Supermarket Price Setting on the Two Sides of the Atlantic - Evidence from Scanner Data*

Peter Karadi† Juergen Amann‡ Javier Sánchez Bachiller§

Pascal Seiler¶ Jesse Wursten‖

This version: September 9, 2022

Abstract

Food-price inflation is more volatile in the US than in the euro area. We utilize a novel supermarket scanner dataset of Germany, Netherlands, France and Italy (EA4), and an equivalent dataset of the US to contrast two contributing factors. First, we document that both the frequency and the size of (sales-filtered) price changes are significantly higher in the US, which indicate a more volatile product-level environment there. Second, we assess the extent of state dependence in price setting. This can be an additional source of price flexibility through the endogenous selection of large price changes (Golosov and Lucas, 2007; Caballero and Engel, 2007). The unparalleled granularity of the data allows us to measure the necessary data-moments directly. We show that price setting is state-dependent, but state dependence raises price flexibility

*All opinions expressed are personal and do not necessarily represent the views of the European Central Bank or the Eurosystem. We would like to thank IRi for making the data available. All estimates and analyzes in this paper, based on data provided by IRi, are by the authors and not by IRi. We thank Ivan Borna Brcina, Matthias Gnewuch and Marta Oliva Riera for excellent research assistance.

†corresponding author, European Central Bank, DG Research, Sonnemannstrasse 20, 60640 Frankfurt am Main, Germany and CEPR, peter.karadi@ecb.europa.eu

‡University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom, Juergen.Amann@nottingham.ac.uk

§jsbachiller.work@gmail.com

¶ETH Zurich, KOF Swiss Economic Institute, Leonhardstrasse 21, 8092 Zurich, seiler@kof.ethz.ch,

‖Faculty of Economics and Business, KU Leuven, Naamsestraat 61 - bus 3535, 3000, Leuven, Belgium, jesse.wursten@kuleuven.be

similarly mildly in both regions. The evidence is well represented by a state-of-the-art price-setting model (Woodford, 2009). We also provide new evidence on the response of supermarket price-setting to the Covid shock in Germany and Italy.

Keywords: food-inflation, state-dependent price setting, generalized hazard, duration hazard, US and euro area comparison, Covid-19

1 Introduction

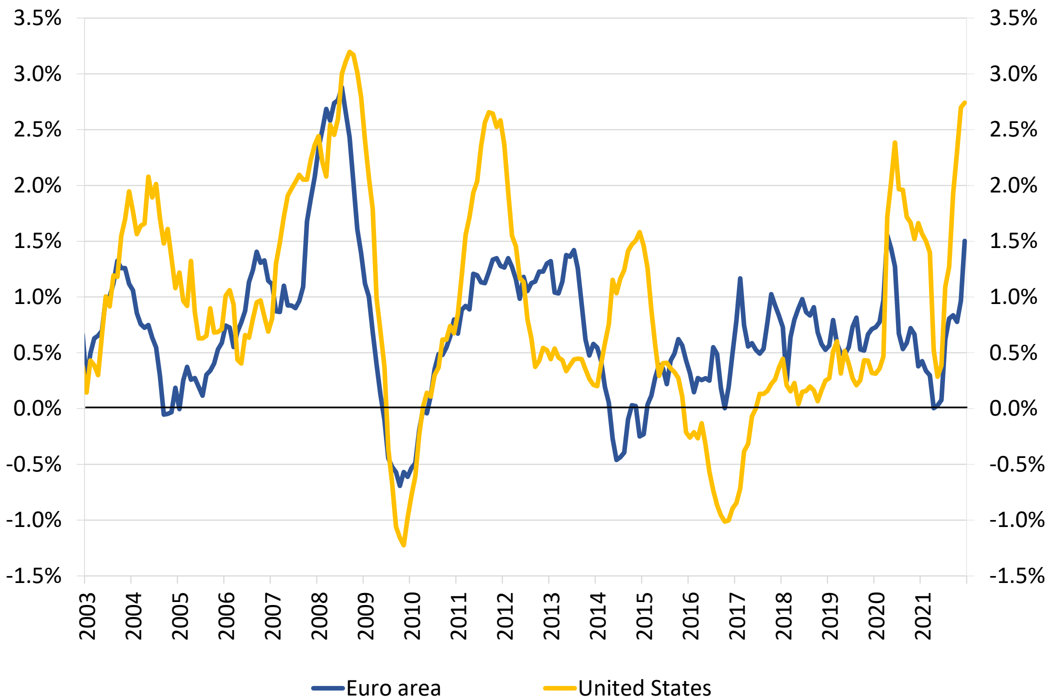
Food inflation is more volatile in the US than in the euro area and, correspondingly, responded more forcefully to the recent Covid-19 pandemic (see Figure 1). Price setting in the food-retail sector has macroeconomic relevance because food consumption accounts for around one-fifth of consumption in both regions, and the salience of grocery prices makes them influence households' aggregate inflation expectations (DAcunto et al., 2021). Previous research has established that price flexibility depends both on the frequency of repricing (*how many* prices change) and the extent of state dependence in price setting (*which* prices change) (Golosov and Lucas, 2007; Caballero and Engel, 2007; Alvarez et al., 2020). In this paper, we use new store-level scanner data from the euro area and a corresponding dataset from the US to carefully measure these two features of supermarket price setting, which can account for the differences in food-inflation volatility.

We find that both the higher frequency and the stronger state dependence of price changes contribute to the higher flexibility of supermarket inflation in the US versus the euro area. We argue that the driving force behind both factors is a more volatile product-level environment in the US. Larger product-level fluctuations both (i) force retailers to adjust prices more frequently and (ii) raise price misalignments, which increase the selection of large price changes. Our conclusions have implications for both model selection and policy.

The paper introduces a novel store-level scanner dataset acquired from the marketing company IRi by the European Central Bank in the context of the Price-setting Microdata Analysis Network (PRISMA). The dataset covers Germany, Netherlands, France, and Italy (EA4), over five years between 2013-2017.[1] The dataset records weekly prices of over 2 million products in over 37 thousand stores in a spatially representative sample. We contrast it to evidence obtained from the US IRi Academic Dataset, an analogous weekly panel over the period 2001-2012 of over 200 thousand products in over 3000 stores covering the 50 most

[1]For the analysis of the Covid shock, we use an auxiliary dataset, which covers the period between mid-February to mid-May in 2019 and 2020 in Germany and Italy for a subset of the stores. For details, see Section 7.

Figure 1: Food and non-alcoholic beverage inflation in the US and euro area, COICOP 01, harmonized prices, year-on-year



Source: OECD

Notes: The figure shows the evolution of year-on-year food and non-alcoholic beverage inflation in the US and euro area between 2003-2021. The series show clear co-movement over most of the period (correlation: 59%), and the US inflation shows higher volatility than euro area inflation (standard deviations: US: 0.95%, EA4: 0.64%).

relevant US markets.

We use the datasets to characterize key features of price setting in the US and the euro area. First, we contrast standard moments about the repricing frequency and the size-distribution of price changes. In our baseline analysis, we filter out temporary sales (Kehoe and Midrigan, 2015; Eichenbaum et al., 2014), which account for the majority of price changes, but contribute only marginally to fluctuations in inflation at regular business cycle frequencies.[2] In line with previous evidence, we find that sales-filtered reference prices change infrequently, and the average absolute size of price changes is large in both regions (Bils and Klenow, 2004; Nakamura and Steinsson, 2008; Gautier et al., 2022a). Previous research has also concluded that one needs price-adjustment frictions and large product-level shocks to account for the low frequency and the large magnitude of price changes (Golosov and Lucas, 2007). We show that both the frequency of price changes (implying price changes once every 7.5 months in the US versus 12 months in the EA4) and the average size of price changes (14% in the US versus 6.7% in EA4) are larger in the US than in the euro area. We conclude that the product-level shocks need to be larger in the US relative to the euro area to account for both the higher frequency and the larger price changes there (see Section 6 for a structural analysis).

Second, we contrast the extent of state dependence in price setting across the two regions. State dependence determines the endogenous selection of large price changes and can raise the volatility of inflation. We use the cross-sectional granularity of the data to generate data moments, which are directly informative about state dependence. In particular, we create a proxy for price-misalignments as the distance of a (log) price from the average price of the same product in those competitor stores that changed their prices in the same month. The average price of price-adjusting stores reveals the optimal reset price in a wide class of models (Calvo, 1983; Dotsey et al., 1999; Golosov and Lucas, 2007). We control for price-level differences among stores caused by differences in amenities and local competitive conditions. To assess the extent of state dependence, we measure both the probability of price adjustment as a function of the misalignment (adjustment hazard) and the density of misalignments following the framework of (Caballero and Engel, 2007). We find that state dependence is higher in the US than in the euro area but raises aggregate price flexibility only mildly in both regions. Notably, the key difference between the extent of state dependence is driven by the more dispersed density of price misalignments, which is strongly influenced by the already established higher volatility of product-level shocks. Our conclusions about the state dependence of price changes are supported by additional data-moments. Specifically, the kurtosis of standardized price changes, which decreases with higher state dependence in a wide class of models (Alvarez et al., 2020), is low in both regions and lower in the US than

[2] In Section 7, we show that after *large* aggregate shocks, like the Covid shock in Italy and Germany, supermarkets do actively adjust the frequency and the size of sales with a sizable impact on inflation.

in the euro area. Furthermore, the duration-hazard of reference price changes is increasing in both regions in line with state dependence in price setting, after we suitably control for unobserved heterogeneity.

Next, we conduct a structural analysis of the price-setting moments, which confirms that higher product-level volatility is one of the key underlying causes of differences in price-setting and food-inflation volatility across the two regions. We use the state-of-the-art state-dependent price-setting model of Woodford (2009) to estimate the underlying structural parameters affecting price setting: (i) the magnitude of price-adjustment (menu) costs, (ii) the standard-deviation of idiosyncratic shocks, and (iii) the magnitude of information-acquisition costs, which, in the model, determines the level of state dependence between the time-dependent Calvo (1983) model and the fixed-menu-cost Golosov and Lucas (2007) model as the two extreme special cases. The most notable difference between the US and euro area is the higher volatility of idiosyncratic shocks in the US, while both the price-adjustment and the information-acquisition costs are quite similar in the two regions.

Finally, we provide evidence on the price-setting response to large shocks by assessing supermarket prices in Germany and Italy during the (first wave of the) Covid-19 pandemic. The shock raised supermarket demand (by restricting access to food-away-from-home) with limited impact on costs (supermarkets were essential sectors sheltered from the lockdowns). We show that supermarkets raised prices both by reducing temporary discounts and raising their reference prices. The inflation response was higher in Italy, where supermarkets' prices are structurally more flexible than in Germany.

Related literature: The paper is related to different strands of the literature. We contribute to the literature that compares price setting in the euro area and the US by introducing a new supermarket scanner dataset and contrasting key price-setting moments, like the frequency and the size of price changes. Dhyne et al. (2006) and, more recently, Gautier et al. (2022a) compare price-setting in the two regions using microdata underlying the Consumer Price Index. They confirm that the frequency and the size of (sales-filtered) price changes are larger in the US not only in the processed food sector, as in our sample, but also in the whole economy, albeit at a somewhat smaller degree.

We contribute to the estimation of the extent of state-dependence in price setting. We calculate moments that are directly informative about state dependence, like the generalized- and the duration hazard functions, utilizing the high granularity of the scanner data. We find that the generalized hazard function, which expresses the probability of price changes as a function of price misalignment, is upward sloping both in the US and in the euro area in line with state dependence. To proxy price misalignments, we use distance from competitors' reset prices (Karadi et al., 2020), which is a valid proxy in a wide range of price-setting models. Our results confirm previous results, which use distance from competitors' prices on more restrictive samples (Gagnon et al., 2012; Campbell and Eden, 2014), and are consistent with complementary estimates using distance from an estimated cost measure (Eichenbaum et al., 2011; Gautier et al., 2022b). We show, furthermore, that the duration hazard, which measures the probability of a price change as a function of the age of the price, is upward sloping in both regions, when we use sales-filtered reference prices and control for unobserved heterogeneity. Upward sloping duration hazard is in line with state-dependent pricing models (see, for example, Dotsey et al., 1999; Nakamura and Steinsson, 2008). Our evidence is different from Nakamura and Steinsson (2008); Klenow and Malin (2010); Campbell and Eden (2014); Alvarez et al. (2021), which find the hazard decreasing, but in line with Fougère et al. (2007), which find it non-decreasing for most disaggregated product-groups.

We assess the implications of our evidence by estimating key structural parameters of a state-of-the-art price setting model (Woodford, 2009) in both regions. The model features fixed (menu) costs of price adjustment (Mankiw, 1985; Caplin and Spulber, 1987), product-level technology shocks (Golosov and Lucas, 2007), and information frictions, which allow it to capture the infrequent and large price adjustments and state-dependence in line with our evidence. As Woodford (2009), Costain and Nakov (2011) and Alvarez et al. (2020), we find that state dependence plays a limited role in raising the flexibility of the price level in both regions. Instead, the key difference contributing to the higher food-inflation volatility in the US is the higher volatility of product-level shocks.

Our paper also contributes to the debate about the role of sales-related price changes as an adjustment margin to aggregate shocks. Previous research has documented conflicting evidence. Anderson et al. (2017), for example, argued that sales are sticky and play an insignificant role as an adjustment margin to aggregate shocks, while Kryvtsov and Vincent (2021) challenged this view and showed that temporary sales do vary over the business cycle. We contribute to this literature by showing that supermarkets in Germany and Italy did respond by adjusting the frequency and the size of their temporary sales to the large demand shock caused by the Covid-19 lockdowns.

The paper is structured as follows. Section 2 describes the data. Section 3 constructs supermarket price indexes and contrasts them with official food-and-beverage subindexes. Section 4 describes conventional moments of price changes in the two regions, including frequency, size, and higher-order dispersion measures. Section 5 presents more complex moments, including the generalized (price-gap) and the duration (price-age) hazard functions, and quantifies the level of state dependence in the two regions. Section 6 conducts structural analysis, and Section 7 contrasts the price-setting response to the Covid shock in Germany and Italy. Section 8 concludes.

2 Data

This section describes key features of the novel euro area dataset and its US counterpart and the data-cleaning steps implemented to improve the informativeness of the data for the analysis of price setting.

2.1 Data coverage

The data covers 4 euro area countries: Germany, Netherlands, France, and Italy between 2013-2017 and the US between 2001-2012.[3] The datasets are weekly panels of total revenues (TR_{psw}) and units sold (Q_{psw}) for each product p in store s in week w. We refer to a product in a store as an item. Unit-value prices of each item are calculated as revenues over units sold ($P^{uv}_{psw} = TR_{psw}/Q_{psw}$). The products are identified with their unique and unmasked barcodes (EANs in the euro area and UPCs in the US).[4] The store IDs are masked to protect the identity of the supermarkets, but they are unique over time, which allows us to track price spells of items over time.

The datasets are representative of the brick-and-mortar sale of participating supermarket chains. The participating chains include regular and discounter supermarkets as well as drug stores.[5] In the euro area countries, our dataset includes 75 percent of the IRi stores.[6]

The euro area datasets are spatially representative in each country. The datasets include the location of the stores up to the first two digits of their ZIP code. The 2-digit ZIP areas partition the countries into around 100 regions (see Table 1). The US dataset covers 50 urban markets across the US. These markets approximately correspond to 50 Metropolitan Statistical Areas (MSA) out of the 384 MSAs in the mainland US in 2010 and cover 73% of the US population.[7]

[3]Even though the US and EA4 datasets do not overlap, this does not hinder the comparison of those features that are stable over time (for example, the frequency of price changes).

[4]The EANs of *private-label* products are masked to protect the identity of the supermarket chain.

[5]The datasets exclude 'hard' discounters like Lidl, Aldi or Walmart.

[6]In some countries (Germany and Italy), additionally, some supermarket chains only share a representative sample of their stores with IRi (i.e., not the census of stores, which IRi obtains for all participating supermarket chains in France and the Netherlands). We 'upweight' sample-stores using projection weights created using information about the population of stores by geographic unit and store type (e.g., large-, small supermarket, discounter, drug store), which is also part of the dataset.

[7]Therefore, even though the US sample is not spatially representative, it covers the most populous areas providing a relevant sample of supermarkets across urban areas.

Table 1: Data coverage

	DE	FR	IT	NL	US
Time series			2013-2017		2001-2012
# 2-digit ZIPs	95	93	103	91	51
# stores	10412	5851	14700	6559	3280
# store types	4	4	6	2	3
# chains	16	43	466	29	147
% in HICP/CPI	18.5	23.3	23.4	20.7	19.6
# products	410276	426153	776521	391507	204519
# categories	216	311	459	140	31
# subcategories	496	1339	1662	891	109
% private labels	21.07	-	19.31	30.93	10.04
% private labels (exp)	23.95	27.98	20.23	36.3	14.3
% internal use	17.91	0.13	0.23	4.85	0.02
% internal use (exp)	24.02	0.15	0.55	15.23	0.04
av. ann. exp. (bn EUR/USD)	24.09	56.19	31.22	30.01	6.2
# observations (bn)	14.26	11.92	11.3	7.66	2.7

Note: Private-label products in France are aggregated by IRi at product-type level.

2.1.1 Product coverage

The product coverage of the datasets is unsurpassable: they include *all* products sold in each store in the sample.[8] The number of unique products ranges from around 390.000 to 776.000 in the euro area and over 200.000 in the US (see Table 1). The products are identified at the barcode level, and the unique product identification numbers (EANs in the euro area and UPCs in the US) are only masked for private-label goods. The dataset includes a product description as well as detailed product characteristics (e.g., size of packaging).

Products sold in supermarkets include food, alcoholic- and non-alcoholic beverages, personal-care products, as well as goods for household maintenance. Figure 2 contrasts the expenditure distribution within our sample (IRi) across nine main product categories for each country, and the corresponding expenditure share of the same category in the official HICP and CPI

[8] The US sample only includes products within a selected 30 broad product categories: beer, blades, carbonated beverages, cigarettes, coffee, cereal, deodorant, diapers, facial tissue, frankfurters, frozen dinner, frozen pizza, household cleaner, laundry detergent, butter, mayonnaise, milk, mustard&ketchup, peanut butter, paper towels, photography supplies, razors, salty snacks, shampoo, spaghetti sauce, sugar substitutes, toilet tissue, toothbrush, toothpaste, yogurt.

indexes in the euro area and the US, respectively.[9] The expenditure distribution in the IRi samples approximates quite well, though not perfectly, the true consumption pattern of households across the product categories. Bread, meat, fruits, and vegetables, for example, are somewhat underrepresented in the IRi supermarket samples, unsurprisingly as these products are regularly purchased also from specialized stores. The match is less tight in the US sample, which only includes a selected set of product categories (for example, only processed sausages 'Frankfurters' as meat products).

