Measuring the impact of retailers' strategies on food price inflation using scanner data

Castellari E.[1,2], Moro D. [1], Platoni S.[3] and Sckokai P. [1]

[1]Dipartimento di Economia Agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza, Italy
[2] Department of Agricultural and Resource Economics, University of Connecticut, Storrs, US
[3] Dipartimento di Scienze Economiche e Sociali, Università Cattolica del Sacro Cuore, Piacenza

Authors' contacts:
elena.castellari@unicatt.it ; daniele.moro@unicatt.it ;
silvia.platoni@unicatt.it ; paolo.sckokai@unicatt.it

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013.

Measuring the impact of retailers' strategies on food price inflation using scanner data

1. INTRODUCTION

According to Eurostat, after a period of relatively stable consumer prices during the 90's, at the beginning of the last decade the European Union has been affected by a price inflation of about 2% per year. In 2008 the price inflation sharply increased to 3.7%. In 2009, following a decline in food prices, the inflation rate stood at 1% to accelerate again its increase in 2010 (2%) and 2011 (3.1%), while recovering to 2.2% at the end of 2012.

Retailers' strategies may play a central role in determining the food prices faced by consumers. Through store and chain marketing strategies retailers are able to somehow influence consumer decisions when choosing among different alternatives. For instance, Chevalier and Kashyap (2011) showed how retailers use promotion strategies to charge consumers different prices influencing their reservation price: "Thus, how consumers update reservation prices for individual goods becomes a critical factor affecting inflation" (Chevalier and Kashyap, 2011). Broda et al. (2009) using household scanner data showed that "poor households systematically pay less than rich households for identical goods". Different explanations why a richer person might pay more can be advanced. First poor people might be willing to invest more time in comparing prices among stores looking for the one which offers deeper discounts. Second, stores in richer neighborhoods might face higher rent costs (Broda et al., 2009), reflecting in higher prices. Last, different stores' characteristics might be a source of differentiation between two goods that otherwise would be identical, meaning the same good may have different value if purchased at different stores because of the related shopping experience (Betancourt and Gautschi, 1992 and 1993; Broda et al., 2009). Although the marketing literature that analyzes the consumer response to marketing strategies employed by retailers is rich, at our knowledge, no study yet has analyzed their effect on food price inflation rates.

Inflation is measured by Central Statistical Agencies, and consumer price indexes (CPI) are computed using collected prices; the computation of the standard CPI does not account for the potential effect of retailers' marketing strategies, that is, prices are not normally identified as 'regular' or 'sale' prices (Nakamura et al., 2011). Price dynamics may produce a bias in the computed CPI, where, according to the definition provided in Nakamura et al. (2011), a bias occurs when the expected value of the price index formula differs from the target index. Silver and Heravi (2001) suggested two main forms of bias in consumer price indexes. First, the substitution bias driven by "the inability of fixed basket Laspeyres/Paasche-type indexes to take into account decreases in the relative expenditure or weight given to goods and services with relatively high price increases, as consumers substitute away from them". Secondly the CPIs do not "properly account for changes in the quality of what we purchase" (Silver and Heravi, 2001). CPIs, as commonly computed, are not cost of living indexes; Boskin et al. (1998) highlighted the challenges of measuring a cost of living index given the considerable dynamics of a modern economy where the number of products is extremely high and their entering and exiting the market might be considerably rapid. In particular the introduction of new products and the instability of the product basket have been found to be a considerable source of bias in the CPI calculation using fixed basket indexes. The Authors suggested the CPI should "abandon the Laspeyres formula and move towards a cost-of-living concept by adopting a "superlative" index formula to account for changing market basket" (Boskin et al., 1998). To this end, the use of retailers' high frequency scanner data allows to first incorporate the actual consumer purchasing behavior in the computation of the CPI by periodically updating the basket of goods. Second, the availability of prices and quantities of all goods allows the construction of superlative weighted price indexes (de Haan and van der Grient, 2011). At the same time, some potential negative implications arise from using scanner data for CPI computation: in particular, the high volatility of prices and quantity due to retailers' sales would generate drifts in the CPI estimation producing a "price and quantity bouncing" bias (de Haan and van der Grient, 2011).

The most recent literature has focused on analyzing different approaches to the computation of price indexes and on establishing the effect of time and store aggregation as well as the drift bias on inflation measurement. Ivanic et al. (2011) showed that the level of data aggregation across time and points of sale becomes relevant when high frequency scanner data are used to estimate price changes through the computation of the CPI. Evidence of a "price bouncing" bias, when estimating price indexes, has also been found by de Haan and van der Grient (2011) using Dutch data. Similarly, Nakamura et al. (2010), using US scanner data, compared price indexes computed using either all prices or only "regular prices", i.e. excluding 'sale prices', and confirmed the insurgence of a chain drift problem. Their suggestion was that "averaging within chains will ameliorate the chain drift problem", although it cannot be the sole solution. A more promising approach is to resort to drift-free multilateral indexes: Ivanic et al. (2011), proposed the use of the Gini-Eltetö-Köves-Szulc (GEKS) index. They showed how the conventional superlative indexes, even calculated at a level of aggregation that seems to minimize the drift bias, "show a troubling degree of volatility when high-frequency data are used" (Ivanic et al., 2011). Differently, the GEKS index provides drift-free estimates. In their empirical tests, de Haan and van der Grient (2011) confirmed the superiority of GEKS indexes with respect to the Dutch method, based on monthly-chained Jevons indexes, when dealing with supermarket scanner data.

