^g_{ct}	0.02	1.05	0.09	0.02	0.02	1.05	0.09	0.02	0.02	1.05	0.09	0.02
$\sum_f w_{gct-4}\delta_{gct}$	0.00	0.99	0.08	0.02	0.03	0.82	0.13	0.02	0.04	0.79	0.17	0.02
$\sum_f w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.02	0.33	0.05	-0.01	-0.01	0.62	-0.02	-0.00	-0.01	0.69	-0.06	0.00

Notes: "Mean" denotes the average inflation rate, "St. Dev." the standard deviation, "Corr" the correlation between the component in the row and actual sample inflation, and "Var(Δp_{ct}) share" denotes the share of the variance of actual inflation accounted for by each component, as measured by the Shapley values. The top panel reports the results computed pooling advanced economies (Austria, Belgium, Germany, Spain, France, the Netherlands, Sweden, UK, US) and the bottom panel emerging markets (Argentina, Brazil, China, Mexico and Russia). Panel A) displays the results using the baseline estimation on a sample using a simpler methodology for matching firms (see Appendix A). Panels B) and C) use the baseline firm matching, but include 2 or 3 factors respectively in the EM PCA.

Table A5: Summary statistics and correlations of factor components

	A) Regional unidentified retailer				B) Firm and category components only			
	Mean	St. Dev	Corr	Var(Δp^r_{ct}) share	Mean	St. Dev	Corr	Var(Δp^r_{ct}) share
Advanced Economies (N. Obs = 457)								
Δp_{ct}	1.09	1.76	1.00	1.00	1.10	1.78	1.00	1.00
U_{ct}	0.15	1.49	0.20	0.24	0.54	1.22	0.66	0.45
Γ^f_{ct}	0.07	0.89	0.63	0.32	0.19	0.89	0.70	0.35
$\sum_f w_{fct-4}\delta_{fct}$	0.08	0.85	0.61	0.29	0.19	0.83	0.67	0.30
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	-0.00	0.26	0.15	0.03	-0.00	0.30	0.23	0.05
Γ^g_{ct}	0.36	0.71	0.48	0.20	0.37	0.73	0.48	0.20
$\sum_g w_{gct-4}\delta_{gct}$	0.31	0.53	0.32	0.09	0.32	0.54	0.32	0.09
$\sum_g w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.05	0.46	0.38	0.11	0.05	0.47	0.37	0.11
Γ^s_{ct}	0.51	1.47	0.38	0.24	-	-	-	-
$\sum_s w_{sct-4}\delta_{sct}$	0.47	1.26	0.38	0.21	-	-	-	-
$\sum_s w_{sct-4}\lambda_{sc}\eta^S_{ct}$	0.04	0.67	0.12	0.03	-	-	-	-
Emerging Markets (N. Obs = 180)								
Δp_{ct}	7.30	10.59	1.00	1.00	7.33	10.74	1.00	1.00
U_{ct}	6.23	10.11	0.20	0.92	6.74	10.36	0.66	0.94
Γ^f_{ct}	0.65	1.17	0.63	0.04	0.55	1.11	0.70	0.05
$\sum_f w_{fct-4}\delta_{fct}$	0.65	1.10	0.61	0.04	0.55	1.09	0.67	0.05
$\sum_f w_{fct-4}\lambda_{fc}\eta^F_{ct}$	0.00	0.35	0.15	-0.00	0.01	0.21	0.23	0.00
Γ^g_{ct}	-0.03	0.92	0.48	-0.00	0.03	0.98	0.48	0.00
$\sum_g w_{gct-4}\delta_{gct}$	-0.04	0.80	0.32	0.04	0.05	0.85	0.32	0.02
$\sum_g w_{gct-4}\lambda_{gc}\eta^G_{ct}$	0.01	0.41	0.38	-0.04	-0.01	0.53	0.37	-0.02
Γ^s_{ct}	0.44	0.75	0.38	0.05	-	-	-	-
$\sum_s w_{sct-4}\delta_{sct}$	0.42	0.70	0.38	0.05	-	-	-	-
$\sum_s w_{sct-4}\lambda_{sc}\eta^S_{ct}$	0.02	0.30	0.12	-0.00	-	-	-	-

Notes: "Mean" denotes the average inflation rate, "St. Dev." the standard deviation, "Corr" the correlation between the component in the row and aggregated sample inflation Δp^r_{ct} using the product-retailer level dataset, and "Var(Δp_{ct}) share" denotes the share of the variance of actual inflation accounted for by each component, as measured by the Shapley values. The top panel reports the results computed pooling all advanced economies and the bottom panel all emerging markets. Panel A) keeps unidentified retailers but assigns it to an artificial regional retailer using the household region information. Panel B) only estimates Γ^f_{ct} and Γ^g_{ct} on the baseline product-retailer level sample. Δp^r_{ct} refers to aggregated inflation computed using the retailer-country-quarter level sample, which slightly differs from the aggregated inflation in the baseline sample Δp_{ct}.