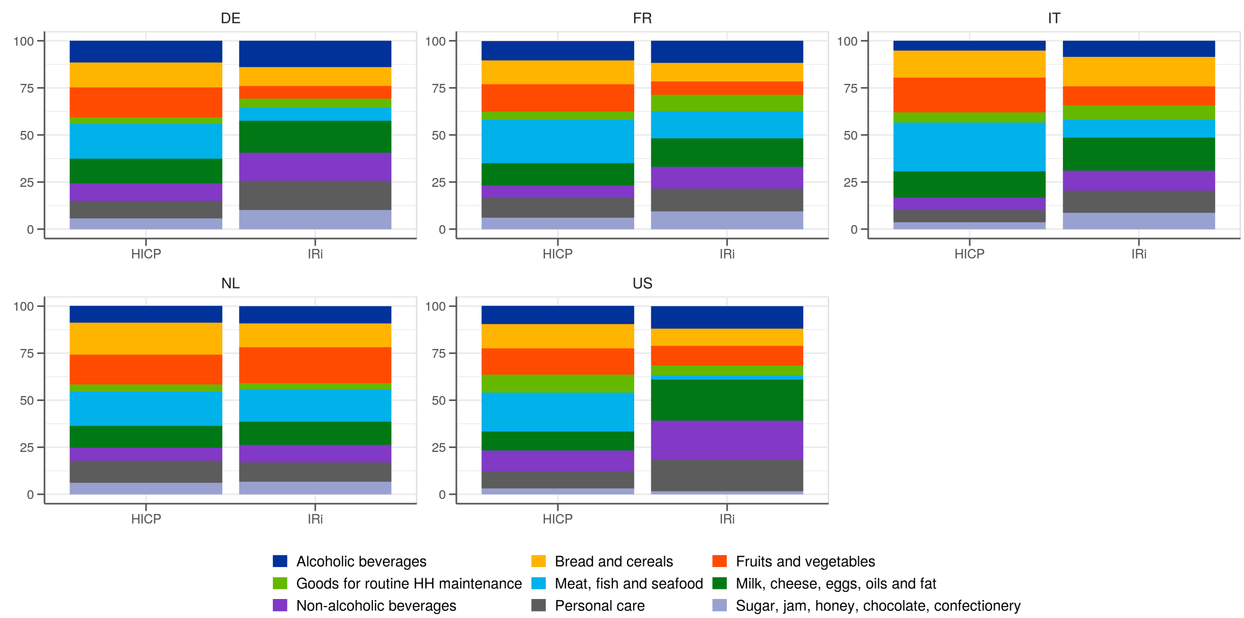


Figure 2: Official vs IRi expenditure shares by category

We conduct the analysis below using a subsample for each country to ease the computational burden. Specifically, we select a 5% random sample of EANs in each EA4 countries, and a 25% random sample of UPCs from the US.[10] The random choice of products ensures that the sample is representative. We include all the stores and time periods in the subsample, where and when the selected products were sold in positive quantities.

[9]The nine categories are constructed to represent large, but still fairly homogeneous groups of products with a sizable share in our sample across all five countries. They are constructed as a suitable combination of 3-digit and 4-digit COICOP categories. We use categories (EA4) and subcategories (US) provided by IRi to allocate products into the nine categories.

[10]The US sample includes fewer products and stores (see Table 1). Choosing a relatively larger subsample makes the number of items in the US sample the same order of magnitude as in the euro area countries).

2.1.2 Store and chain coverage

The datasets cover brick-and-mortar[11] supermarkets and drug stores. The available store types vary across countries. They are the broadest in Germany and Italy, where the sample distinguishes between hypermarkets, supermarkets, drug stores, and discounters.[12] In the US and the Netherlands, the dataset includes supermarkets and drug stores, while in France, there are four groups categorized as large- and small hypermarkets and large- and small supermarkets. The discounters in the sample, where available, refer to soft discounters (like Penny in Germany), but it always excludes hard discounters (like Lidl or Aldi in Europe, and Walmart in the US).

The euro area dataset includes a representative 75 percent subsample of the stores of the IRi sample. In the Dutch, French, and US samples, chains report a full census of their stores to IRi.[13] In the German and Italian samples, in contrast, some supermarket chains only share a representative sample of their stores with IRi. To maintain the representativeness of our sample, we need to adjust the weight of the sample stores suitably.

In particular, we 'upweight' stores (s) by projection weight ν_s. The weights for the stores that appear as census are $\tilde{\nu}_{sw} = 4/3$, which offsets the impact of the dataset being a 75 percent share of the full sample.

To obtain projection weights for sample stores in Germany and Italy, we need to estimate the overall number of stores by store type. This data is part of the IRi dataset.[14] In Italy, the overall number of stores by store type is available annually at the end of each year: here, we use linear projections between end-of-year observations to obtain estimates of the weekly number of stores by store type $\tilde{N}_{Sw}$. In Germany, the number of stores by store type is only available at the end of 2017 (N_{ST}). Here we use the evolution of the stores in the census population in our sample to obtain estimates of the weekly number of stores by store type. First, we calculate average weekly entry (γ_S) and exit rates (δ_S) by census store type (S) over the 2013-2017 sample. Second, the estimated number of stores by store types are obtained by assuming constant entry and exit rates by store type $\tilde{N}_{Sw} = (1 + \gamma_S - \delta_S)\tilde{N}_{Sw-1}$, $\tilde{N}_{ST} = N_{ST}$.

[11] The dataset does not have information about online retail. Online retail is growing, but it is still a small share of overall expenditure.

[12] In Italy, there is a fifth category for self-service stores.

[13] To guard the identity of the stores, store information is only included in our sample if there is a sufficient number of stores (for example, at least three in France) by geographical area and store type. In most cases (in France and the US, for example), store information is withdrawn in these cases. In other cases (in Italy, for example), the geographical granularity becomes coarser (1-digit as opposed to 2-digit ZIP areas).

[14] The number of stores is available by store type and geographical area, but the latter we ignore in the current analysis.

The weights for sample stores are obtained as

$$\tilde{\nu}_{sw} = \frac{4}{3} \frac{\text{non-census population by type in week w}}{4/3 \cdot (\text{non-census sample by type in week w})}, \tag{1}$$

where (non-census population in type in week w) = (estimated aggregate population by type in week w ($\tilde{N}_{Sw}$))-4/3(the number of census stores by type in week w).

Finally, we normalize the projection weights to make sure they sum to unity in each week:

$$\nu_{sw} = \frac{\tilde{\nu}_{sw}}{\sum_s \tilde{\nu}_{sw}}, \text{for each } w. \tag{2}$$

2.2 Data cleaning

Table 2: Data-cleaning moments

	DE	**FR**	**IT**	**NL**	**US**
% same-direction changes	2.15	5.39	8.1	3.58	6.03
% also fractional	1.66	3.71	5.36	1.65	3.31
% fractional price	7.6	8.05	11.66	5.91	6.96
% below closest integer	68.93	53.83	59.48	62.33	58.95
% missing (obs)	43.91	42.09	46.58	42.97	38.49
% missing (exp)	55.37	46.88	59.12	38.9	55.5
% missing (exp >4w)	22.16	21.04	26.27	16.7	13.49

Note: 'Missing (exp)' refers to the expenditure share of products that record zero sales in a single (over four consecutive) week(s).

2.2.1 Posted-price filter

Unit-value prices do not necessarily reflect posted prices. There are two main reasons for this. First, mid-week price changes generate unit values that are in-between actual prices. Second, coupons and other buyer-specific discounts can reduce the average revenue from a product below its posted price. We transform unit prices to estimated posted prices using the following filtering rules.[15]

[15]There is a potential third reason: a within-week temporary price discount. These within-week price changes would be recorded as (smaller) changes in the weekly average price, potentially distorting the price-setting moments at the highest frequencies. As our focus is monthly frequency, we do not expect such changes to influence our conclusions.

First, to reduce the impact of mid-week price changes, we filter out same-direction consecutive price changes. A one-time mid-week permanent price change necessarily generates such same-direction consecutive unit-price changes. The mid-week price change increases the average weekly unit price only partially in the initial week, and pass-through fully only during the second week. Formally, if we observe two consecutive price increases ($I^{+}_{psw,w-1} > 0, I^{+}_{psw+1,w} > 0$) or decreases ($I^{-}_{psw,w-1} > 0, I^{-}_{psw+1,w} > 0$), we conclude that there was a mid-week price change during week w. We set the end-of-the-week posted price during this week as the unit-value price in the following week $P_{psw} = P^{uv}_{psw+1}$. As Table 2 shows, 2-8 percent of the prices are affected by the same-direction filter. Out of these filtered prices, usually over half are fractional (fractions of a cent). As fractional unit values cannot be posted prices, their presence strongly confirms mid-week price changes. Their high share suggests that the filter recovers the true posted prices in most cases. And even though some of the filtered same-direction price changes could have been true adjustments, filtering them out biases our results only marginally, especially at the monthly frequency, which is going to be our focus.

Second, to mitigate the impact of buyer-specific discounts, we round prices upwards to the nearest cent. Posted prices need to be integers in cent units. However, 6-12 percent of unit-value prices are fractional even after controlling for same-direction price changes (see Table 2). As most of the deviations from the posted price result from discounts, we round the prices upwards. Indeed, the closest integer is higher than the price in over 60 percent of the fractional prices. A higher than 50 percent share is expected when the fractional prices are caused by discounts paid by a small fraction of the buyers. In cases when the discounts are paid by such a small fraction of the buyers that the unit-value price deviates from the posted price by at most a cent, our filter picks up the actual posted price. Even when the discounts reduce the average price by more than a single cent, upward rounding brings us closer to the posted price. However, there can still be many cases when the share of buyers paying a discount is large enough to reduce the unit-value price away from the posted price by more than a single cent. In these cases, the filter does not recover the actual posted price. Therefore, we show the robustness of our results below when we exclude fractional prices from the analysis.

The posted price also remains unobserved when there is no sale of the item in a particular week. Zero-sale weeks ($I^{\emptyset}_{psw}$) for existing items[16] are frequent in the data. In particular 38-47 percent of the observations are missing (see Table 2). Furthermore, the expenditure share of items with missing observations is also high, so the issue does not only affect rarely-sold unpopular products with a small expenditure share. In particular, the annual expenditure share of products with at least one missing observation over a year is 40-60 percent. The issue is somewhat less pressing, if we realize that consecutive missing observations are usually

[16] We consider a product p in store s existing in week w if $M_{ps} \leq w \leq T_{ps}$, where M_{ps} is the date of entry (first week when product p was sold in store s), and T_{ps} is the date of exit (the last week it was sold).

short, much shorter than a month. In particular, the annual expenditure share of products with at least one case of 4 or more consecutive missing observations is between 15-25 percent. This is the relevant metric in our analysis, which focuses on monthly price developments: as monthly prices are missing only if weekly prices are missing for 4 weeks consecutively. The presence of a not insignificant fraction of missing prices is still can be considered a caveat of our dataset, and its potential impact needs to be carefully assessed in analyses below.

The dataset requires careful treatment during the rare occasions when the product identifiers stop referring to the same product over time. This happens in the US sample in 2007:01, 2008:01, and 2012:01, when the identifiers of some private-label products get reassigned by IRi. We lack additional information about the rules followed during the reassignment, so, conservatively, we assume that new private-label products replaced old private label products during these three months, and we do not link price spells of private-label products over these months. We treat similarly a subset of German beer- and beer-cocktail products in 2014:01, when their EAN got reassigned to refer to a crate instead of a bottle (which could occasionally generate artificial 24-fold price increases): we treat them as separate products and do not link their prices over the 2014:01 period. Lastly, we drop from the Dutch dataset over the 2013-2014 period a subset of (overwhelmingly fresh) products, which had inconsistent unit treatment resulting in unreliable price development. In particular, we drop products with 'internal use' EANs and 'random weight' volume measurements over the 2013-2014 period. The internal use EANs are assigned by the stores to products packaged internally (e.g., fresh meat). The 'random weight' volume measurement implied a non-standardized unit treatment before 2014, which could have resulted in random unit variation over time if the store changed its reporting. To avoid artificial variation in our data, we drop these products from the analysis. The treatment impacts a small subset of the products (around 12 percent share of annual expenditures) over only two years and only in the Dutch data, so we expect it to have a marginal impact on our analysis.

2.2.2 Time aggregation

We transform the weekly data to a monthly frequency. Our reasons for this are manifold. First, micro price data underlying the official price indices are available at the monthly frequency, so calculating monthly moments facilitates comparison. Second, business cycle fluctuations and inflation dynamics are influenced predominantly by persistent price adjustments, which monthly data will capture. Indeed, the monthly aggregation automatically cleans the data from some of its high-frequency variation unrelated to aggregate fluctuations; therefore, it can improve the efficiency of the analysis. Third, as indicated before, some of the caveats of the data become much less pressing at the monthly frequency, for example, the lack of price observations in weeks with zero sales.

We define monthly posted $i = p$ price of product p in store s as the (highest) mode of the posted prices observed over the weeks of the month:

$$P^i_{psm} = \text{mode}_{w \in m} P^i_{psw}. \tag{3}$$

Using the mode guarantees to choose one of the actual posted prices, so the time aggregation does not introduce artificial prices. This would happen if one instead used the mean or calculated monthly unit prices. Picking the *highest* mode in case of multi-modality tilts the monthly prices towards the (more persistent) reference prices, which tend to be above the sales prices. We calculate monthly reference prices from weekly data analogously.

To calculate expenditure weights, which we detail later, we also need estimates of the monthly expenditures TR_{psm}. We transform weekly expenditures to normalized monthly expenditures as

$$TR_{psm} = \frac{52}{12} \frac{\sum_{w \in m} \sum TR_{psw}}{\sum_{w \in m} 1}, \tag{4}$$

where the normalization controls for the number of weeks in the month (either 4 or 5). We first divide the sum of expenditures by the number of weeks in the particular month and then multiply it by the average number of weeks in the year.

3 Inflation

In this section, we construct an inflation index and compare its dynamics to the official food-at-home inflation subindices.

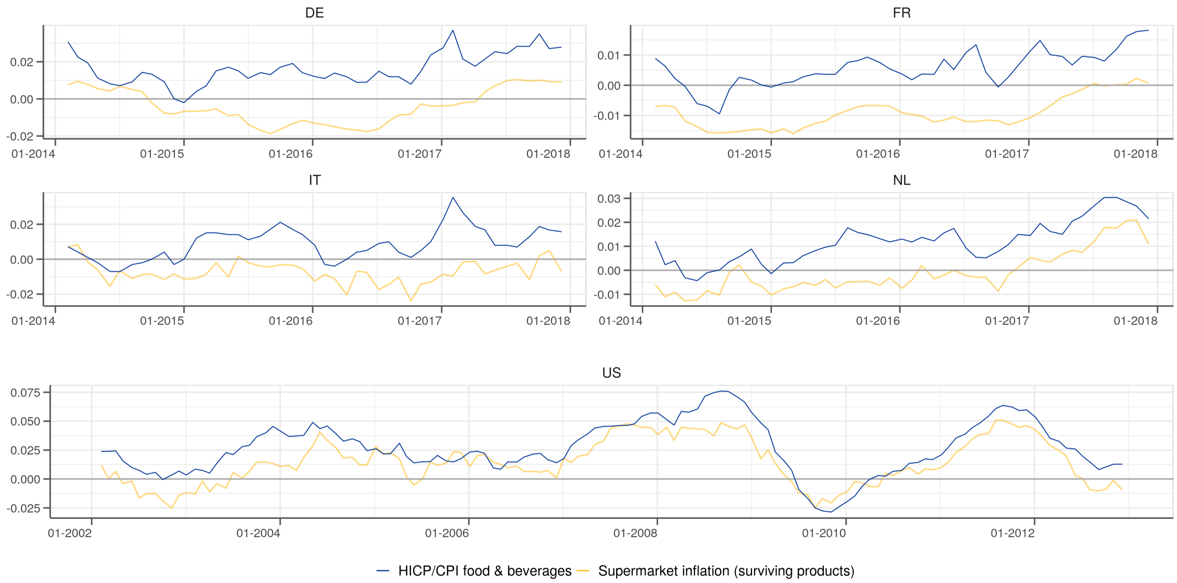
Our baseline inflation index is constructed as a geometric average of price changes weighted by their annual expenditures. Formally:

$$\Pi_t = \prod_{ps} \left(\frac{P_{pst}}{P_{pst-1}} \right)^{\omega_{pst-1,t}}, \tag{5}$$

where Π_t is the gross inflation rate in month t, and P_{pst} is the posted-price of product p in store s in month t. The weight is the the annual expenditure on product p in store s as a share of the annual expenditures. Formally, $\omega_{psy} = \sum_{t \in y} TR_{pst} / \sum_{ps} \sum_{t \in y} TR_{pst}$, where TR is the total revenue (nominal expenditure), and y is the year of month t.[17] The price index $P_t = \prod_{s=0}^{t} \Pi_s$ is simply a chained product of the inflation index.

[17] The index only considers items, which exist both in periods $t-1$ and t, therefore the actual expenditure

Figure 3: The year-on-year change of the IRi supermarket inflation index, and the official food and beverage subindexes.



Note: The figure shows the evolution of the year-on-year IRi supermarket inflation and the official food and beverage subindexes. The comovement between the series is apparent, especially at low frequencies.

Our baseline inflation index captures the business cycle fluctuations in official food and beverage inflation reasonably well. Figure 3 shows the evolution of the year-on-year inflation rates of the two series in each country. The comovement is apparent, especially at low frequencies.[18]

In the upcoming analysis, we decompose the baseline index into key components in order to establish relevant stylized price-setting facts.

---

weights used are

$$\omega_{pst-1,t} = \frac{I_{pst-1,t}\omega_{psy}}{\sum_{ps} I_{pst-1,t}\omega_{psy}},$$

where $I_{pst-1,t}$ is an indicator function that takes the value 1 if product p in store s exists in both months $t-1$ and t.

[18] At the same time, our inflation index underestimates the level of official inflation. As we detail in the appendix, the primary reason for this is that our index excludes the impact of new product introductions. These tend to have small impact on inflation variability at business cycle frequencies (see also Argente and Yeh, 2199), but can substantially raise the level of inflation.