This work mainly focuses on determining the effect of some retailers' strategies, such as promotion, assortment and the presence of Private Label (PL), on food inflation. In particular, focusing on the Italian dairy market, we want to identify if some observable retailers' strategies may affect inflation rates. Moreover, it is of our interest to test if retailing chains and types of store play a role in influencing food inflation. To do so, we employ high frequency scanner data from different retail chains, concerning seven different dairy product categories from 400 points of sales during 156 weeks from January 2009 to January

2012. Points of sales are all located in Italy and belong to fourteen different retailing chains. Each of the fourteen chains can have different types of store (hypermarkets, supermarkets and superettes) for a total of 33 chain-format combinations. We also resort to the drift-free CPI estimate proposed by Ivanic et al. (2011) to measure price dynamics.

After computing the GEKS index for each product and chain-format combination along the 156 weeks, we adapt a three-way ECM estimator (Davis, 2002) to capture for the unobservables due to chain, type of store and time variation. Moreover, for each product of our dataset, we estimate the effect on CPI due to observed retailers' strategies such as promotional activities, PL presence, retailers' assortment, PL line extension.

Results show that promotional activities and PL market shares have a positive role in restraining price increases. Promotional activities on national brands (NB) seem to be the most effective in moderating the price rise compared to promotions on PL. Moreover, a high PL presence in different market segments for the same product categories reflects a relatively higher price index. The effect due to the intensity of assortment is mixed depending on the product category analyzed.

Our paper proceeds as follow. Section 2 provides a description of the data and of the computation of the variables used in the econometric model. Section 3 describes the price index and its computation while giving some descriptive statistics. Section 4 describes the econometric technique and the model specification. Results and conclusions follow in section 5 and 6.

2. DATA DESCRIPTION

We use the SymphonyIRI dataset to compute the price index and to measure retailers' strategies by chain and type of store. Our scanner database provides brand level weekly prices and sales, with and without promotion, for four-hundred points of sales which belong to fourteen different retailing chains along one-hundred and fifty-six weeks from January 2009 to January 2012. All points of sales in the sample are located in Italy although we do not observe any geographical cluster among them. For each point of sales we observe the retail chain it belongs to and the store format among hypermarkets, supermarkets and superettes; discount stores are not included in our sample. Furthermore, in our dataset the retailing chains, manufacturers and brands, beside the indication of PL, are blinded by letters code for confidentiality. In this way we can distinguish different chains, manufacturers and brands among each other, but we are not able to link them to real market entities.

Our data cover seven different dairy product categories: refrigerated and ultra-high temperature (uht) liquid milk, yogurt, cheese[1], mozzarella cheese, fresh cream, and butter. Not observing the Universal Product Code (UPC), we define 'a product' as the interaction of segments, manufacturers, brands and packaging attributes.

[1] Processed spreadable cheese and cheese flakes.

For instance, any product on the category uht milk will be defined by the interaction of four different segments (whole, semi-skimmed, skimmed and vitamin enriched), twenty-seven different manufacturers, one hundred and twenty-two brands and two attributes related to product packaging (Table 1). We define the latter level of interaction as brand unit (BU).

Table 1: Number of Segments, Manufactures, Brands, Products types and Brand units for each product category.

Products	Segments	Manufactures	Brand	Type	Brand Units
Yogurt	15	36	199	11	739
Uht milk	4	27	122	2	261
Refrigerated milk	7	30	139	2	364
Uht Cream	5	33	84	1	121
Ricotta cheese	5	40	70	2	106
Butter	3	45	118	5	251
Cheese	2	23	54	4	65
Mozzarella cheese	4	39	163	7	315

Source: Our elaboration on IRI data

The four-hundred points of sales are distributed among types of store (hypermarkets, supermarkets and superettes) and retailing chains (A, B…P). For each of the thirty-tree different chain and type-of-store combinations we compute a weekly price index using the points of sales of the sample. Details are provided in Table 2. From the dataset we build other variables to proxy the intensity of some marketing strategies at the category level for each point of sales. For instance, taking the assortment strategy, for any of the 33 combinations of retailing chains and types of store we construct the related variable by computing first the strategy at each point of sales, using the number of BU in a given week and product category, and then averaging across the points of sales.

Similarly, we compute the share in value of PL in total sales to measure their market competitiveness with respect to other brands. Furthermore, to capture the PL line extension, we resort to the ratio between the number of market segments where PL are present and the total number of segments for a given product category. For example, within the uht milk, where we observe four different market segments (whole, semi-skimmed, skimmed and vitamin enriched), the presence of PL for a given point of sale on a given week in two of the four market segments will correspond to a PL line extension of 50%. Finally, we use the share in value sold under sales to measure the intensity of promotion activities. We compute all the variables for each of the 33 combinations of retailer chains and types of store along the line sketched above for the computation of the assortment variable.

Table 2. Distribution of the point of sales in the sample among different retailing chains and types of stores.

Retailer Chain	Hyper	Super	Superette	Total
A	-	10	4	14
B	8	8	-	16
C	16	22	4	42
D	12	26	4	42
E	16	29	5	50
F	-	30	5	35
G	10	10	-	20
H	16	35	4	55
I	10	8	-	18
L	-	8	4	12
M	-	8	4	12
N	12	28	6	46
O	-	20	6	26
P	-	8	4	12
Total	**100**	**250**	**50**	**400**

Source: Our elaboration on IRI data

Among the dairy product categories, we observe a lot of variability in terms of product differentiation. For instance, the yogurt category is the most differentiated within the sample with almost seven-hundred and forty different BU, while in other products categories BU range from around three-hundred and sixty for refrigerated milk to sixty-five for cheese (table 1).

Descriptive statistics show that marketing strategies strongly differ among product categories (table 3). The yogurt market is characterized not only by a more intense overall product differentiation but also by a high variability across weeks, chains and stores, with a standard deviation of more than forty BU; uht milk and mozzarella cheese follow the yogurt category in the degree of intensity of assortment strategies, with an average number of BU for chain and type of store higher than twenty, while the remaining categories have an average assortment ranging from twelve to fifteen units.