3.1 Temporary sales

A salient feature of price spells is their high-frequency variation. Prices regularly get reduced (or increased) temporarily for a couple of weeks, after which they tend to return to exactly the initial price. As we show momentarily, most price changes in our sample are due to such temporary sales. Previous research has established that the nature of high-frequency price changes is distinct from those of more persistent reference price changes (Nakamura and Steinsson, 2008; Kehoe and Midrigan, 2015; Eichenbaum et al., 2014). While reference prices are driven primarily by costs, sales are used as a marketing tool to trigger households to try out new products and stores, to gain the trade of bargain-hunter households, as well as a tool to fine tune inventory. The high-frequency variation influences inflation dynamics differently than the evolution of reference prices, therefore it is instructive to analyze them separately.

We employ a state-of-the-art sales-filtering technique on the weekly data. We create weekly reference prices (P^f_{psw}) as a 13-week running modal price,[19] which we iteratively update to align the reference-price change with the actual price change as in Kehoe and Midrigan (2015).[20] As an additional step, we control for clearance and introductory sales in the first and last 5 weeks of the spell. We do this by carrying forward (backward) the reference price in the 6th week before the last (after the first) price of the spell.[21] A key advantage of the reference-price filter over a more conventional regular-price filter that controls for V-shaped temporary price cuts (Nakamura and Steinsson, 2008) is that it also controls for temporary increases (spikes) in price spells. Such increases can be rationalized, for example, by inventory management: higher prices temporarily reduce demand and make sure the store does not run out of the product until a new delivery arrives. Spikes can account for as high as one third of high-frequency price changes (Eichenbaum et al., 2011; Kehoe and Midrigan, 2015).

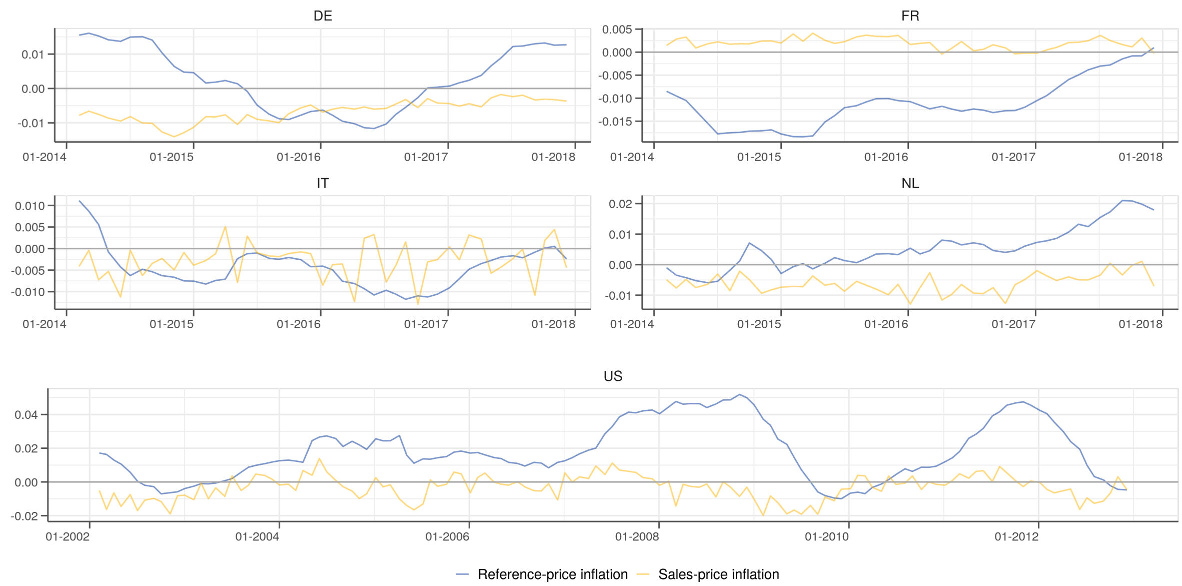
Figure 4 decomposes our baseline inflation series into a reference price inflation series and a sales inflation series. The reference price inflation series is constructed analogously to our baseline series (equation 5) with reference prices replacing posted prices. Sales inflation is

[19]As in Kehoe and Midrigan (2015) the mode is accepted as reference price if at least 50 percent of the observations in the window is non-missing (accuracy) and more than 33 percent of prices are at the mode (cutoff). If the accuracy or cutoff conditions are not met, the previous reference price is carried forward.

[20]Due to the nature of the algorithm, a change to a new reference price is sometimes picked up with a delay (it takes a while till the new price becomes a mode within the rolling window). The algorithm corrects for this by aligning the change in the reference price with the change in the posted price. It achieves this by iteratively replacing the current reference price to the subsequent one if both the current and the subsequent posted prices are equal to next-period reference price.

[21]Argente and Yeh (2199) documents that the frequency of sales start to increase significantly around 5 weeks before the exit of a product. They also document that the sales behavior is special in the first 6 weeks of a product introduction (sales are actually less frequent than later).

Figure 4: The year-on-year reference-price and sales inflation



Note: The figure decomposes inflation into reference-price inflation and a residual sales inflation. The figure shows that the reference-price inflation evolves smoothly and it accounts for most inflation variation at business cycle frequencies. Sales inflation, in contrast, varies little at business cycle frequencies, while it is subject to a sizable high-frequency noise.

defined as the difference between posted-price inflation and reference-price inflation. The figure shows that sales-price inflation is subject to large high-frequency noise and it explains a small share of inflation variability at business cycle frequencies.[22] This is one of the key reasons why we concentrate on reference-price changes in most of the subsequent sections.

Even though sales-inflation accounts for a small share of inflation variability at business cycle frequencies, a large fraction of price changes are due to sales. To show this, we calculate the average monthly frequency of price changes for both posted- and reference prices. We weigh the item-level frequency with annual expenditure weights analogously to our baseline inflation index, and take a simple average over time. Formally, the frequency (ξ_t) of monthly price changes is

$$\xi_t = \sum_s \sum_p \omega_{pst-1,t} I_{pst-1,t}, \tag{6}$$

[22]Previous research has documented that sales-inflation does not respond significantly, or responds only marginally to small aggregate shocks (Anderson et al., 2017; Karadi et al., 2020; Gautier et al., 2022a).

where $I_{pst-1,t}$ is an indicator that takes the value 1 if the posted price of product p in store s in month t changed from the previous month, and 0 otherwise. Frequency of reference-price changes are calculated analogously with an indicator function that takes the value 1 in case of a reference-price change. Table 3 shows the frequency of posted- and reference-price changes in the 4 euro area countries and the US in rows 1 and 2 and shows their ratio in row 3. The table shows that around 2/3 of price changes are due to sales and this share is fairly stable across countries.

Table 3: Frequency of posted- and reference-price changes

Frequency (monthly, mean)	DE	FR	IT	NL	EA4	US
Posted	12.41%	42.23%	27.56%	24.77%	25.18%	39.35%
Reference	4.53%	12.78%	9.04%	10.06%	8.41%	13.34%
Ratio	2.74	3.31	3.05	2.46	2.93	2.95

Note: The table presents the frequency of posted- and reference price changes and their ratio. It shows that around 2/3 of price changes are due to sales.

4 Key moments of price changes

In this section, we characterize key features of reference price changes in supermarkets across the 4 euro area countries and the US. We focus on conventional moments, including frequency, size and kurtosis of price changes, which were found to be relevant by the theoretical literature to influence the flexibility of the aggregate price level, subject to aggregate shocks.

4.1 Frequency

The frequency of reference-price changes is a key indicator of price flexibility. As Table 3 shows, the average frequency in EA4 supermarket prices is fairly low, only 8.4 percent monthly. This suggests that reference prices change infrequently, only once in every 12 months, on average. The low frequency indicates that supermarkets face price-adjustment frictions, which hinder them to adjust prices flexibly to changes in costs. The price rigidity in EA4 supermarkets is higher than in the US, where the frequency of reference price changes is 13.3

percent, implying an average duration of 7.5 months.[23]

There is a notable heterogeneity in frequency across euro area countries. The frequency varies from 9 to 13 percent in most countries (implying a duration between 9-11 months). It is particularly low, 4.5 percent, in Germany (22 months average duration), where the number of competing supermarket chains is also uncharacteristically low (see Table 1).

A sizable share of reference price changes are price decreases, as Table 4 shows. The frequency of decreases substantially exceeds those of the increases in France, where the reference-price inflation is negative over our sample.

Table 4: Frequency of reference-price increases and decreases

Frequency	DE	FR	IT	NL	EA4	US
Increase	2.56	5.65	4.78	5.20	4.24	7.88
Decrease	1.98	7.12	4.26	4.86	4.17	5.46

Note: The table presents the frequency of reference-price increases and decreases across countries. It shows that there are regular price decreases in the data.

4.2 Size

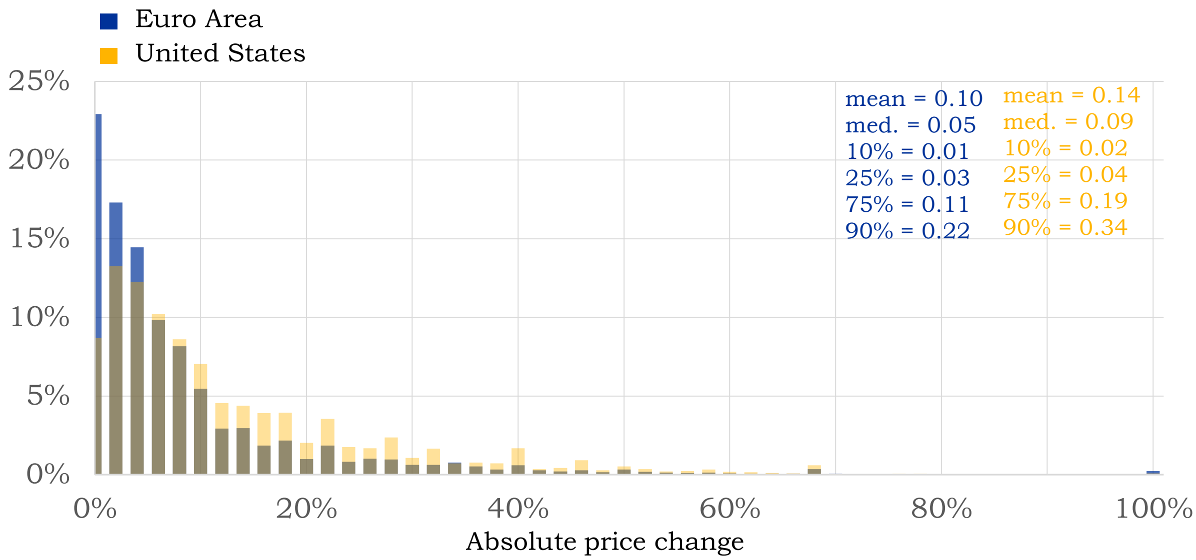
The average absolute size of reference price changes is large: 10 percent in the EA4 countries, on average. Its magnitude way exceeds what could be explained by trend inflation or aggregate fluctuations, which are both small during our sample period. Instead, they indicate an important role for idiosyncratic, product-level shocks. The size of price changes is lower in the euro area than in the US, where it reaches as high as 14 percent. This indicates a more prominent role of product-level shocks in the US than in the EA4. The larger size of idiosyncratic shocks also contribute to the higher frequency of price changes in the US.

Figure 5 shows the histograms of the absolute price change distributions in both areas with some percentiles. The size of price changes in both regions is dispersed, with many small as well as large price changes. A quarter of the prices are smaller than 3 and 4 percent, while around quarter are larger than 11 and 19 percent in the euro area and in the US respectively.

[23] As Figure 19 in the appendix shows, the frequency is stable over time, so the issue of non-overlapping US-EA4 samples should not hinder the international comparison.

The dispersion is larger in the US, where the interquartile range is 15 percent, while it is only 8 percent in the euro area.

Figure 5: Absolute reference-price-change distributions



Note: The figure shows the absolute reference-price-change distributions in both regions together with the means and key percentiles. It shows that the size is large and dispersed in both regions, and it is larger and more dispersed in the US than in the euro area.

4.3 Higher-order moments

The shape of the price-change distribution can be informative about the extent of state dependence in price-setting in a wide-class of models (Alvarez et al., 2014). Figure 6 shows the shape of reference-price change distribution for both regions. The reference-price changes are standardized at the product-store level so as to minimize the potential bias caused by cross-product heterogeneity in the mean or standard deviation of price changes.[24]

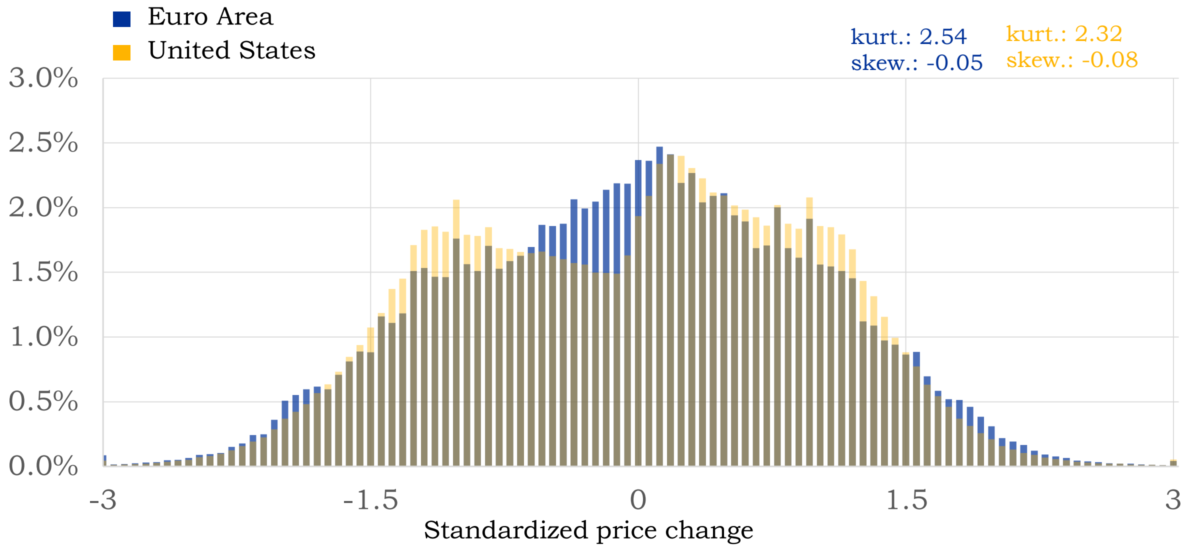
The figures indicate a relatively low kurtosis between 2.5, and 2.3 in the euro area and in the US, respectively, below kurtosis of 3 of the Gaussian distribution.

The distribution shows some pronounced bimodality in the US with some 'missing' mass close to zero, which is in line with the presence of fixed costs of price adjustment. At the same

[24]We only include product-stores that have at least 5 reference-price changes over the sample period.

time, the share of small reference-price changes stays high also in the US, much higher than models with strong state dependence would predict (Golosov and Lucas, 2007).

Figure 6: Standardized reference-price-change distributions



Note: The figure shows the standardized reference-price-change distributions in the euro area and in the US. It shows that the kurtosis of price changes is low in both the US and in the euro area, and it is lower in the US.

5 Evidence on state dependence - generalized and duration hazards

The conventional moments described in the previous section provides only indirect information about an important feature of price setting: the extent of its state dependence. Previous research has established that state dependence, which influences *which* prices adjust, can have as large an impact on the aggregate price flexibility as the frequency, which determines *how many* prices adjust. For example, in realistic models of price setting with strong state dependence (for example, Golosov and Lucas, 2007), the price level can respond almost completely flexibly to monetary policy shocks even though only a few prices adjust. The reason is that in these models, firms face a small fixed menu cost to change prices, so they find it optimal to adjust prices with large misalignments. When these prices change, they change by a lot, which can offset the impact of price rigidity and make the price level flexible.

In this section, we present two sets of moments that are more directly informative about the extent of state dependence than conventional moments. The first moment is the generalized (price gap) hazard function, which expresses the probability of price adjustment as a function of the price misalignments, or price gaps. The slope of the hazard function is directly informative about the extent of state dependence: the higher the slope the more sensitive the probability of adjustment to the price misalignment. A key challenge to measuring the generalized hazard is to obtain a valid proxy for the unobserved price gap. To obtain these, we use the unparalleled granularity of the supermarket scanner data, which helps us to identify the optimal price from the behavior of close substitutes, as we explain below. The second moment is the *price-age* hazard function, which expresses the probability of price adjustment as a function of the time elapsed since the last price adjustment. In models with high state dependence, the age hazard function is upward sloping: the probability of price adjustment increases with the age of the price. The reason is that as time elapses the optimal price tends to drift away from the posted price giving stronger and stronger reasons for a price adjustment. A key empirical challenge in measuring the age hazard is to control for cross-sectional heterogeneity, which biases the slope estimate downward. The granularity of the scanner data allows us to control for this at the lowest, product-store level.

5.1 Generalized hazard

The generalized (or price-gap) hazard expresses the probability of price adjustment as a function of the price gap. The price gap is the distance between the posted price from the optimal 'reset' price the store would set in case all price-adjustment frictions were temporarily absent. The gap, therefore, influences the strength of the product-level price-adjustment impetus: the larger it is, the further the price is from its optimal level causing a potentially larger profit loss through either suboptimal demand (if the price is too high) or suboptimal markup (if the price is too low).

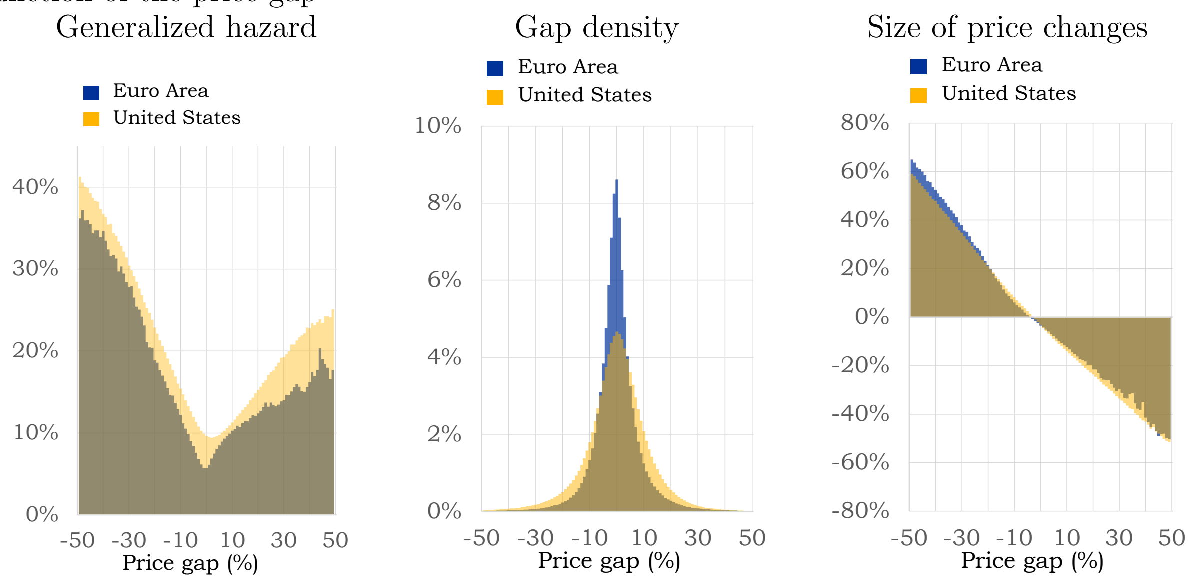
A key empirical challenge is that the optimal reset price is unobservable. As a proxy, we calculate the competitors' reset price (Karadi et al., 2020). It is the average reference price[25] of the same products in those competing stores that also changed the price of the same product in the same month. The measure also controls for permanent store-and-category-level price differences caused by heterogeneity in amenities, geography or market power. The proxy relies on the assumptions that (i) the price of the same good among price-changing competitors tracks well the evolution of the product's wholesale price and aggregate demand conditions, which are the primary drivers of the optimal reset price, (ii) differences in amenities and

[25] By concentrating on reference prices, the measure controls for the impact of temporary sales.

market power between stores cause permanent store-and-category-level differences between prices, and (iii) chains follow national price-setting strategies (DellaVigna and Gentzkow, 2019), so local demand conditions have insignificant impact on the optimal reset prices. We validate our proxy by showing that the size of the price change has a very tight, an almost exactly one-to-one negative relationship with the price gap.

Formally, we formulate the competitor-reset-price gap x_{pst} for product p in store s in month t in three steps. First, we take the (logarithm of) the sales-filtered reference prices p^f_{pst}. Second, we calculate an unadjusted gap as $\tilde{x}_{pst} = p^f_{pst} - \bar{p}^f_{p-st}$, where $\bar{p}^f_{p-st}$ is the average reference-reset-price of the same product across those alternative stores that changed the price of the same product in month t. Third, we deal with the persistent heterogeneity across stores (i.e., chains, locations) by subtracting the average store-and-category-level gap α_{cs} and reformulate the price gap as $x_{pst} = \tilde{x}_{pst} - \alpha_{cs}$, where product p belongs to category c.

Figure 7: Generalized hazard, price-gap density and the size of non-zero price changes as a function of the price gap



Note: The figures shows the frequency of reference price changes (generalized hazard, left panel) and the average size of non-zero reference price changes (right panel) as a function of price gap, and the density of the price-gap (middle panel) in EA4 and the US. The V-shape hazard indicates the presence of state dependence in price setting, albeit at a moderate level in both regions. The density indicates wide dispersion of price gaps, higher on average in the US. The size figure validates the price gap measures showing a tight relationship between the gap and the eventual price-change size.

The middle panel on Figure 7 shows the density of the price gap distributions in the four euro area countries (EA4, Germany, France, Italy, the Netherlands) and the US. To arrive at the densities, we control for unobserved heterogeneity across items and the common impact of aggregate fluctuations by estimating item- and time fixed effects in a panel regression of the form

$$x_{pst} = \alpha_{ps} + \alpha_t + \varepsilon_{pst}, \tag{7}$$

and calculating the share of normalized gaps ($x_{pst} - \hat{\alpha}_{ps} - \hat{\alpha}_t$) in the 101 unit-percentage-point ranges between -50.5 and 50.5 percents. We censor the normalized gaps at -50.5 and 50.5 percents.

The figure shows that the gaps are high, on average, and higher in the US than in the EA4. The average absolute size of gaps is 10% in EA4 and 14% in the US. At the same time, the distribution of the gaps are dispersed in both regions with a high mass of small gaps and a fat tail of large gaps. This is true, even though we control for sales-related price changes as well as permanent differences between the store-specific prices. This evidence is broadly in line with the observations we documented in Section 4 concerning price changes, which are also larger and more dispersed in the US than in EA4. This is not surprising as there is a tight relationship between the price gaps and the size of non-zero price changes, as we show next.

We now turn to assess the relationship between the price gap in period $t-1$ and the probability and average size of price adjustment in the following month t. Our aim is to estimate these relationship non-parametrically with a minimal set of structural assumptions. First, we allocate price gaps into 101 bins, each covering a unit percentage-point range between -50.5 and 50.5 percents. The indicator function $I_{pst-1}^{[x_{j-1},x_j)}$ for bin j takes the value 1 in case the gap $x_{pst} \in [x_{j-1}, x_j)$, and 0 otherwise. Second, we estimate a relationship coefficient (β_y^j) between the gap x and a variable of interest $y_{pst,t+1}$ (frequency or size) for each bin j using the following panel specification:

$$y_{pst,t+1} = \sum_{j=1}^{J} \beta_y^j I_{pst-1}^{[x_{j-1},x_j)} + \alpha_{ps} + \alpha_t + \varepsilon_{pst} \tag{8}$$

where α_{ps} are product-store-, and α_t are time fixed effects. The fixed effects help us to control for unobserved heterogeneity across items and common co-movement caused by aggregate fluctuations. And, third, we obtain the estimated relationship as a sum of two components. The first component is the β_y^j coefficients for $j = [1, 101]$. The second component is the average of the estimated fixed effects $\text{mean}_{ps}\hat{\alpha}_{ps} + \text{mean}_t\hat{\alpha}_t$ added to each bin j. Adding the second component makes sure that the weighted average across bins approximates the sample average of the variable of interest y.

The right panel on Figure 7 shows the average size of non-zero price changes as a function of the price gap in EA4 and the US. It is estimated following the above described steps, when the dependent variable is the non-zero reference-price changes $y_{pst,t+1} = \Delta p^f_{pst+1}|_{\Delta p^f \neq 0}$. The figures show a tight, negative, almost exactly one-to-one relationship between the gap and the average non-zero price changes in the subsequent month. This validates our price gap measure by showing that stores choose to close the gap, on average, when they adjust the price.

We are now ready to turn to one of the key empirical moments we are interested in: the generalized hazard function, shown on the left panel of Figure 7. They are estimated for each region following the steps outlined above, when the dependent variable is an indicator function that takes the value 1 in case the reference price of product p in store s changed in period t+1, and 0 otherwise $y_{pst,t+1} = I^f_{pst+1}$. The figures show clear evidence for state dependence in price setting in both regions: the probability of price adjustment clearly increases with the price gap as illustrated by the V-shape of the hazard functions. The (weighted) average slope is 0.51 in EA4 and is 0.38 in the US, which suggests that the state dependence is moderate in both regions (see Section 5.3 for further discussion), and somewhat larger in EA4. The difference between the regions is caused by the larger slope at lower gaps, where the largest mass of price gaps are concentrated. The height of the hazard function is larger in the US, in line with the higher frequency of price changes there, as we have already documented above.

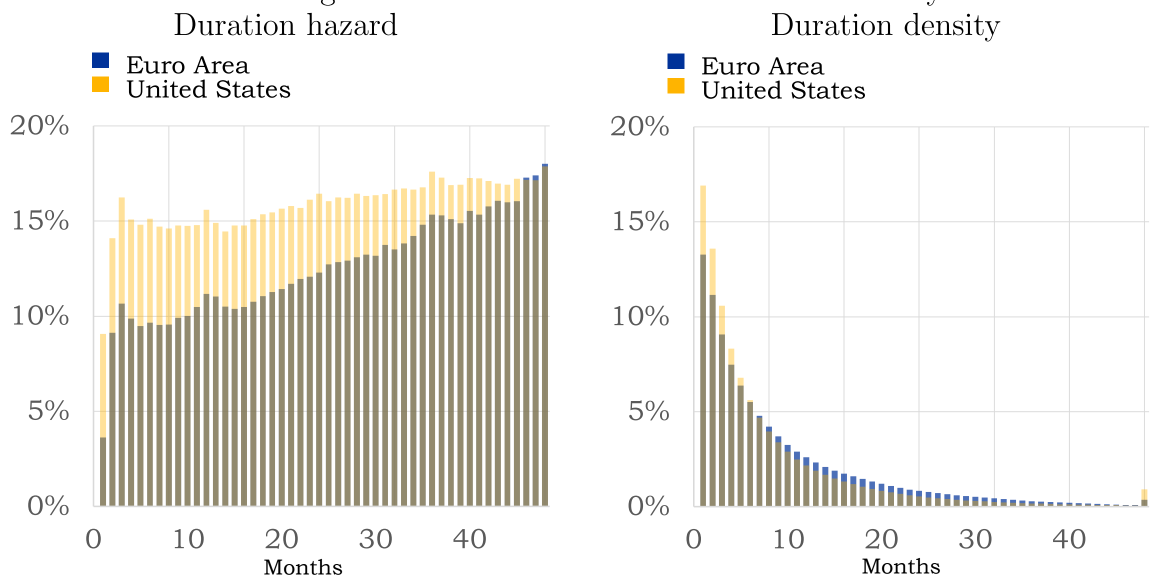
5.2 Duration hazard

An alternative way of looking at state dependence is the duration hazard, which expresses the probability of price adjustment as a function of the months elapsed since the last price adjustment. In the presence of state dependence, the duration hazard is upward sloping as the the optimal price drifts further and further away from the posted price. The advantage of using granular scanner data to estimate the hazard function is that we can control for the cross-item heterogeneity, which can bias the slope estimate downward We estimate the following panel regression

$$I_{pst,t+1} = \sum_{j=1}^{48} \beta^j I^j_{pst-1} + \alpha_{ps} + \alpha_t + \varepsilon_{pst}, \tag{9}$$

where the indicator function I^j_{pst-1} takes a value 1 if the reference price of product p in store s in month $t-1$ is j months old, and 0 otherwise. As with the generalized hazard, we add the average of the estimated item- and time fixed effects to the β^j coefficients in order to make the weighted average of the coefficients approximate the frequency of reference-price changes.

Figure 8: Duration hazard function and density



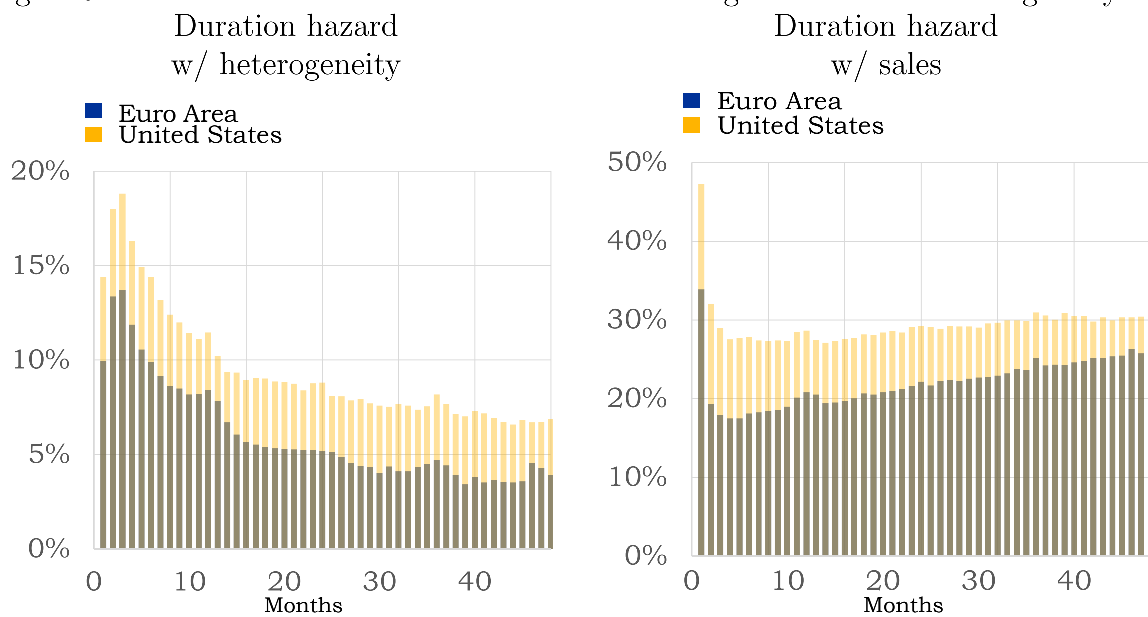
Note: The figures show the probability of reference price change as a function of the months elapsed since the last reference-price change (duration hazard, left panel) in EA4 and US, and the density of prices with various age (duration density, right panel). The figures indicates the presence of state dependence in price setting, somewhat stronger in the EA4 than in the US.

The left panel of Figure 8 shows the results for EA4 and the US. It shows that the duration hazard is upward sloping in both regions: the probability of adjustment increases with the age of the product. The slope of the adjustment hazard is higher in EA4 than in the US. Notably, the hazard function is approximately linear. This is especially true, if we disregard the low estimated adjustment frequencies of recently adjusted prices, where the sales-filtering might introduce uncertainty and a potential downward bias into the estimation by mechanically identifying as sales high-frequency price fluctuations.

Controlling for both cross-item heterogeneity as well as sales-related price changes is important for the results. The left and right panels of Figure 9 show, respectively, our estimates without controlling for fixed effects in equation (9) and using posted-price changes, as opposed to reference price changes. The figures show that both factors bias the estimated slope downward, so much so that in both cases we would erroneously conclude that the hazard function is downward sloping.

The right panel on Figure 8 shows the density of the price-age distribution. The figure shows

Figure 9: Duration hazard functions without controlling for cross-item heterogeneity and sales



Note: The figures show the duration hazards without controlling for cross-item heterogeneity (left panel) and including sales-related price changes (right panel). The figures show that controlling for both factors in important to conclude that the hazard function is upward sloping.

that the price-age is dispersed in both regions: most reference prices are young, but there are a noticeable share of reference prices that are older than 2 years. Prices in the US are younger than in EA4 in line with the more frequent reference-price adjustments there.

5.3 State dependence and price-level flexibility

In the previous section, we have argued that the V-shaped generalized hazard function and the upward sloping duration hazard function are in line with state dependence in price setting. In this section, we quantify the extent of this state dependence. A natural measure of state dependence is how much it contributes to the price flexibility, specifically to the price-level impact of a permanent money shock. To measure this, we follow the framework of Caballero and Engel (2007) and its extension by Karadi et al. (2020), who showed that under mild conditions, the generalized hazard function and the density provide sufficient information to quantify the contributions of the intensive and extensive margins of adjustment. We first describe the framework and explain how the relevant objects in the model relate to our empirical moments before turning to use it to decompose an aggregate money shock to adjustment margins.

In the flexible price-setting framework of Caballero and Engel (2007), there are a continuum of firms each producing a single product i. Firms set the (log nominal) prices of their product (p_{it}) subject to a price-adjustment friction. If these frictions were temporarily absent, the optimal price in period t would be p^*_{it}. The optimal price is driven by both aggregate and idiosyncratic factors $p^*_{it} = m_t + \nu_{it}$. For simplicity, we assume that shocks to both m_t and ν_{it} are permanent. The aggregate shock m_t shifts the optimal nominal price of all firms, whereas the idiosyncratic shock ν_{it} affects only firm i. The gap between the price and its optimal value $x_{it} = p_{it} - p^*_{it}$ is the relevant state variable and is sufficient to characterize each firms' price-setting choice. Assuming that the product i is sold in a continuum of stores, the average price set by price-changing stores reveals the optimal price p^*_{it}, in line with our empirical application.

The firms' price adjustment decision can be described by a generalized hazard function $\Lambda(x)$. The function takes values between 0 and 1, and its value expresses the probability of price adjustment for a firm with a price gap x. The hazard function is constant in the time-dependent Calvo (1983) model: there, the probability of adjustment is independent of the price gap. At the other extreme, in the fixed menu cost model (Caplin and Spulber, 1987; Golosov and Lucas, 2007), the hazard function is a step function, which takes the value 0 when the gap is within the inaction band, and 1 otherwise. Caballero and Engel (2007) shows that a continuum of intermediate hazard functions can arise when the menu cost is an i.i.d. random variable Dotsey et al. (1999), and when the firm is subject to rational inattention friction as in Woodford (2009) (see also Alvarez et al., 2020).

In this economy, inflation can be expressed as

$$\pi = \int -x\Lambda(x)f(x)dx \tag{10}$$

where $f(x)$ is the density of price gaps across firms, and we suppressed subscripts for notational convenience. The expression is intuitive: the inverse price gap $(-x)$ is the size of the price adjustment, when it takes place, and the hazard is the probability of a price adjustment taking place. Their product summed across the gap distribution and weighted by the density of the gap is, therefore, equal to the inflation rate.

The question is how flexibly the inflation rate reacts to a small aggregate money increase m? Caballero and Engel (2007) points out that the aggregate shock increases the optimal price of all firms, so it reduces the price gaps of each firms uniformly. The response to the aggregate shock can be therefore expressed as a derivative of the expression on the right-hand side of equation (10) with respect to x, which implies

$$\frac{\partial \pi}{\partial m} = \underbrace{\int \Lambda(x) f(x) dx}_{\text{intensive}} + \underbrace{\int x \Lambda'(x) f(x) dx}_{\text{extensive}}, \tag{11}$$

where $\Lambda'(x)$ is the slope of the hazard function. The expression has two terms. The first term, which Caballero and Engel (2007) dubs the intensive margin results for each adjusting firm changing their prices by marginally more to incorporate the impact of the aggregate shock. Notably, it is exactly equal to the frequency of price adjustment and this is the only margin that is active in the time-dependent Calvo (1983) model, which has a constant hazard. The second term is the extensive-margin effect, which takes into account any shifts in the identity of price-adjusting firms. The slope of the hazard function appears in this expression, because it measures the mass of new price-adjusters as the aggregate shock shifts the price-gap density. The extensive margin is powerful if the new adjusters are primarily those with large price gaps. This tends to be the case with strongly state-dependent (S,s)-type menu cost models (Golosov and Lucas, 2007), where it is optimal to adjust prices with the largest gaps in the presence of fixed menu costs of price adjustment.

Our empirical estimates on the hazard function and the density of price gap shown on Figure 7 allow us to conduct the Caballero and Engel (2007) decomposition described by equation (11). The intensive margin effect is the average frequency, which is also the average of the hazard function weighted by the density at each bin. To obtain the extensive-margin effect, we first calculate the slope of the hazard function at each bin as the centered finite difference between subsequent bins. Second, we multiply the slope with the size of the misalignment and, third, we calculate a weighted average using the density weight of each bin.

The first row of Table 5 shows the overall impact effect of a permanent money shock in the euro area, in the US and in each four euro area countries. The table shows that stronger state dependence further increases price flexibility in the US relative to EA4. [26] Taking into account state dependence mitigates the heterogeneity in the flexibility in supermarket prices across euro area countries, but supermarket price-level flexibility in Germany remains noticeably below those in other euro area countries.