Significant differences across product categories can be found also for the PL share and its product extension. Butter is the category with the higher PL share (28.2%) followed by cream (20.6%), mozzarella cheese (19.3%), uht milk (16.3%) and refrigerated milk (11.0%), while yogurt and cheese have an average PL share under 10% (respectively 8.6% and 2.6%). The product extension is around 30% for cheese, refrigerated and uht milk; it reaches around 40% in butter, mozzarella cheese and uht cream, and 58% in yogurt.

Table 3. Descriptive Statistics.

	Mean	Std Dev	Minimum	Maximum
		Butter		
GEKS-Price	7.124	0.770	4.708	10.779
Assortment	15.273	7.743	2.800	40.061
Share PL	0.282	0.145	0.007	0.706
PL line extension	0.399	0.170	0.000	1.000
Promotion	0.225	0.132	0.000	0.897
PL Promotion	0.172	0.231	0.000	1.000
		Cheese		
GEKS-Price	9.058	1.192	5.876	14.134
Assortment	12.575	6.947	2.556	37.067
Share PL	0.026	0.042	0.000	0.292
PL line extension	0.297	0.246	0.000	0.500
Promotion	0.324	0.169	0.000	0.868
PL Promotion	0.151	0.290	0.000	1.000
		Milk		
GEKS-Price	1.427	0.109	1.064	1.684
Assortment	15.846	5.034	6.280	28.738
Share PL	0.110	0.113	0.000	0.467
PL line extension	0.283	0.193	0.000	0.571
Promotion	0.052	0.079	0.000	0.696
PL Promotion	0.118	0.231	0.000	1.000
		Mozzarella cheese		
GEKS-Price	7.704	0.832	5.220	11.080
Assortment	23.540	13.522	1.500	57.381
Share PL	0.193	0.124	0.000	0.635
PL line extension	0.412	0.152	0.000	0.500
Promotion	0.309	0.154	0.000	0.761
PL Promotion	0.231	0.239	0.000	1.000
		Cream		
GEKS-Price	4.476	0.527	1.947	6.452
Assortment	13.147	5.559	4.500	28.146
Share PL	0.206	0.091	0.000	0.601
PL line extension	0.399	0.192	0.000	1.000
Promotion	0.197	0.125	0.000	0.819
PL Promotion	0.188	0.264	0.000	1.000
		UHT milk		
GEKS-Price	0.958	0.102	0.673	1.432
Assortment	26.472	8.272	9.872	50.274
Share PL	0.163	0.112	0.000	0.620
PL line extension	0.696	0.160	0.000	1.000
Promotion	0.319	0.143	0.000	0.820
PL Promotion	0.219	0.226	0.000	0.999
		Yogurt		
GEKS-Price	4.215	0.366	3.351	5.447
Assortment	86.668	40.692	29.281	207.093
Share PL	0.086	0.061	0.000	0.350
PL line extension	0.581	0.186	0.000	0.857
Promotion	0.262	0.118	0.000	0.758
PL Promotion	0.216	0.222	0.000	1.000

Source: Our elaboration on IRI info-scan database

The intensity of promotion activities is quite high in almost all categories ranging from 20% to over 30% of the whole value; only refrigerated milk strongly differs from the others with an average value of product sold in merchandising of only 5.2%. Consistently, in refrigerated milk also PL have a more contained promotion strategy with respect to other product categories. However, PL sell under promotion a higher share than the overall category average: in fact the average share of PL sold under promotion reaches 11.8% while the overall average is 5.2%. In the other categories, the percentage of PL sold under promotion ranges from 15.1% in cheese to 23.1% in mozzarella cheese. In all categories, except for refrigerated milk, the share in value of PL sold in merchandising is smaller than the correspondent average of the overall category: this might be explained with the degree of merchandising, suggesting PL promotion activity is in value much lower with respect to promotion on national brands.

3. INDEX COMPUTATION

CPI measures the changes of the price of a basket of goods purchased in a given market during a given time interval. There are several different numeric formulae to compute the CPI and they can be classified in different ways. A first distinction is made if the basket of good entering the index computation is hold constant over time (fixed basket index), or if it changes over time (flexible basket index). A flexible basket approach permits different baskets to enter the index computation over time accounting for the introduction of new products and/or the change in their characteristics. Another distinction is made on the way the base period is updated in the index computation. While chain index updates the base period over time, using a direct index approach the base period is held constant. Laspeyres and Paschee indexes are the most common price indexes used. Being p_{i0} the base period price for item i, p_{it} its price at time t for $\mathrm{t} = 1 \dots \mathrm{T}$, and w_{it} the good i's share of total expenditure at time $\mathrm{t} = 0{,}1 \dots \mathrm{T}$, the fixed basket Laspeyres index can be written as follow:

$$Laspeyres_t = \sum_i w_{i0} \frac{p_{it}}{p_{i0}} \tag{1}$$

while its counterpart, the fixed-basket Paasche index, can be written as:

$$Paasche_t = \left[\sum_i w_{it} \frac{p_{i0}}{p_{it}} \right]^{-1} \tag{2}$$

A Laspeyres index measures the cost of a fixed basket of goods with respect to the cost in the base period. This index tends to overestimate the cost of living not allowing substitutions among goods and new products' introduction. Conversely, a Paasche index tends to understate the cost of living weighting prices by current consumption pattern (Diewert, 1998; Boskin et al.,1998). Thus, when either Laspeyres or Paasche indexes are used, the resulting CPI cannot be considered a good measure of the cost of living. The use of superlative indexes has been supported as preferable with respect to Laspeyres-type indexes as they have been found to "approximate the true cost-of living index under certain assumptions" (Boskin et al., 1998). Moreover, the use of superlative indexes can handle the potential dimensionality problem which can arise when estimating the cost of living through a demand equations approach (Boskin et al. 1998). The (unobservable) Pollak-Konüs true cost of living index has been found to be between the Paasche and Laspeyres price indexes (Diewert, 1998). This result suggests that taking an average of Paasche and Laspeyres price indexes can closely approximate the true cost of living (Diewert,1998). In particular the Fisher index, which is the geometric mean of the Paasche and Laspeyres index ($Fisher_i = [Laspeyres_t * Paasche_t]^{1/2}$) can be a good candidate to measure the cost of living (Diewert,1998)[2].

The use of scanner data in the computation of consumer price indexes has been proposed since the late 90's. For example, Boskin et al. (1998) recommended their use as a possible way to reduce the bias carried by the CPI on measuring the cost of living. In particular, scanner data are appealing for two main reasons. First, the cost related to their collection is relatively cheap compared to other types of data. Second, their use can facilitate the computation of a flexible basket index. In addition, scanner data record real purchasing consumer decision, so they implicitly store the effects of marketing activities on consumer choices allowing their incorporation when accounting for substitution patterns among different products.

However, the use of scanner data in the CPI computation has some drawbacks. In particular, the intensity of price and quantity bouncing due to sales has been found to cause chain drift bias when computing price indexes using high frequency data (de Haan and van der Grient , 2011; Ivanic et al., 2009, Nakamura et al., 2011). Nakamura et al. (2011) show that specific aggregation over time and over stores can help in reducing the price bouncing effect and consequently the chain drift bias in the index computation. A more promising alternative is to resort to a drift-free chain index such as the GEKS index, as suggested by Ivanic et. al (2009).

Using the SymphonyIRI dataset we compute drift-free GEKS indexes as proposed by Ivanic et al. (2009). The GEKS is a multilateral index, usually used in international trade to compare several entities. Differently from some bilateral indexes, multilateral indexes satisfy the Fisher's circularity test (Fisher, 1922) which allows the comparison of entities directly among each other or through their relationship with a third one.

For example, consider P_{ij} to be the Fisher index between entities i and j ($j = 1 \dots M$) and P_{kj} to be the Fisher index between entities k and j. The GEKS index between i and k will be the geometric mean of the two Fisher indexes (Ivanic et al., 2009). We can, then, write the $GEKS_{i,k}$ as follows:

$$GEKS_{i,k} = \prod_{j=1}^{M} [\frac{P_{ij}}{P_{kj}}]^{1/M} \qquad (3)$$

Ivanic et al. (2009) proposed to use the GEKS index to make comparison among T different time periods $j = 1 \dots T$. Considering the reference time period $t = 0$ the GEKS price index between 0 and t, as in Ivanic et al. (2009), will be:

$$GEKS_{o,t} = \prod_{t=0}^{T} \left[\frac{P_{0l}}{P_{lt}}\right]^{1/(T+1)} = \prod_{t=1}^{t} GEKS_{t-1,t} \qquad (4)$$

[2] Diewert (1998) states several reasons to support the use of a Fisher index as ideal candidate to approximate the cost-of living index. Please refer to it for a deeper analysis

The circularity property of the GEKS index allows to write the GEKS index between time period 0 and t ($GEKS_{o,t}$) as a period-to-period chain index ($\prod_{t=1}^{t} GEKS_{t-1,t}$).

In addition, the GEKS index is free of chain drift as it satisfies the multi-period identity test. The multi-period identity test has been proposed by Walsh (1901) and Szulc (1983) as a method to test for the presence of drift chain bias. Given price indexes among all different time periods, the price index will not suffer from chain drift bias if the product of indexes among all possible time combinations is equal to one. For example, in the case of a three-time period, given the price indexes between periods 1 and 2, $p(p_1, p_2, q_1 q_2)$, periods 2 and 3, $\mathrm{p}(\mathrm{p}_2, \mathrm{p}_3, \mathrm{q}_2\mathrm{q}_3)$, and periods 3 and 1, $\mathrm{p}(\mathrm{p}_3, \mathrm{p}_1, \mathrm{q}_3\mathrm{q}_1)$, if the product of the three indexes is equal to one, the price index formula is not affected by drift chain bias. The GEKS index satisfies the multi-period identity by construction.

Further, another advantage of the GEKS index is its suitability when using a flexible basket. Thus, the GEKS index is a good candidate for CPI computation using high frequency data, given that scanner data present a high heterogeneity in product assortment over time.

In computing the GEKS price index for our product categories we choose to aggregate at the lower time level of aggregation (week). Previous studies have shown that the GEKS indexes are free of drift bias, but, to our knowledge, no study has compared GEKS indexes calculated under different level of time aggregation. For this reason, we choose the level of aggregation which is considered the most informative to identify retailers' strategies. For instance, many retailers' strategies, like promotion activities, have usually a weekly or bi-weekly time horizon.

We repeat the GEKS computation for each product category. For any of the eight product categories our final dataset is represented by indexes along the one-hundred and fifty-six weeks for the thirty-three different chain-type of store combinations.

To understand the source of variation on the GEKS index we first implement a variance decomposition using a three-way ANOVA which accounts for chain, type of store and time for each product category. Results (table 4) show how the ANOVA model sum of squares over the total sum of squares ranges from around 30% in uht milk and cream to over 60% in refrigerated milk, thus indicating strong variability among different product categories. This might be due to different marketing strategies and different competitive interactions among manufactures, but also to the characteristics of the products, like the shelf life, and to their degree of differentiation. This analysis provides two major suggestions. First, accounting for the variability among chains, types of store and time periods can be a good estimation strategy to identify the contribution of retailers' strategies. Second, a model that considers products categories separately allows us to better understand if the impact of a retailer strategy has a general consistent effect among products categories.