The second and the third rows of Table 5 show relative contributions of each adjustment margins relative to the overall effects. The relative contribution of the extensive margin effect is 25 percent in the euro area, approximately equal to that in the US. This means that

[26] This might be surprising given we have found that the average slope of the hazard is higher in the euro area than in the US. The reason of the weaker extensive margin effect lies in the fact that the slope is higher mostly at the low gap ranges, which contribute little to the extensive margin effect.

accounting for state dependence raises the price-level flexibility by around 33%=25%/(1-25%) relative to a time-dependent benchmark (Calvo, 1983). This is a meaningful increase, but it is small relative to an (S,s)-type menu cost model, where the price-level flexibility with the same frequency is predicted to be 6 times that of a time-dependent benchmark (Golosov and Lucas, 2007).

As Table 5 also shows there is a sizable heterogeneity among euro area countries in the extent of the contribution of state dependence to the aggregate price-level flexibility. It is the lowest in France, where it only raises aggregate price flexibility by around 15% relative to the time-dependent benchmark, and highest in Germany, where it raises the aggregate price flexibility by 70%, albeit from a lower level.

Karadi et al. (2020) further decomposed the extensive margin effect into two components. The first is the so called gross extensive margin effect. This comes from a shift between price increases versus price decreases that are uniform across the price-gap distribution. The second is the selection effect, which measures whether new adjusters are disproportionately coming from products with large price gaps. In case of linear generalized hazard, the gross extensive margin is the only extensive-margin active. Selection becomes significant, if the generalized hazard function is sufficiently convex. The fourth- and fifth rows of Table 5 shows the relative contributions of the gross extensive margin and the selection effects to the overall price-level impact. The results show that the overwhelming majority of the extensive margin effect is accounted for by the gross extensive margin effect, and the selection effect is uniformly small, and occasionally negative. These results are not surprising given the close-to-linearity of the generalized hazard functions, especially at the most relevant regions of the state space.

6 Structural analysis

In this section, we interpret the evidence through the lens of a state-of-the-art price-setting model (Woodford, 2009). We ask which structural features drive the differences between the US and euro area price setting in the food retail sector.

6.1 Structural model

The model we use is a quantitative price-setting model with price-adjustment costs and information frictions. It provides a microfoundation for the popular 'random menu cost' models (Dotsey et al., 1999; Alvarez et al., 2020) and includes the time-dependent Calvo (1983) model and the fixed-menu-cost model of Golosov and Lucas (2007) as special cases.

Table 5: Overall impact effect and relative contributions of adjustment margins

Margins	EA4	US
Overall impact effect	11.5%	17.2%
Intensive (relative)	74.6%	75.1%
Extensive (relative)	25.4%	25.0%
Gross extensive (relative)	25.5%	19.5%
Selection (relative)	-0.1%	5.5%

Note: The table presents the overall impact effect of a marginal money shock and the relative contributions of the intensive- and extensive margin effects (Caballero and Engel, 2007), the latter further decomposed into gross-extensive-margin and selection effects (Karadi et al., 2020). The table shows that stronger state dependence further increases price flexibility in the US relative to EA4.

We sketch the key features of the model here, and direct the interested reader to the original paper for details and derivations. The paper generalizes the fixed menu cost model of Golosov and Lucas (2007). There is a continuum of differentiated goods (i), which are sold in a market with monopolistic competition. This market structure gives the producer of each good market power to set prices at a markup above the marginal cost. The market power is determined by the elasticity of demand, which, in turn, is governed by the (constant) elasticity of substitution parameter ε.

The production requires labor, and the product-specific productivity is subject to idiosyncratic shocks. As argued by Golosov and Lucas (2007), these shocks are necessary to explain the large absolute size of price changes. Specifically, productivity follows a random walk, with an idiosyncratic shock $z_t(i)$ with standard deviation σ_z ($A_t(i) = A_{t-1}(i) + z_t(i), z_t(i) \sim N(0, \sigma_z^2)$). As in Woodford (2009), we concentrate on the stationary equilibrium without aggregate fluctuations. All the relevant information for the firm is incorporated into its price gap, defined as the distance of its (log) price from its (log) optimal price $x(i) = p(i) - p^*(i)$.[27] In particular,

[27] It is easy to see that the price gap can be equivalently expressed as the difference of the normalized price ($q(i)$) as defined in Woodford (2009) from its optimum ($x(i) = p(i) - p^*(i) = q(i) - q^*$).

its profit is a function of the price gap and it is maximized when the price gap is zero. The price gap fluctuates as the idiosyncratic shocks hit the optimal prices, and the firm does not necessarily reset it to zero because the adjustment of the product price ($p(i)$) is costly.

The firms face two types of adjustment costs. First, as in Golosov and Lucas (2007), the firm needs to pay a fixed (menu) cost κ in case it conducts a price review. After paying the cost, the firm obtains full information, thereby it learns its price gap and optimally closes it. Second, the firm needs to decide about the timing of its price review under imperfect information about the state of the economy, therefore about its price gap. The imperfect information is modelled as rational inattention, whereby the firm can obtain a costly signal $f(x)$ about the price gap, and the cost increases linearly with the informativeness (I) of the signal with a coefficient θ ($\theta I = -\theta E[\log f(x)]$)). Woodford (2009) establishes two useful results. First, optimal policy is described by a hazard function $\Lambda(x)$: a firm chooses to obtain a signal with probability $\Lambda(x)$ as a function of its price gap x and conducts a price review in case it receives a signal. Second, the functional form of the hazard function is well defined, it is (weakly) increasing with the (absolute value of the) price gap and its slope depends on the information cost parameter θ. As the cost parameter $\theta = \infty$ increases without limit, the hazard function approaches a constant, which is the time-dependent Calvo (1983) case, and the cost parameter is zero $\theta = 0$, the hazard function approaches a step function as in the fixed menu cost Golosov and Lucas (2007) case.

6.2 Estimation

Our goal in this section is to identify the most relevant structural features, which account for the differences between the US and EA4 price setting. We do this by using the empirical moments obtained in previous sections to match key structural parameters in the model.

We calibrate some parameters to levels used in the literature exactly as Woodford (2009), with one difference. We set the elasticity of substitution parameter (ε) to 2, which is lower than the parameter (6) used by Woodford (2009). The lower parameter implies weaker competition between firms and a flatter profit function, which helps us to match the consistently low slope of the empirical hazard function for large gaps. Our conclusions are robust to reverting to the same elasticity parameter as Woodford (2009), primarily because the fit of the hazard function is similarly good under both parameters for the range of gaps, where most of the mass is.

The three parameters we estimate are the (i) standard deviation of the idiosyncratic shocks (σ_z), which affects the volatility in the product-level environment, and the two parameters governing the price-adjustment costs: the (ii) review (menu) cost (κ) and (iii) the information cost (θ). We estimate these parameters by targeting three moments: the shape of the generalized hazard[28], and the frequency and the size of the price changes.[29] These moments well identify the parameters because, first, the hazard function estimates pin down the information cost parameter, which governs the slope of the hazard function. Second, the frequency and the size of price changes jointly pin down the review cost, which reduces the frequency and increases the size, and the standard deviation of idiosyncratic shocks, which increases both the frequency and the size of price changes. We also check how the model matches some untargeted moments, like the duration hazard and the standardized price change distribution.

6.3 Results

Figure 10 shows the match of the theoretical and empirical generalized hazards and densities for EA4 and the US. The fit is good for both the hazards and the densities, especially over the range where most of the mass concentrates as indicated by the shaded areas. The distribution of the gaps in the euro area is more concentrated than in the US, which the theoretical distribution can only partially capture. Specifically, even though the theoretical distribution captures the lower standard deviation of the gap distribution (a moment, which is targeted through the size of price changed - the inverse price gap) in the EA4, it underestimates the kurtosis of the distribution. One reason for the high kurtosis can be the heterogeneity in the dispersion of price gaps and, consequently, price changes across products. Indeed, the fit to the *standardized* price-change distribution (see the second row of Figure 11), which controls for this heterogeneity, is superior, and actually somewhat better for the euro area than for the US. The model is also reasonably good at matching the duration hazard, even though the moment was not directly targeted.

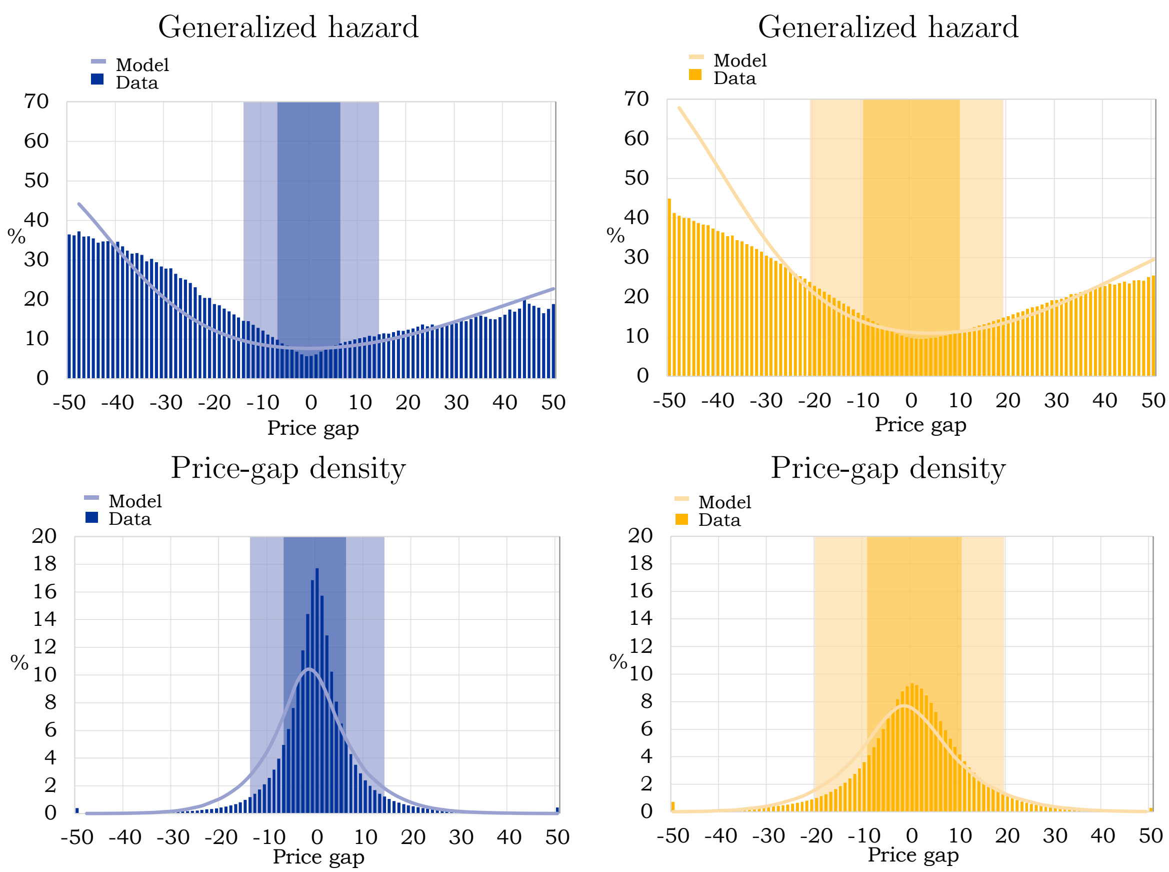
Table 6 shows the estimated structural parameters for the euro area and the US. Several results are worth pointing out. First, the information cost parameters are finite (infinite costs would imply no state dependence) indicating the presence of state dependence, in line with an increasing hazard function. Second, the information costs are fairly high, which indicate a mild state dependence, which is quantitatively closer to the time-dependent Calvo

[28]The estimation algorithm minimizes the squared difference between the empirical and the theoretical hazard functions at each bin, weighted by the price-gap density over the same bin.

[29]For internal consistency of our quantitative exercise, the frequency and the size measures we match here are derived from (unweighted, truncated at +-50%) generalized hazard and density estimates. In particular, frequency is measured as $\sum_j \Lambda_j f_j$, and size as $\sum_j |x_j| \Lambda_j f_j / \sum_j \Lambda_j f_j$, where Λ_j is the height of the generalized hazard, f_j is the relative share of products in the price-gap bin j and x_j is its midpoint. These measures are not equal to the (weighted, untruncated) frequency (EA4: 8.4% vs. 8.6%, US: 13.0% vs. 13.3%) and size (EA4: 8.6% vs. 10%, US: 11.6% vs. 14%) measures reported in Sections 4.1 and 4.2, but they are close with comparable relative magnitudes.

Figure 10: Estimation, targeted moments

$EA4$ US

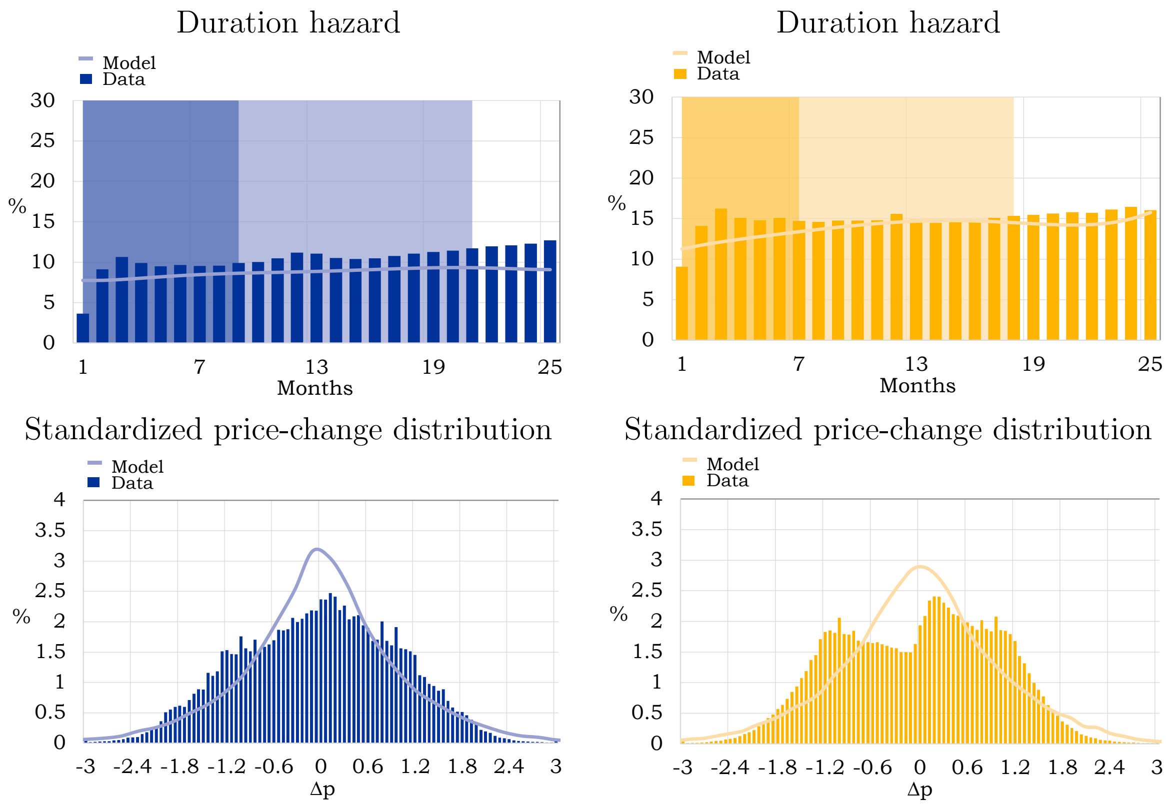


Note: The figures show the match of the theoretical and the empirical generalized hazards and densities in the euro area (EA4) and the US. Shaded areas cover the 67% (darker) and 90% (ligher) mass of the price-gap density.

(1983) model than the strongly state-dependent fixed-menu-cost Golosov and Lucas (2007) model. This is in line with flat hazard functions. Third, even though the information cost parameters are not equal across the regions, they both indicate mild state dependence, which contribute very weakly to the flexibility of the food inflation. Fourth, the estimated review (menu) cost parameters are very close across the regions, suggesting that price-adjustment costs do not contribute to the differences in the frequency of price adjustments across regions.

Figure 11: Estimation, untargeted moments

*EA*4 *US*



Note: The figures show the match of the theoretical and the empirical duration hazards and the standardized price change distributions in the euro area (EA4) and the US. Shaded areas cover the 67% (darker) and 90% (ligher) mass of the price-age density.

Instead, fifth, the key structural reason for the differences across the regions is the distinct standard deviation of the idiosyncratic shocks. In other words, the volatility of the product-level environment is higher in the US, which leads to at the same time (i) higher frequency of price changes, (ii) higher size of price changes and (iii) more dispersed price gap distribution.

Table 6: Estimated parameters

Parameters	EA4	US
Review cost (κ)	9.0%	9.2%
Stdev. of idiosyncratic shocks (σ_z)	3.3%	5.5%
Information cost (θ)	0.72	0.46

Note: The table shows that state dependence is present, but mild in both regions (information frictions are high). Higher idiosyncratic-shock variation in US plays a prominent role in explaining higher frequency and size of price changes.

7 Price setting during the first wave of the Covid-19 in Germany and Italy

In this section, we analyse the price-setting response of German and Italian supermarkets to the first wave of the Covid-19 lockdowns. The shock had a large, persistent, and broadly similar effect on supermarket demand in both countries. Contrasting the response in the two countries is relevant because price setting is heterogeneous across the countries: the frequency of price changes is persistently higher in Italy than in Germany (see Section 4), while the extent of state dependence is mild and not too different. Price-setting models, therefore, predict higher flexibility of the Italian supermarket inflation conditional on demand shocks, which we can test in the data.

7.1 Data

The analysis in this section uses an auxiliary dataset, which covers large German and Italian supermarkets over three months encompassing the first wave of the Covid-19 pandemic from mid-February till mid-May in 2020. The dataset also covers the analogous period in 2019, which we will use as the base period in our index calculations. The dataset covers 20 2-digit ZIP areas.[30]

Our analysis uses the 2013-2017 German and Italian pre-Covid sample as a benchmark to assess the significance of changes observed over the 2019-2020 Covid period. To minimize the impact of compositional shifts over time, we restrict our baseline sample to stores and products which appear with positive sales in both the first quarter in 2013 and the sample quarter in 2020. The majority of stores are such 'established' stores[31]. A sizable fraction of the products are such 'established' products[32].

7.2 Supermarkets and the first wave of the Covid-19 pandemic

The Covid-19 pandemic and the accompanying lockdown measures had a large and persistent impact on supermarket demand. During the lockdowns, access to food-away-from-home was

[30] The ZIP areas in the sample cover 16% and 40% of the population and a share of supermarket expenditures of 22% and 46% throughout 2013-2017 in Germany and Italy, respectively.

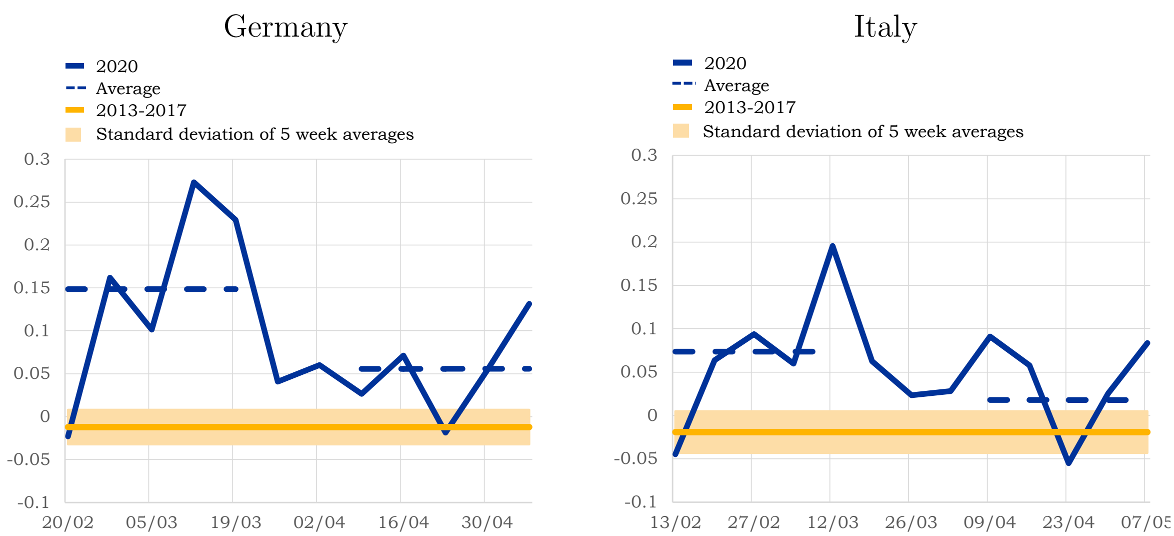
[31] 668 out of 815 unique stores in Germany and 1486 out of the 2387 unique stores in Italy.

[32] 57.000 out of 266.000 unique products in Germany and 83.800 out of 535.500 unique products in Italy with an expenditure share in Germany of 43.43% and in Italy of 42.43%.

severely restricted, while supermarkets were deemed essential and sheltered from the impact of the lockdowns. The Italian government imposed a national lockdown on 9 March and gradually eased it only after mid-May. In Germany, a federal lockdown was introduced on 22 March and was gradually eased from early May. In both countries, supermarkets stayed open during the lockdowns, while alternative access to food and beverages was restricted, as restaurants, canteens, and bars were deemed unessential and were ordered to be closed.

Our data allows us to quantify the magnitude of the demand change in supermarkets because the scanner data includes the weekly quantity sold as well as the weekly expenditures for each product in each store. We restrict attention to the demand change for established products in established stores, which are the focus of our analysis. We measure year-on-year nominal expenditure growth as the 52-week change in overall expenditure on items (which we define as product-store combinations) sold in positive quantities both in the current and base weeks. Real expenditure growth is the difference between nominal expenditure growth and the inflation rate (for the details of inflation measurement, see the next section).

Figure 12: Real expenditure growth in supermarkets during the first wave of the Covid-19 pandemic, year-on-year



Note: The figure shows the weekly, year-on-year real expenditure growth (blue solid line) between mid-February and mid-May in 2020 in Germany and Italy. It shows that the 5-week-average expenditure growth (blue-dashed line) exceeded the average long-term expenditure growth (yellow dashed lines) by more than a standard deviation in both Germany and Italy. The expenditure growth was particularly high in the weeks preceding the lockdowns ('stock-up' shock), but stayed persistently high also during the lockdowns.

Figure 12 shows the evolution of real expenditure growth (blue solid line) between mid-February to mid-May in German and Italian supermarkets. The figure shows that the expenditure growth significantly exceeded its long-term average (yellow dashed line). The increase was particularly pronounced during the weeks preceding the introduction of the lockdowns. The growth rate reached as much as 18-28 percent during this 'stock-up shock,' as households increased their home-stock of non-perishable groceries for precautionary reasons. The expenditure growth during the lockdowns stayed persistently well above average. It stabilized at around 7.5 percent in Germany and at 3.5 percent in Italy, which significantly exceeded the below-zero long-term real expenditure growth experienced over the 2013-2017 period (among the established products, which are the focus of our analysis).

7.3 Supermarket inflation

We measure inflation using the year-on-year change in the Tornqvist price index with quarterly expenditure weights. The Tornqvist index is a superlative price index with desirable welfare-theoretical properties[33]. Quarterly expenditure weights reduce the impact of high-frequency variation in the composition of products both due to seasonal factors and temporary sales. Additionally, concentrating on year-on-year indexes minimizes the impact of seasonal variation as well as the potential impact of the 'chain drift,' which can be present with higher-frequency indexes relying on scanner data (Ivancic et al., 2011).

Formally, we calculate inflation as

$$\pi_w = \sum_{ps} \gamma_{psw} (\log P_{psw} - \log P_{psw-52}), \tag{12}$$

where P_{psw} is the posted price of product p in store s in week w and the weights are

$$\gamma_{psw} = \frac{I_{psw,w-52}(\omega_{psq-4} + \omega_{psq})/2}{\sum_{ps} I_{psw,w-52}(\omega_{psq-4} + \omega_{psq})/2}, \tag{13}$$

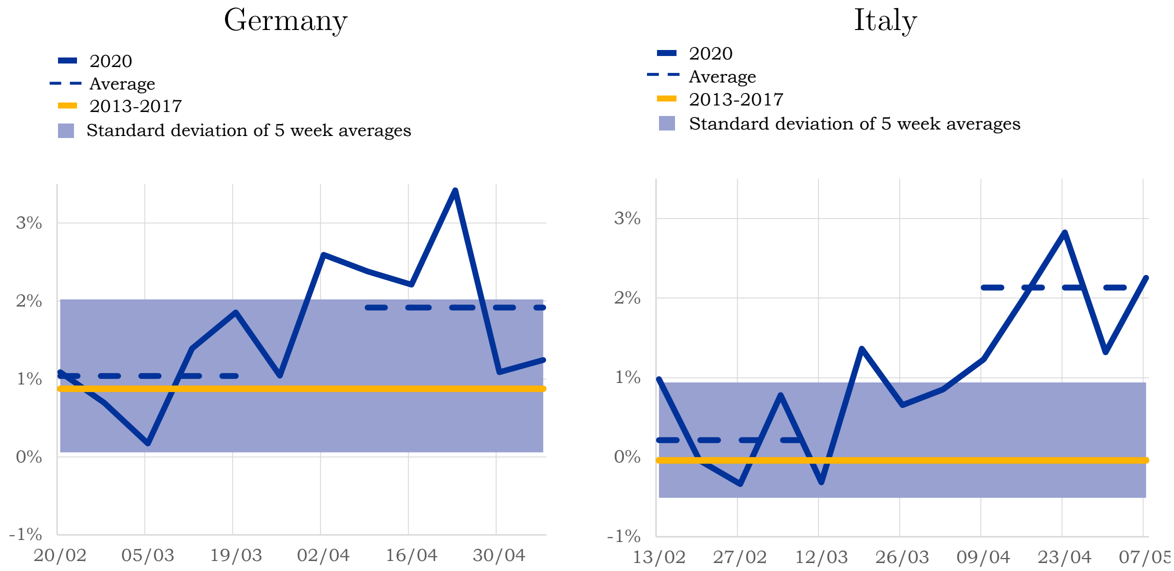
where $I_{psw,w-52}$ is an indicator function that takes the value 1 if product p in store s is sold in strictly positive quantities in both w and $w - 52$ and 0 otherwise[34], and ω_{psq} is the quarterly expenditure share of product p in store s in quarter q.[35]

Figure 13 shows the weekly, year-on-year supermarket inflation in Germany and Italy between mid-February and mid-May. We concentrate on the 5-week-average inflation, which smooths

[33]It is the second-order approximation of the welfare-relevant price index under an arbitrary homothetic utility function.

[34]We match weeks with previous-year weeks based on their distance from the Easter week, the strongest

Figure 13: Supermarket inflation during the first wave of the Covid-19 pandemic, year-on-year



Note: The figure shows the weekly, year-on-year supermarket inflation (blue solid line) between mid-February and mid-May in 2020 in Germany and Italy. It shows that the average inflation in the first 5 weeks (blue-dashed line) stayed close to the average inflation rate during the first two quarters of 2013-2017 (yellow dashed lines). Over the course of the quarter, the 5-week-average inflation (blue dashed line) increased sizeably in both Germany (0.95%) and Italy (1.89%). The change stayed within a +- one-standard-deviation band in Germany, but exceeded it in Italy.

out some high-frequency variability of the weekly series. The 5-week-average inflation started at around its long-term average in 2020 in both Germany and Italy and increased throughout the quarter in both countries. The increase was higher and clearly exceeded a one-standard-deviation band[36] in Italy (1.89 percentage points), while it was smaller and stayed within a one-standard-deviation band in Germany (0.95 percentage points). The increases are comparable to the HICP food-and-beverage subindexes between February and May in Italy (1.96

seasonal factor over the mid-February-mid-May period we have data for in 2019 and 2020.

[35]The ensuing supermarket inflation rates in both countries co-move with the respective HICP food and beverages subindices. The correlation coefficients of the monthly inflations are 43% in Germany and 54% in Italy. The level of supermarket inflation is below the HICP subindices. The main reason is that we concentrate on surviving products and ignore the impact of new product introductions, which generate a major share of trend inflation.

[36]The band shows the standard deviation of 5-week-inflation rates over the first two quarters of the years between 2013-2017.

percentage points) and Germany (0.79 percentage points).

The evidence points to notable heterogeneity between Germany and Italy in price responses to large aggregate demand shocks. If we attribute all the changes to a demand shift, the supply elasticities implied by the price and quantity indexes are 0.95%/9.5%=0.1 for Germany and 1.89%/6.6%=0.29 for Italy. For these calculations, we measured the changes in the quantity indexes as the difference between the real expenditure growth during the last 5 weeks of our sample relative to the long-term real expenditure growth. The supply elasticity in Germany is comparable to the low elasticity (0.07) reported by Gagnon and López-Salido (2020) after large local demand shocks in the US. However, the measured supply elasticity is much higher in Italy despite the similarity of the shock, the type of retailers and the basket of products. In the next two sections, we analyse some features of price setting that contributed to the inflation impact of the Covid-19 shock and can explain some of the differences between Germany and Italy.

7.4 Temporary Sales

A sizable fraction of price changes in our sample is due to temporary sales, which are fully reversed within a short time span. Previous research has established that the nature of such high-frequency price changes is distinct from those of more persistent reference price changes (Nakamura and Steinsson, 2008; Kehoe and Midrigan, 2015; Eichenbaum et al., 2014). While reference prices are driven primarily by costs, sales are used as a marketing tool to trigger households to try out new products and stores and to gain the trade of bargain-hunter households. Therefore, the frequency of sales-related price changes, mostly driven by cross-product and cross-store competition, has a more muted impact on the inflation-effect of an aggregate demand shift than the frequency of reference price changes. Still, a key outstanding question in the literature is whether sales-related price changes remain an active adjustment margin that retailers use to respond to aggregate shocks.

We identify reference prices as the (highest) mode within a centered rolling window, and, in turn, we define sales as temporary downward deviations from this reference price (Kehoe and Midrigan, 2015; Eichenbaum et al., 2014).[37] We set the size of the rolling window to 5 weeks. This is a conservative choice. It categorizes fewer price cuts as sales than Kehoe and Midrigan (2015) or Eichenbaum et al. (2014), which used 11 and 13 weeks windows, respectively. However, the shorter window has minimal impact on the time variation of the

[37] There are also frequent temporary *upward* deviations from the reference price (spikes), but they are not the focus of the analysis.

frequency and size of sales[38], and allows us to assess changes in reference prices over our 13-week sample period in 2020 (see next section).

Table 7: Average moments, 2013-2017

	Surviving-product Tornqvist inflation	Annual change-frequency Posted	Annual change-frequency Reference	Temporary sales Share	Temporary sales Size
Germany	0.87%	56%	48%	17%	13%
Italy	-0.04%	75%	68%	17%	12%

Note: The table lists some relevant moments of posted- and reference prices and temporary sales. It confirms that reference price changes explain most of the posted-price changes at the annual frequency. Price changes are more frequent in Italy than in Germany. Furthermore, a sizeable fraction of products are on sale at any given time in both Germany and Italy and the frequency and the size of sales are similar in the two countries.

Table 7 shows some relevant price-setting moments in Germany and Italy measured over the 2013-2017 sample. Its fifth and sixth columns show the expenditure share of products on sale and the expenditure-weighted average size of sales. Both moments are very similar in the two countries. The share of sales are 17 percent in both countries and the size of sales are 12 and 13 percent in Germany and Italy, respectively.

To assess how changes in sales contributed to inflation, we measure the annual change in sales frequency as

$$\Delta\xi_w = \xi_w - \xi_{w-52} = \sum_{ps} \gamma_{psw} I^s_{psw} - \sum_{ps} \gamma_{psw} I^s_{psw-52},$$

where γ_{psw} are annual Tornqvist weights defined in equation (13) and I^s_{psw} is an indicator function that takes the value 1 in case product p in store s is on sale in week w, i.e., the posted price is strictly below the reference price ($P_{psw} < P^f_{psw}$), and 0 otherwise. Additionally, we measure the annual change in the average size of sales as the 52-week difference between the average percentage distance between the reference and the posted prices among products

[38] The correlations between the series based on 5-week and 13-week reference price filters are 0.78 for the frequency and 0.76 for the size of sales.

on sale. Formally,

$$\Delta\psi_w^s = \psi_w^s - \psi_{w-52}^s = \frac{\sum_{ps}\gamma_{psw}I_{psw}^s\left(\log P_{psw}^f - \log P_{psw}\right)}{\sum_{ps}\gamma_{psw}I_{psw}^s} - \frac{\sum_{ps}\gamma_{psw}I_{psw-52}^s\left(\log P_{psw-52}^f - \log P_{psw-52}\right)}{\sum_{ps}\gamma_{psw}I_{psw-52}^s} \quad (14)$$

The contribution of changes in sales-related price setting can be expressed as a 'sales inflation,' formally defined as

$$\pi_w^s = -\left(\xi_w^s\psi_w^s - \xi_{w-52}^s\psi_{w-52}^s\right). \quad (15)$$

Fewer and smaller sales in the current week relative to the base period necessarily *increases* inflation, which explains the negative sign on the right-hand side of the expression.

Figure 14 shows the annual change in the frequency and the size of sales in Germany and Italy.[39] The figures show that the frequency and the size of the sales started below their long-term average already during the early weeks of the pandemic. This suggests that retailers promptly responded to the elevated demand during the stock-up shock by reducing both the frequency and the magnitude of their sales. Furthermore, both the frequency and the size of sales declined gradually further during the quarter in both Germany and Italy. The decline in the frequency and the size of sales contributed to the increase in inflation over our sample by 1.4 percentage points in Germany and 0.6 percentage points in Italy.[40] These results indicate that the retailers actively adjusted their temporary sales to respond to the large demand shock. This is broadly in line with the evidence in Gautier et al. (2022a) using euro area CPI microdata, which finds significant sales-inflation response to large (e.g., their global demand shock) aggregate shocks.

7.5 Reference-price inflation

In this section, we turn our attention towards the evolution of (sales-filtered) reference prices. The more flexible inflation response to the Covid-19 shock in Italy relative to Germany is explained predominantly by the differences in reference-price inflation. As Figure 15 shows, the increase in reference price inflation in Germany (0.54%) was only around one-third of that in Italy (1.65%), albeit from a higher initial level.

[39]The figures exclude the first and the last two weeks of the sample because it is particularly difficult to estimate both reference prices and sales so close to the endpoints.

[40]The change in sales-related inflation is smaller in Italy than in Germany primarily because it is already above its long-term average in mid February, possibly already as a response of the ongoing stock-up shock.

The more flexible response in Italy is consistent with the structurally more frequent changes in reference prices between 2013-2017 reported in Table 7. The frequency of reference price changes is a relevant statistic determining the flexibility of the price level in most price-setting models (Calvo, 1983; Alvarez et al., 2020). The table shows that 48% of the reference price change in Germany annually, but much more, 68% change in Italy. The difference between the repricing frequencies stays robust if we restrict attention to a subsample of goods that are sold in both countries, or if we measure reference-price changes at the monthly frequency (see Table 3). The less frequent price changes in Germany might be partly related to differences in competitive environment with fewer and larger retailers in Germany (16 chains in our sample) than in Italy (466 chains).

Figure 16 shows the evolution of the annual frequency of price changes during the Covid-19 shock. It reveals no significant change in the overall frequency of price changes in either Germany or Italy as a response to the large increase in demand, and, if anything, the aggregate frequency declined in both countries. The lack of an increase in aggregate frequency indicates that even the sizable Covid-19 shock was insufficient to trigger sizable state-dependent adjustment on the (net) extensive margin.[41] The constant aggregate frequency masks an apparent shift away from price decreases towards price increases in both countries.

8 Conclusion

The paper contrasted price setting in the euro area and the US using a novel supermarket-scanner dataset in four euro area countries and in the US. It has found that the higher flexibility of food inflation in the US is driven both by the higher frequency of repricing and the stronger state dependence of price changes. It argues that the driving force behind both factors is a more volatile product-level environment in the US.

Our conclusions have implications for both model selection and policy. First, the evidence is in line with models with sizable nominal rigidities in both regions, amplifying the impact of monetary and fiscal policy on the real economy. The higher nominal rigidities in the euro area imply that, at least in the food sector, changes in nominal expenditure growth have a smaller impact on inflation and a larger impact on quantities than in the US. Second, the evidence presented in the paper supports state dependence in price setting. Even though we find that the estimated magnitude of state dependence has a mild impact on price flexibility in response

[41] We have to keep in mind, though, that remaining sales-related price changes, coming from the 5-week window, as opposed to a more standard 13-week window in our implemented filter, might bias our estimates downward.

to small aggregate shocks, state dependence necessarily implies that prices endogenously become more flexible after large aggregate shocks and higher trend inflation (Karadi and Reiff, 2019; Alvarez et al., 2019; Costain et al., 2022). Third, the sizable differences in the implied product-level volatility between the US and the euro area raise important questions for future research. Although in the simplest class of price-setting models, product-level volatility matters only as far as it influences frequency and state dependence (Alvarez et al., 2020), in more complicated models, it can have an independent impact on price flexibility, as high product-level volatility can make retailers limit their attention to aggregate fluctuations (Mackowiak and Wiederholt, 2009), which could mitigate their responsiveness to aggregate shocks. Its key role in driving differences across regions also highlights the importance of further research to understand better the underlying sources of the product-level volatility.