Table 4. Results of three-ways ANOVA of the GEKS index for each product category. Sum of squares and its percentage contribution on the total variance.

	Butter		Cheese		UHT milk		Refr. milk		Yogurt		Mozzarella		UHT Cream	
Model	**23.1**	**53.9%**	**19.7**	**34.4%**	**6.4**	**30.1%**	**8.9**	**61.8%**	**7.0**	**50.9%**	**16.4**	**43.0%**	**7.4**	**30.3%**
Chain	5.9	13.6%	10.3	18.1%	2.8	13.3%	5.0	34.5%	3.8	27.9%	13.0	34.1%	3.8	15.7%
Type of store	3.4	7.9%	1.7	3.0%	1.3	6.1%	0.9	6.0%	1.6	11.9%	0.4	1.1%	0.4	1.5%
time	13.9	32.3%	7.6	13.3%	2.3	10.7%	3.1	21.3%	1.5	11.1%	3.0	7.8%	3.2	13.1%
Residual	**19.8**	**46.1%**	**37.5**	**65.6%**	**14.8**	**69.9%**	**5.5**	**38.2%**	**6.7**	**49.1%**	**21.7**	**57.0%**	**17.0**	**69.7%**
Total	**43.0**		**57.2**		**21.2**		**14.4**		**13.7**		**38.2**		**24.4**	
Observations	5148		5148		5148		5148		5148		5148		5148	
R-squared	0.54		0.34		0.30		0.62		0.51		0.43		0.30	

All p-values are < 0.01. Significant at 1% level

4. MODEL SPECIFICATION AND ECONOMETRIC STRATEGY

We consider the following regression model:

$$y_{itj} = \mathbf{x}'_{itj}\boldsymbol{\beta} + \mu_i + \nu_t + \lambda_j + u_{itj} = \mathbf{x}'_{itj}\boldsymbol{\beta} + \varepsilon_{itj}, \qquad (5)$$

where y_{itj} is the dependent variable, given by the GEKS index multiplied by the chain and type-of-store specific average price at time $t = 1$; $\mathbf{x}_{itj}$ is a k vector of explanatory variables, proxying retailers' strategies, and $\boldsymbol{\beta}$ a k vector of parameters, μ_i is the chain-specific effect (indexed $i = 1, \dots, N$), ν_t the time-specific effect (indexed $t = 1, \dots, T$), λ_j the type of store-specific effect (indexed $j = 1, \dots, L$), u_{itj} the remainder error term and ε_{itj} the composite error term. Table 5 shows the description of the explanatory variables used in the model and their rationale.

Hence, defining the $n \times N$ matrix $\boldsymbol{\Delta}_\mu$, the $n \times T$ matrix $\boldsymbol{\Delta}_\nu$ and the $n \times L$ matrix $\boldsymbol{\Delta}_\lambda$ and using matrix notation, we can write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}_\mu\boldsymbol{\mu} + \boldsymbol{\Delta}_\nu\boldsymbol{\nu} + \boldsymbol{\Delta}_\lambda\boldsymbol{\lambda} + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad (6)$$

where $\mathbf{X}$ is a $n \times k$ matrix of explanatory variables, $\boldsymbol{\mu}$ the $N \times 1$ vector of chain-specific effects, $\boldsymbol{\nu}$ the $T \times 1$ vector of time-specific effects, $\boldsymbol{\lambda}$ the $L \times 1$ vector of format-specific effects, $\mathbf{u}$ the $n \times 1$ vector of residual disturbances and $\boldsymbol{\varepsilon}$ the $n \times 1$ vector of composite error terms.

Davis (2002) develops simple matrix algebra techniques that simplify and unify much of the previous literature on estimating Error Component Models (ECMs). In fact, the simple analytic results provided by Davis (2002) are useful for analyzing a very broad set of models with complex error structures (multi-way ECMs).

Table 5: Explanatory variables descriptions, rationale and expected impacts.

Variable Name	Computation	Rationale	Expected impacts
$\boldsymbol{Assortment_{itj}}$	Weekly average number of BU by point of sales for each chain-type of store unit	The variable aims to capture the assortment strategy of each chain-type of store pair over time.	*Mixed effect:* (+) higher assortment might lead to higher costs for the retailer, thus higher prices. (-) higher assortment might lead to an increase of price competition among manufactures.
$\boldsymbol{Share\ PL_{itj}}$	Weekly average of PL share in value by points of sales for each chain-type of store unit	The variable aims to capture the PL market competitiveness with respect to the other brands.	*Negative effect:* (-) as the PL prices are usually lower than the NB counterpart we expect to reduce an upward inflation rate trend.
$\boldsymbol{PL\ line\ extension_{itj}}$	Weekly average of the share of segments where the PL is present by points of sales for each chain-type of store unit. The shares are computed as number of segments where the PL are present over all segments present in the market.	The variable aims to measure the line extension of the PL in the market, in particular their expansion strategy in different market segments.	*Mixed effect:* (+) The presence of PL in "*premium*" segments may lead consumers to shift their consumption to relatively more expensive segments. (-) (+) The presence of PL in a market segment may have mixed effects on NB prices.
$\boldsymbol{Promotion_{itj}}$	Weekly average of the value share sold under promotion for each chain-type of store unit.	The variable aims to capture the intensity of promotion activity over time and for each chain-type of store unit.	*Negative effect:* (-) we expect a more intense promotion strategy to reduce an upward inflation rate trend.
$\boldsymbol{PL-Promotion_{itj}}$	Weekly average of the PL value share sold under promotion for each chain-type of store unit. The shares are computed as value of PL sold under promotion over the total PL sold in value.	The variable aims to capture the intensity on promotion activity by PL. Moreover we want to capture the intensity of PL promotion activity compared to the overall average promotion intensity in the market.	*Mixed effect:* (-) (+) PL promotion activity might be more or less effective in reducing an upward inflation trend with respect to the NB counterpart.