References

Alvarez, Fernando, Martin Beraja, Martin Gonzalez-Rozada, and Pablo Andrés Neumeyer (2019) "From Hyperinflation to Stable Prices: Argentinas Evidence on Menu Cost Models," *The Quarterly Journal of Economics*, Vol. 134, pp. 451–505.

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi (2014) "Small and Large Price Changes and the Propagation of Monetary Shocks," Working Paper 20155, National Bureau of Economic Research.

Alvarez, Fernando E, Katarína Borovičková, and Robert Shimer (2021) "Consistent Evidence on Duration Dependence of Price Changes,"Technical report, National Bureau of Economic Research.

Alvarez, Fernando E., Francesco Lippi, and Aleksei Oskolkov (2020) "The Macroeconomics of Sticky Prices with Generalized Hazard Functions," NBER Working Papers 27434, National Bureau of Economic Research, Inc.

Anderson, Eric, Benjamin A Malin, Emi Nakamura, Duncan Simester, and Jón Steinsson (2017) "Informational Rigidities and the Stickiness of Temporary Sales," *Journal of Monetary Economics*, Vol. 90, pp. 64–83.

Argente, David and Chen Yeh, "Product Life Cycle, Learning, and Nominal Shocks,"Technical report.

Bils, Mark and Peter J. Klenow (2004) "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, Vol. 112, pp. 947–985.

Broda, Christian and David E Weinstein (2010) "Product Creation and Destruction: Evidence and Price Implications," *American Economic Review*, Vol. 100, pp. 691–723.

Caballero, Ricardo J and Eduardo MRA Engel (2007) "Price Stickiness in Ss models: New Interpretations of Old Results," *Journal of Monetary Economics*, Vol. 54, pp. 100–121.

Calvo, Guillermo A. (1983) "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, Vol. 12, pp. 383 – 398.

Campbell, Jeffrey R and Benjamin Eden (2014) "Rigid Prices: Evidence from US Scanner Data," *International Economic Review*, Vol. 55, pp. 423–442.

Caplin, Andrew S. and Daniel F. Spulber (1987) "Menu Costs and the Neutrality of Money," *The Quarterly Journal of Economics*, Vol. 102, pp. 703–726.

Costain, James and Anton Nakov (2011) "Distributional Dynamics under Smoothly State-Dependent Pricing," *Journal of Monetary Economics*, Vol. 58, pp. 646 – 665.

Costain, James, Anton Nakov, and Borja Petit (2022) "Flattening of the Phillips Curve with State-Dependent Prices and Wages," *The Economic Journal*, Vol. 132, pp. 546–581.

DellaVigna, Stefano and Matthew Gentzkow (2019) "Uniform Pricing in US Retail Chains," *The Quarterly Journal of Economics*, Vol. 134, pp. 2011–2084.

Dhyne, Emmanuel, Luis J Alvarez, Hervé Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffmann, Nicole Jonker, Patrick Lunnemann, Fabio Rumler, and Jouko Vilmunen (2006) "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data," *Journal of Economic Perspectives*, Vol. 20, pp. 171–192.

Dotsey, Michael, Robert G. King, and Alexander L. Wolman (1999) "State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output," *The Quarterly Journal of Economics*, Vol. 114, pp. 655–690.

DAcunto, Francesco, Ulrike Malmendier, Juan Ospina, and Michael Weber (2021) "Exposure to Grocery Prices and Inflation Expectations," *Journal of Political Economy*, Vol. 129, pp. 1615–1639.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo (2011) "Reference Prices, Costs, and Nominal Rigidities," *American Economic Review*, Vol. 101, pp. 234–62.

——— (2014) "How Frequent Are Small Price Changes?" *American Economic Journal: Macroeconomics*, Vol. 6, pp. 137–155.

Feenstra, Robert C (1994) "New Product Varieties and the Measurement of International Prices," *American Economic Review*, pp. 157–177.

Feenstra, Robert and Matthew Shapiro eds. (2002) *Scanner Data and Price Indexes*: University of Chicago Press.

Fougère, Denis, Hervé Le Bihan, and Patrick Sevestre (2007) "Heterogeneity in Consumer Price Stickiness: A Microeconometric Investigation," *Journal of Business & Economic Statistics*, Vol. 25, pp. 247–264.

Gagnon, Etienne and David López-Salido (2020) "Small Price Responses to Large Demand Shocks," *Journal of the European Economic Association*, Vol. 18, pp. 792–828.

Gagnon, Etienne, David López-Salido, and Nicolas Vincent (2012) "Individual Price Adjustment along the Extensive Margin," *NBER Macroeconomics Annual 2012, Volume 27*, pp. 235–281.

Gautier, Erwan, Cristina Conflitti, Riemer P Faber, Brian Fabo, Ludmila Fadejeva, Valentin Jouvanceau, Jan-Oliver Menz, Teresa Messner, Pavlos Petroulas, Pau Roldan-Blanco, Fabio Rumler, Sergio Santoro, Elisabeth Wieland, and Helene Zimmer (2022a) "New Facts on Consumer Price Rigidity in the Euro Area," Working Paper 2669, European Central Bank.

Gautier, Erwan, Magali Marx, and Paul Vertier (2022b) "The Transmission of Nominal Shocks when Prices are Sticky," unpublished manuscript.

Golosov, Mikhail and Robert E. Lucas (2007) "Menu Costs and Phillips Curves," *Journal of Political Economy*, Vol. 115, pp. 171–199.

Ivancic, Lorraine, W Erwin Diewert, and Kevin J Fox (2011) "Scanner data, time aggregation and the construction of price indexes," *Journal of Econometrics*, Vol. 161, pp. 24–35.

Karadi, Peter and Adam Reiff (2019) "Menu Costs, Aggregate Fluctuations, and Large Shocks," *American Economic Journal: Macroeconomics*, Vol. 11, pp. 111–46.

Karadi, Peter, Raphael Schoenle, and Jesse Wursten (2020) "Measuring Price Selection in Microdata-It's Not There,"Technical report, CEPR Discussion Papers.

Kehoe, Patrick and Virgiliu Midrigan (2015) "Prices are Sticky After All," *Journal of Monetary Economics*, Vol. 75, pp. 35–53.

Klenow, Peter J. and Benjamin A. Malin (2010) "Microeconomic Evidence on Price Setting," Vol. 3 of Handbook of Monetary Economics: Elsevier, pp. 231 – 284.

Kryvtsov, Oleksiy and Nicolas Vincent (2021) “The Cyclicality of Sales and Aggregate Price Flexibility,” *The Review of Economic Studies*, Vol. 88, pp. 334–377.

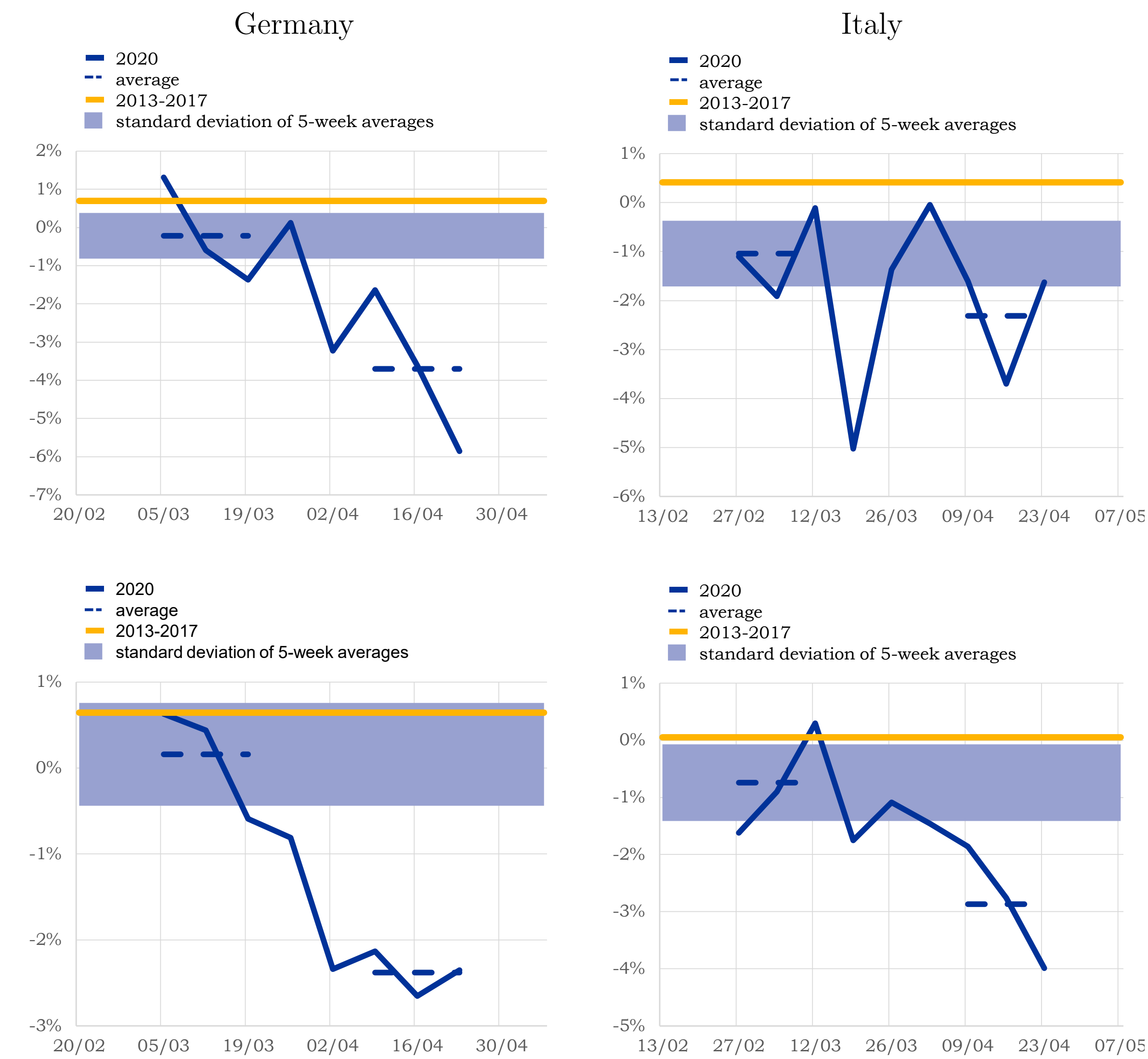
Mackowiak, Bartosz and Mirko Wiederholt (2009) “Optimal Sticky Prices under Rational Inattention,” *The American Economic Review*, Vol. 99, pp. 769–803.

Mankiw, N Gregory (1985) “Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly,” *The Quarterly Journal of Economics*, Vol. 100, pp. 529–537.

Nakamura, Emi and Jón Steinsson (2008) “Five Facts about Prices: A Reevaluation of Menu Cost Models,” *The Quarterly Journal of Economics*, Vol. 123, pp. 1415–1464.

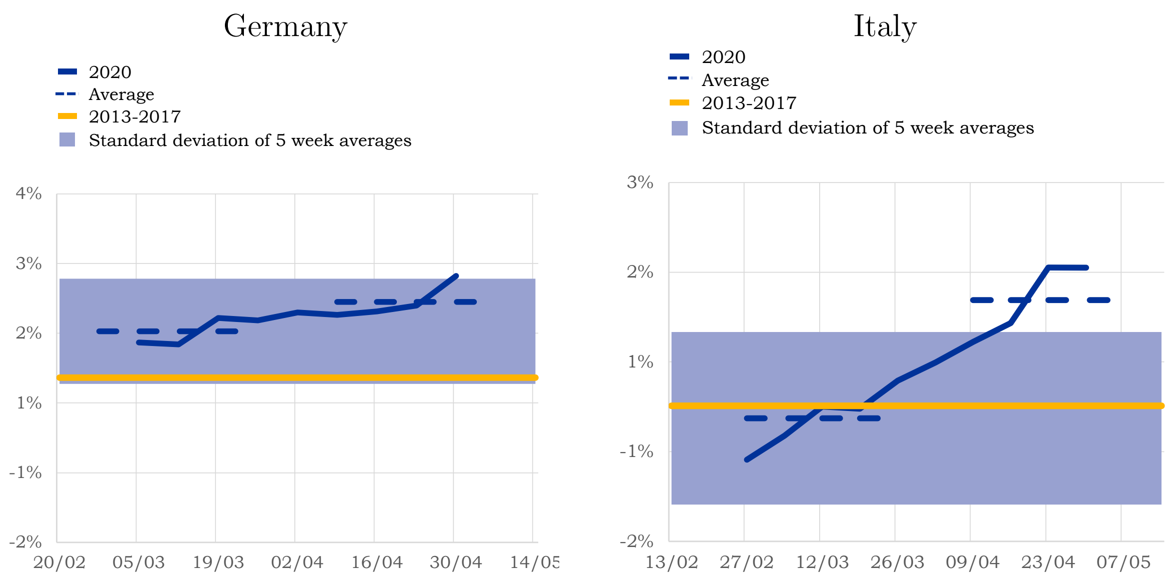
Woodford, Michael (2009) “Information-Constrained State-Dependent Pricing,” *Journal of Monetary Economics*, Vol. 56, pp. S100–S124.

Figure 14: Annual change in frequency and size of sales during the first wave of the Covid-19 pandemic



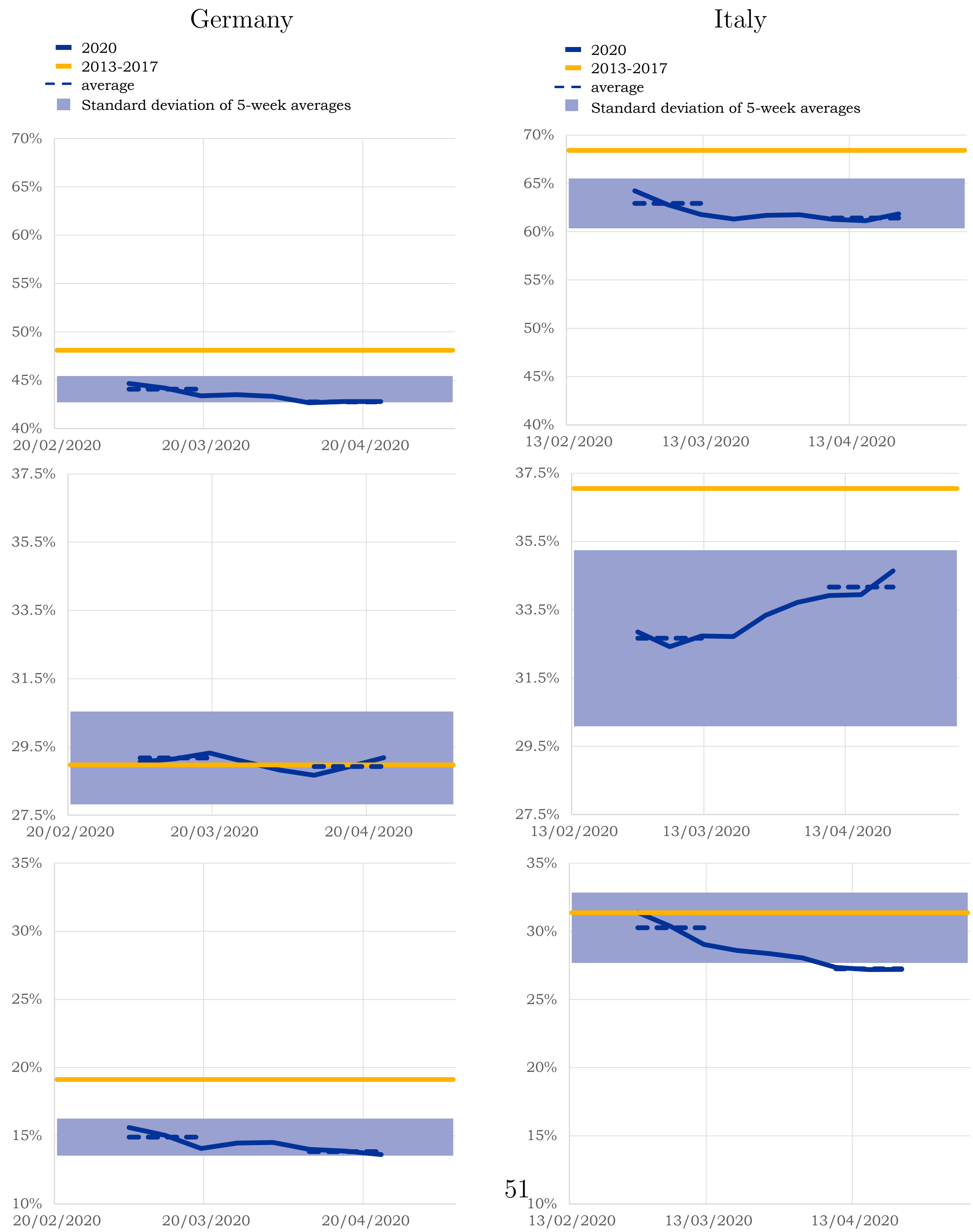
Note: The figure shows the weekly evolution of the annual change in the frequency (top row) and size (bottom row) of temporary sales between mid-February and mid-May in 2020 (blue solid line) in Germany (left column) and Italy (right column). It shows that both the average frequency and the average size of sales in the first three weeks (blue-dashed line) started out below their long-term average measured during the first two quarters of 2013-2017 (yellow dashed lines). Over the course of the quarter, both the sales frequency and the sales size declined markedly both in Germany and Italy, exceeding the +- one-standard-deviation bands in both countries.

Figure 15: Sales-filtered (reference-price) inflation during the first wave of the Covid-19 pandemic, year-on-year



Note: The figure shows the weekly, year-on-year reference-price inflation (blue solid line) between mid-February and mid-May in 2020 in Germany and Italy. It shows that the increase in the average 5-week-inflation over the quarter was smaller (0.54%) and within a +- one-standard-deviation band in Germany, while it was three times as large in Italy (1.65%) and clearly exceeded the standard-deviation band.

Figure 16: Reference-price changes, increases and decreases during the first wave of the Covid-19 pandemic, year-on-year



Note: The figure shows the weekly share of reference price changes between mid-February and mid-May in 2020 in Germany and Italy and decomposes it to the share of increases and decreases. The figure shows that the overall frequency declined, because the more frequent price increases could not offset the impact of the less frequent price decreases.