The within transformation of the three-way ECM is:

$$\mathbf{Q}_{\Delta_{3w}} = \mathbf{Q}_A - \mathbf{P}_B - \mathbf{P}_C \tag{7}$$

with

$$\begin{aligned} \mathbf{P}_A &= \boldsymbol{\Delta}_\mu \boldsymbol{\Delta}_N^{-1} \boldsymbol{\Delta}'_\mu && \rightarrow \mathbf{Q}_A = \mathbf{I}_n - \mathbf{P}_A \\ \mathbf{P}_B &= \mathbf{Q}_A \boldsymbol{\Delta}_\nu (\boldsymbol{\Delta}'_\nu \mathbf{Q}_A \boldsymbol{\Delta}_\nu)^- \boldsymbol{\Delta}'_\nu \mathbf{Q}_A && \rightarrow \mathbf{Q}_B = \mathbf{I}_n - \mathbf{P}_B \\ \mathbf{P}_C &= \mathbf{Q}_A \mathbf{Q}_B \boldsymbol{\Delta}_\lambda (\boldsymbol{\Delta}'_\lambda (\mathbf{Q}_A \mathbf{Q}_B) \boldsymbol{\Delta}_\lambda)^- \boldsymbol{\Delta}'_\lambda \mathbf{Q}_A \mathbf{Q}_B \end{aligned} \tag{8}$$

with $\boldsymbol{\Delta}_N = \boldsymbol{\Delta}'_\mu \boldsymbol{\Delta}_\mu$ and $\mathbf{I}_n$ is the identity matrix of dimension n and where $\mathbf{Q}_A \mathbf{Q}_B = \mathbf{I}_n - \mathbf{P}_A - \mathbf{P}_B$ (Davis, 2002). Therefore the fixed effect (FE) estimator is:

$$\boldsymbol{\beta}^{WT} = \left(\mathbf{X}' \mathbf{Q}_{\Delta_{3w}} \mathbf{X}\right)^{-1} \left(\mathbf{X}' \mathbf{Q}_{\Delta_{3w}} \mathbf{y}\right). \tag{9}$$

where μ_i, ν_t and λ_j are assumed to be fixed parameters and $u_{it} \sim IID(0, \sigma_u^2)$. In the three-way random effect (RE) model all error components are random variables: $\mu_i \sim IID(0, \sigma_\mu^2)$, $\nu_t \sim IID(0, \sigma_\nu^2)$, $\lambda_j \sim IID(0, \sigma_\lambda^2)$ and $u_{it} \sim IID(0, \sigma_u^2)$. The covariance matrix of the composite error ε_{itj} is:

$$\boldsymbol{\Omega}_{3w} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\varepsilon^2 \cdot \mathbf{I}_n + \sigma_\mu^2 \cdot \boldsymbol{\Delta}_\mu \boldsymbol{\Delta}'_\mu + \sigma_\nu^2 \cdot \boldsymbol{\Delta}_\nu \boldsymbol{\Delta}'_\nu + \sigma_\lambda^2 \cdot \boldsymbol{\Delta}_\lambda \boldsymbol{\Delta}'_\lambda. \tag{10}$$

Following Davis (2002), we define the following matrices

$$\mathbf{V}_N = \mathbf{I}_n - \mathbf{\Delta}_\mu \left(\mathbf{\Delta}_N + \frac{\sigma_u^2}{\sigma_\mu^2} \cdot \mathbf{I}_N\right)^{-1} \mathbf{\Delta}'_\mu \quad \rightarrow \mathbf{\Omega}_{1\text{w}}^{-1} = \frac{1}{\sigma_u^2} \cdot \mathbf{V}_N$$
$$\begin{cases} \mathbf{W}_{TN} = \frac{\sigma_u^2}{\sigma_\nu^2} \cdot \mathbf{I}_T + \mathbf{\Delta}'_\nu \mathbf{V}_N \mathbf{\Delta}_\nu \\ \mathbf{V}_{TN} = \mathbf{V}_N - \mathbf{V}_N \mathbf{\Delta}_\nu \mathbf{W}_{TN}^{-1} \mathbf{\Delta}'_\nu \mathbf{V}_N \end{cases} \quad \rightarrow \mathbf{\Omega}_{2\text{w}}^{-1} = \frac{1}{\sigma_u^2} \cdot \mathbf{V}_{TN} \tag{11}$$
$$\begin{cases} \mathbf{W}_{LTN} = \frac{\sigma_u^2}{\sigma_\lambda^2} \cdot \mathbf{I}_L + \mathbf{\Delta}'_\lambda \mathbf{V}_{TN} \mathbf{\Delta}_\lambda \\ \mathbf{V}_{LTN} = \mathbf{V}_{TN} - \mathbf{V}_{TN} \mathbf{\Delta}_\lambda \mathbf{W}_{LTN}^{-1} \mathbf{\Delta}'_\lambda \mathbf{V}_{TN} \end{cases} \quad \rightarrow \mathbf{\Omega}_{3\text{w}}^{-1} = \frac{1}{\sigma_u^2} \cdot \mathbf{V}_{LTN}$$

where $\mathbf{I}_N$ is the identity matrix of dimension N, $\mathbf{I}_T$ is the identity matrix of dimension T and $\mathbf{I}_L$ is the identity matrix of dimension L. Then the GLS estimator is:

$$\boldsymbol{\beta}^{GLS} = (\mathbf{X}'\mathbf{\Omega}_{3\text{w}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}_{3\text{w}}^{-1}\mathbf{y}). \tag{12}$$

We derive Quadratic Estimations (QUEs) for σ_u^2, σ_μ^2, σ_ν^2 and σ_λ^2, and, by using the FE residuals, we average them over chains, type of stores and periods. Since we are considering a constant term, with the FE residuals $\mathbf{e} \equiv \mathbf{y} - \mathbf{X} \cdot \boldsymbol{\beta}^{WT}$ and with $\text{f} \equiv \text{E}_\text{n} \cdot \text{e} = \text{e-}\bar{\text{e}}$, where $\mathbf{E}_n = \mathbf{I}_n - \bar{\mathbf{J}}_n$, $\bar{\mathbf{J}}_n = \frac{\mathbf{J}_n}{n}$ and $\mathbf{J}_n$ is a matrix of ones of dimension n, we equate:

$$\begin{aligned} q_n &= \mathbf{f}'\mathbf{Q}_{\Delta_{3\text{w}}}\mathbf{f} \\ q_N &= \mathbf{f}'\mathbf{\Delta}_\mu \mathbf{\Delta}_N^{-1} \mathbf{\Delta}'_\mu \mathbf{f} \\ q_T &= \mathbf{f}'\mathbf{\Delta}_\nu \mathbf{\Delta}_T^{-1} \mathbf{\Delta}'_\nu \mathbf{f} \\ q_L &= \mathbf{f}'\mathbf{\Delta}_\lambda \mathbf{\Delta}_L^{-1} \mathbf{\Delta}'_\lambda \mathbf{f} \end{aligned} \tag{13}$$

with $\mathbf{\Delta}_T = \mathbf{\Delta}'_\nu \mathbf{\Delta}_\nu$ and $\mathbf{\Delta}_L = \mathbf{\Delta}'_\lambda \mathbf{\Delta}_\lambda$, to their expected values:

$$\begin{aligned} \text{E}(q_n) &= (n - N - (T-1) - (L-1) - k) \cdot \sigma_u^2 \\ \text{E}(q_N) &= (N + k_N - k_0 - 1) \cdot \sigma_u^2 + (n - \lambda_N) \cdot \sigma_\mu^2 + (k_{NT} - \lambda_T) \cdot \sigma_\nu^2 + (k_{NL} - \lambda_L) \cdot \sigma_\lambda^2 \\ \text{E}(q_T) &= (T + k_T - k_0 - 1) \cdot \sigma_u^2 + (k_{TN} - \lambda_N) \cdot \sigma_\mu^2 + (n - \lambda_T) \cdot \sigma_\nu^2 + (k_{TL} - \lambda_L) \cdot \sigma_\lambda^2 \\ \text{E}(q_T) &= (L + k_L - k_0 - 1) \cdot \sigma_u^2 + (k_{LN} - \lambda_N) \cdot \sigma_\mu^2 + (k_{LT} - \lambda_T) \cdot \sigma_\nu^2 + (n - \lambda_L) \cdot \sigma_\lambda^2 \end{aligned} \tag{14}$$

Where $k_N = tr\left(\left(\mathbf{X}' \cdot \mathbf{Q}_{\Delta_{3\text{w}}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}' \cdot \mathbf{\Delta}_\mu \cdot \mathbf{\Delta}_N^{-1} \cdot \mathbf{\Delta}'_\mu\right)$, $k_T = tr\left(\left(\mathbf{X}' \cdot \mathbf{Q}_{\Delta_{3\text{w}}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}' \cdot \mathbf{\Delta}_\nu \cdot \mathbf{\Delta}_T^{-1} \cdot \mathbf{\Delta}'_\nu\right)$, $k_L = tr\left(\left(\mathbf{X}' \cdot \mathbf{Q}_{\Delta_{3\text{w}}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}' \cdot \mathbf{\Delta}_\lambda \cdot \mathbf{\Delta}_L^{-1} \cdot \mathbf{\Delta}'_\lambda\right)$, $k_0 = \frac{\boldsymbol{\iota}'_n \cdot \mathbf{X} \cdot \left(\mathbf{X}' \cdot \mathbf{Q}_{\Delta_{3\text{w}}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}' \cdot \boldsymbol{\iota}_n}{n}$, $\lambda_N = \frac{\boldsymbol{\iota}'_n \cdot \mathbf{\Delta}_\mu \cdot \mathbf{\Delta}'_\mu \cdot \boldsymbol{\iota}_n}{n}$, $\lambda_T = \frac{\boldsymbol{\iota}'_n \cdot \mathbf{\Delta}_\nu \cdot \mathbf{\Delta}'_\nu \cdot \boldsymbol{\iota}_n}{n}$ and $\lambda_L = \frac{\boldsymbol{\iota}'_n \cdot \mathbf{\Delta}_\lambda \cdot \mathbf{\Delta}'_\lambda \cdot \boldsymbol{\iota}_n}{n}$. Moreover $k_{NT} = \text{tr}(\mathbf{\Delta}_{TN} \cdot \mathbf{\Delta}_N^{-1} \cdot \mathbf{\Delta}'_{TN})$ and $k_{NL} = \text{tr}(\mathbf{\Delta}_{LN} \cdot \mathbf{\Delta}_N^{-1} \cdot \mathbf{\Delta}'_{LN})$, with $\mathbf{\Delta}_{TN} = \mathbf{\Delta}'_\nu \cdot \mathbf{\Delta}_\mu$ and $\mathbf{\Delta}_{LN} = \mathbf{\Delta}'_\lambda \cdot \mathbf{\Delta}_\mu$, $k_{TN} = \text{tr}(\mathbf{\Delta}'_{TN} \cdot \mathbf{\Delta}_T^{-1} \cdot \mathbf{\Delta}_{TN})$ and $k_{TL} = \text{tr}(\mathbf{\Delta}_{LT} \cdot \mathbf{\Delta}_T^{-1} \cdot \mathbf{\Delta}'_{LT})$, with $\mathbf{\Delta}_{LT} = \mathbf{\Delta}'_3 \cdot \mathbf{\Delta}_2$, and $k_{LN} = \text{tr}(\mathbf{\Delta}'_{LN} \cdot \mathbf{\Delta}_L^{-1} \cdot \mathbf{\Delta}_{LN})$ and $k_{LT} = \text{tr}(\mathbf{\Delta}'_{LT} \cdot \mathbf{\Delta}_L^{-1} \cdot \mathbf{\Delta}_{LT})$.