A Features of the baseline inflation index

The construction of our baseline index resembles that of the HICP and CPI, which can also be characterized as chained, annual-expenditure-weighted price indices. One difference is that the weights are contemporaneous in our index, while HICP and CPI relies on lagged expenditures. The advantage of setting contemporaneous weights is that we do not need to restrict our analysis to products that exist also in the preceding year. This is a relevant advantage in the fast-moving supermarket-product category, where there is a sizable turnover between products. (Linking closely related products with distinct barcodes over time is out of the scope of this paper, while statistical offices put substantial effort to regularly replace exiting with entering products after suitable quality adjustment.)

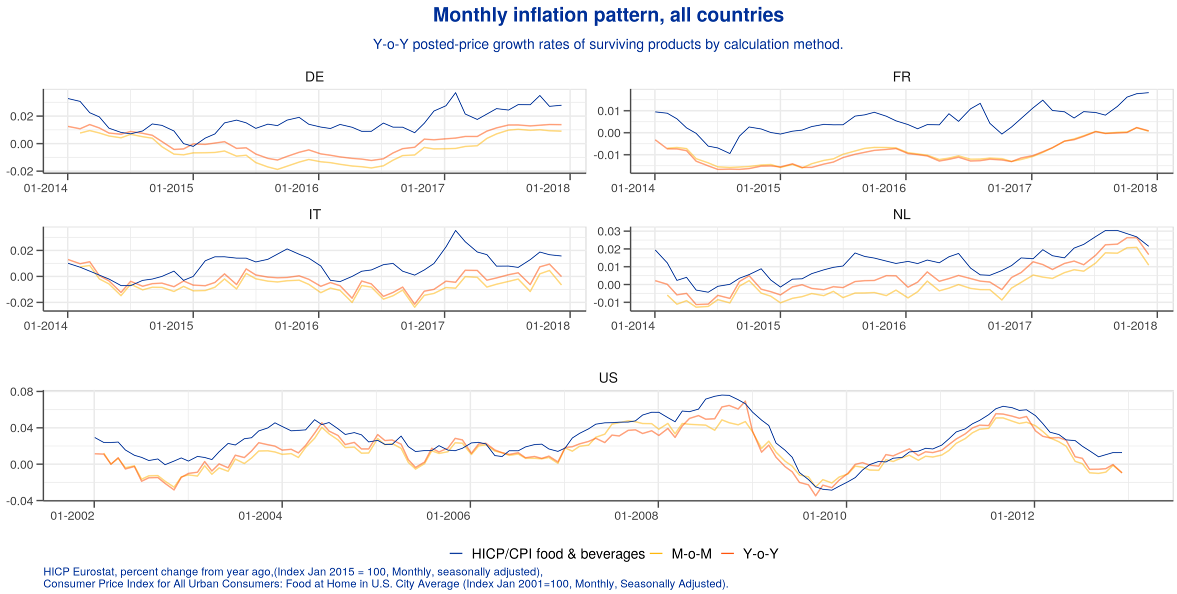
Annual weighting has multiple advantages over schemes that use more frequent weighting. First, because our focus is on price setting (as opposed to the measurement of the welfare-relevant inflation), annual weights minimize the impact of high-frequency quantity changes on the measurement of price dynamics, while still correctly tracking trend changes in the relative importance of different products. For example, it mitigates the seasonality of the index that more frequent weighting schemes would cause. This is particularly relevant among supermarket goods, where seasonal sales generate large seasonal variation in expenditures. Second, annual weighting mitigates a bias called the chain-drift, which impacts chained indices with high-frequency weighting schemes. Chain drift is present if the index does not return to 1 when the price returns to its initial level. One cause of the chain drift is the inventory behaviour of households, who stock up during temporary promotions. As a result, the quantity of purchases declines below its initial level after it increased during the promotions. In the presence of such dynamic behavior, even superlative chained indices (e.g., Tornqvist[42]) with high-frequency weights measure deflation even though the price returned to its initial level (Ivancic et al., 2011). Lower-frequency weights mitigate the impact of the chain drift by taking into account the longer planning horizon of the households (Feenstra and Shapiro, 2002), and bringing the index closer to fixed-base indexes, which are free of chain drift. To assess the remaining impact on chain drift on our index, we compare our index with unchained year-on-year price index of existing products.

The results indicate that even annual weighting is not sufficient to completely eliminate chain drift from our baseline inflation measure. At the same time, the close to one correlation between the unchained and the chained series shows that the chain drift has insignificant impact on the variation of measured inflation at business cycle frequencies, which is the main

[42]The Tornqvist index between month t and $t-1$ equals to $\Pi_t = \prod_{ps} \left(\frac{P_{pst}}{P_{pst-1}}\right)^{(\omega_{pst}+\omega_{pst-1})/2}$, where ω_{pst} is the expenditure share of product p in store s in month t.

focus of our analysis.

Figure 17: Chain drift: the year-on-year change of the baseline chained inflation index and the unchained 12-month inflation index.



Note: The figure compares the year-on-year change of the baseline chained inflation index and the unchained 12-month inflation index. The level difference between the series indicates that our baseline inflation index suffers from some remaining chain drift, however, the close comovement of the series show that the chain drift has insignificant impact on the fluctuations of the series.

Table 8: Chain drift: Average inflation of and correlation between the year-on-year change of the baseline chained inflation index and the unchained 12-month inflation index.

Series	moment	**DE**	**FR**	**IT**	**NL**	**US**
Inflation - chained 1m	average	-0.40	-0.92	-0.71	-0.10	1.47
Inflation - unchained 12m	average	0.13	-0.97	-0.31	0.39	1.67
1m-12m inflation	correlation	0.99	0.99	0.99	0.98	0.96

B New product introductions

A feature of the inflation index described by equation (**??**) is that it includes the price change of existing products, but excludes the impact of product entry and exit on the price level. We justify our focus on the price-setting of existing products by showing below that it is this component of the price level that is mostly responsible for its responsiveness to business-cycle fluctuations. The component of inflation caused by product entry and exit, which we call new-product inflation, in contrast, is broadly stable and insensitive to business-cycle fluctuations. Notably, this letter component, through the sizable difference between the price of the entering and exiting products, is what is responsible for the level of inflation.

The role of new-product inflation is not the focus of this paper. However, we find it useful to show some indicative evidence on its behaviour by comparing the evolution of an 'all-product' inflation index to our baseline existing-product index. We measure the former as the monthly change in the price levels defined as the geometric average of prices weighted by the annual expenditure weights. Formally,

$$\Pi_t^{ap} = \frac{\prod_{ps} P_{pst}^{\omega_{pst}}}{\prod_{ps} P_{pst-1}^{\omega_{pst-1}}}. \tag{16}$$

where the weights ω_{pst} are annual expenditure-share of items that are present in month t. [43] The key difference between the existing-product and the all-product inflations is in the set of products they consider. The existing-product inflation only includes products that exist in both periods (the weights $\omega_{pst-1,t}$ are positive only for products that exist both in months $t-1$ and t), the all-product price levels include all existing prices (the weights ω_{pst} are positive for all products existing in period t), therefore, the all-product inflation is taking account the price differences between exiting and entering products.

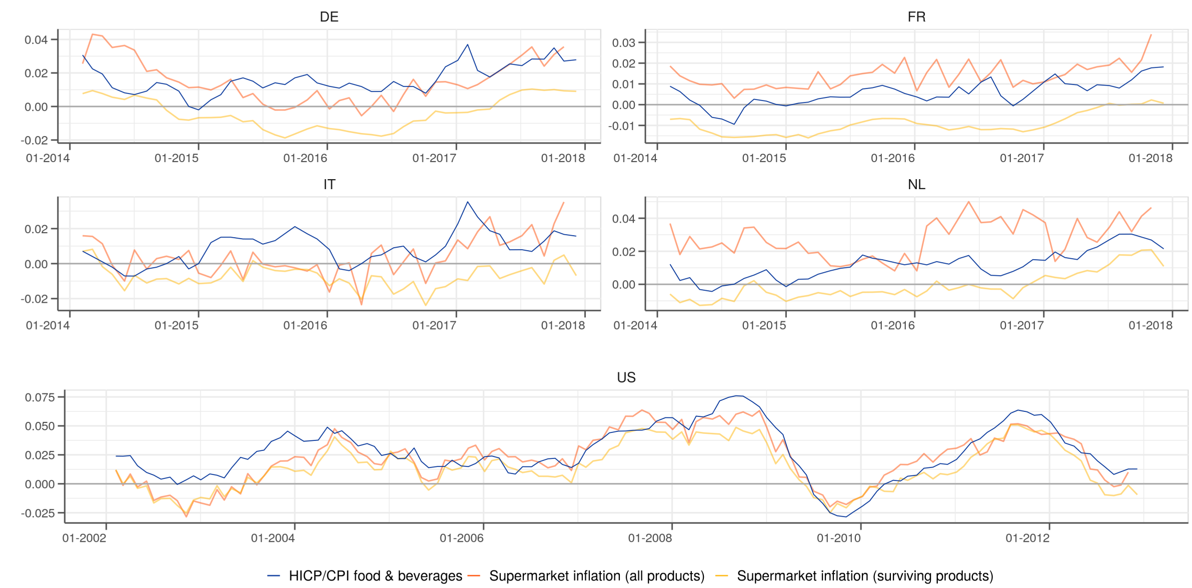
The all-product inflation is the relevant inflation index, if the exiting products are all replaced by similar quality entering products. This is admittedly a strong assumption. There are both 'true' product exits as well as 'true' product introductions without a matching entry or exit respectively in our dataset. But the replacements constitute arguably the majority of exits and entries.[44] These could include pure barcode changes (when the exact same product is reintroduced with a different barcode), changes in packaging, in volume or minor flavor/color/form upgrades. The true entries and the replacements require different treatment

[43]Formally, the weights are given by $\omega_{pst} = I_{pst}\omega_{psy} / \sum_{ps} I_{pst}\omega_{psy}$, where I_{pst} is an indicator function that takes a value 1 if product p in store s is sold in positive quantities.

[44]Argente and Yeh (2199) using the US IRi Academic Dataset find, consistently with our statement, that 'product line extensions, such as flavor or form upgrades or novelty and seasonal items, are much more prevalent than the introduction of new brands.'

by the price index. With some reasonable assumptions about the utility function (including how quality affects demand), Feenstra (1994) and Broda and Weinstein (2010) show how the quality of true introductions can be assessed by relying on their market share relative to existing products. Intuitively, if their quality is high relative to their price (which is observable) their relative market share (which is also observable) is also going to be high. The same techniques, however, are not applicable in case a producer, which influences the supply of both the old and the new products, replaces an old product with a new. In this case the market share is influenced by the producer's choices and is not informative about the quality of the new product. There are arguably a lot of product replacement among supermarket goods, where what is changing is the packaging and the price, but not the quality of the good. In these cases the all-product inflation is the valid index.

Figure 18: New product introductions: The year-on-year change of the baseline inflation index, the all-product index and the HICP/CPI subindexes.



Note:

Figures 18 presents the all-product inflation rates for all countries together with the relevant official inflation subindex. Table 9 shows the average year-on-year inflation rates. The figures/tables show that the all-product inflation rates gets closer to the level of the HICP inflation than our baseline index and it captures the variation of the official indices well across different countries. This suggests that the all-product inflation, despite its simplicity, captures most of the information inherent in official price indices, which are based on much more

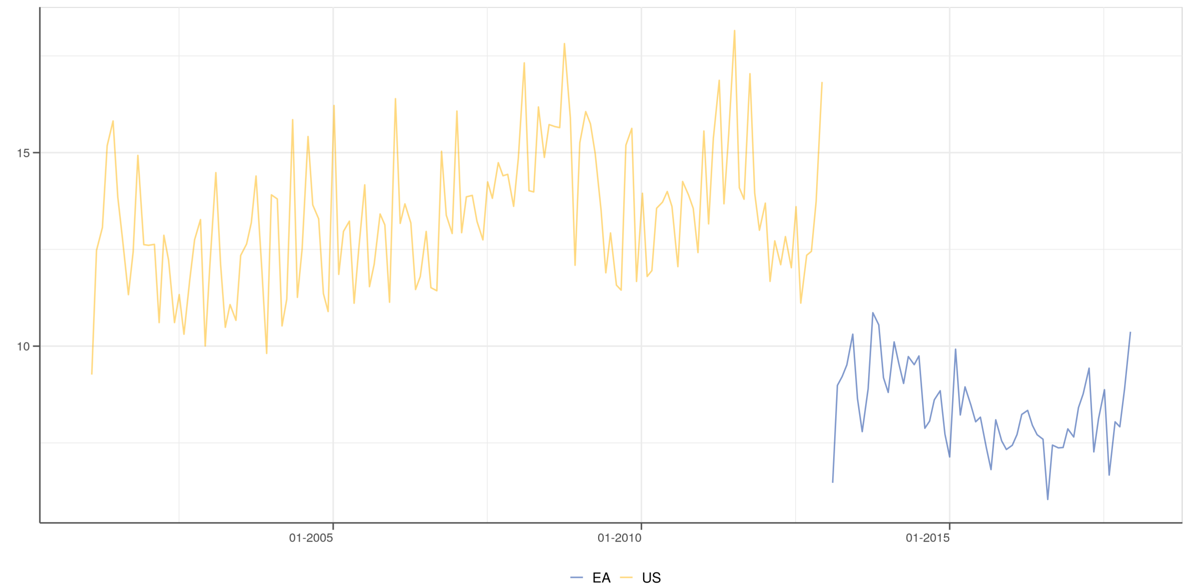
Table 9: Level and comovement of key inflation indexes

Series	moment	**DE**	**FR**	**IT**	**NL**	**US**
All-product inflation	average	1.51	1.39	0.37	2.78	2.25
Surviving-product inflation	average	-0.40	-0.92	-0.71	-0.10	1.47
Official food and beverage inflation	average	1.63	0.56	0.91	1.21	2.72
All-Surviving inflation	correlation	0.93	0.74	0.71	0.52	0.96
Official-Surviving inflation	correlation	0.49	0.71	0.32	0.87	0.89

careful judgement of product replacement and quality adjustment.

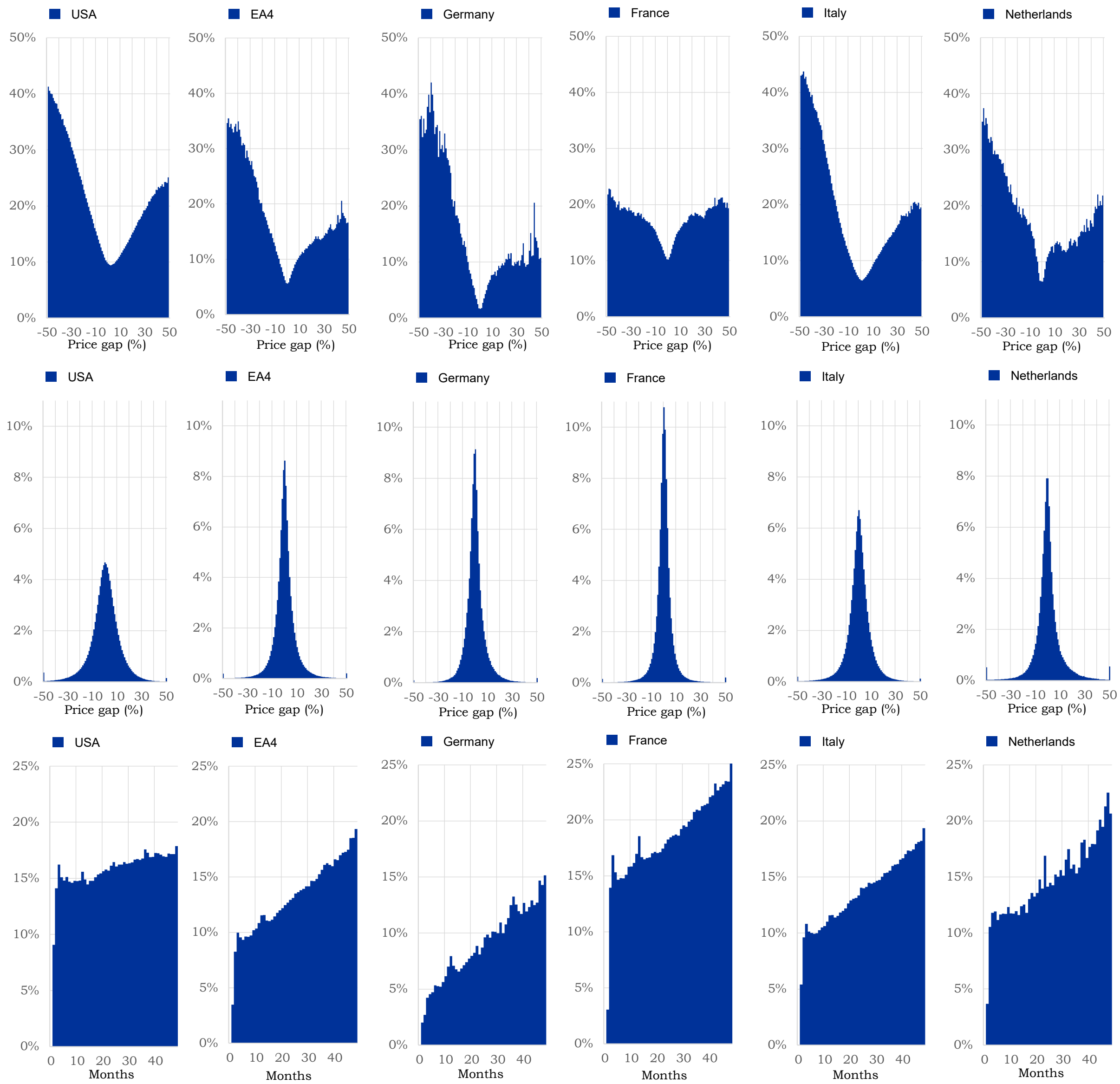
C Time variation of EA4 and US frequency

Figure 19: Frequency of reference-price changes, EA4 vs US



Note: The shows the evolution of the frequency of reference-price changes in the US (2001-2012) and in EA4 (2013-2017). The figure shows that the frequency is robustly lower, implying higher price rigidity in EA4 relative to the US.

Figure 20: Generalized hazard functions and price-gap densities and duration hazard functions across euro area countries and the US



Note: The figure shows the generalized hazard functions (first row), the price-gap densities (second row) and the duration hazard functions (third row) across four euro area countries and the US. The figures show evidence for moderate state dependence in all countries, with notable heterogeneity across countries.