To check for the validity of assumptions made on the structure of the three-way ECM we use the Lagrange Multiplier test statistic based on components of the loglikelihood evaluated at parameters estimates (Boumahdi et al., 2004). The loglikelihood function under normality of the disturbances is

$$\text{L} = \text{constant} - \frac{1}{2} \cdot \log|\mathbf{\Omega}_{3\text{w}}| - \frac{1}{2} \cdot \mathbf{e}'\mathbf{\Omega}_{3\text{w}}^{-1}\mathbf{e} \tag{15}$$

where $\boldsymbol{\vartheta} = \left(\sigma_\mu^2, \sigma_\nu^2, \sigma_\lambda^2, \sigma_u^2\right)'$. Under H_0 $\boldsymbol{\vartheta} = \bar{\boldsymbol{\vartheta}} = (0,0,0,\sigma_{OLS}^2)'$ where σ_{OLS}^2 is the variance of the OLS residuals $\mathbf{e}_{OLS}$. Then we compute the restricted score vector

$$\mathbf{D}(\boldsymbol{\vartheta}) = -\frac{n}{2\cdot\sigma_{OLS}^2}\cdot\bar{\mathbf{D}}(\boldsymbol{\vartheta}) = -\frac{n}{2\cdot\sigma_{OLS}^2}\cdot\begin{bmatrix} 1-\frac{\mathbf{e}'_{OLS}\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}'_\mu\mathbf{e}_{OLS}}{\mathbf{e}'_{OLS}\mathbf{e}_{OLS}} \\ 1-\frac{\mathbf{e}'_{OLS}\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}'_\nu\mathbf{e}_{OLS}}{\mathbf{e}'_{OLS}\mathbf{e}_{OLS}} \\ 1-\frac{\mathbf{e}'_{OLS}\boldsymbol{\Delta}_\lambda\boldsymbol{\Delta}'_\lambda\mathbf{e}_{OLS}}{\mathbf{e}'_{OLS}\mathbf{e}_{OLS}} \\ 0 \end{bmatrix} \quad (16)$$

and the information matrix

$$\mathbf{J}(\boldsymbol{\vartheta}) = \frac{1}{2\cdot\sigma_{OLS}^4}\cdot\bar{\mathbf{J}}(\boldsymbol{\vartheta}) = \frac{1}{2\cdot\sigma_{OLS}^4}\cdot\begin{bmatrix} \mathrm{tr}(\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}'_\mu\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}'_\mu) & \mathrm{tr}(\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}'_\mu\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}'_\nu) & \mathrm{tr}(\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}'_\mu\boldsymbol{\Delta}_\lambda\boldsymbol{\Delta}'_\lambda) & n \\ & \mathrm{tr}(\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}'_\nu\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}'_\nu) & \mathrm{tr}(\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}'_\nu\boldsymbol{\Delta}_\lambda\boldsymbol{\Delta}'_\lambda) & n \\ & & \mathrm{tr}(\boldsymbol{\Delta}_\lambda\boldsymbol{\Delta}'_\lambda\boldsymbol{\Delta}_\lambda\boldsymbol{\Delta}'_\lambda) & n \\ & & & n \end{bmatrix} \quad (17)$$

with $\mathbf{J}(\boldsymbol{\vartheta}) = E\left(\frac{\partial^2 L}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right) = [\mathbf{J}_{rs}]$ and $\mathbf{J}_{rs} = E\left(-\frac{\partial^2 L}{\partial\vartheta_r\partial\vartheta_s}\right)$. Hence the LM statistic under H_0 $\boldsymbol{\vartheta} = \bar{\boldsymbol{\vartheta}}$ is given by

$$\begin{aligned} \mathrm{LM} &= \mathbf{D}(\boldsymbol{\vartheta})'\mathbf{J}(\boldsymbol{\vartheta})^{-1}\mathbf{D}(\boldsymbol{\vartheta}) = \\ &= \left(-\frac{n}{2\cdot\sigma_{OLS}^2}\cdot\bar{\mathbf{D}}(\boldsymbol{\vartheta})'\right)\cdot(2\cdot\sigma_{OLS}^4\cdot\bar{\mathbf{J}}(\boldsymbol{\vartheta})^{-1})\cdot\left(-\frac{n}{2\cdot\sigma_{OLS}^2}\cdot\bar{\mathbf{D}}(\boldsymbol{\vartheta})\right) = \\ &= \frac{n^2}{2}\cdot\bar{\mathbf{D}}(\boldsymbol{\vartheta})'\bar{\mathbf{J}}(\boldsymbol{\vartheta})^{-1}\bar{\mathbf{D}}(\boldsymbol{\vartheta}) \end{aligned} \quad (18)$$

which under H_0 is asymptotically distributed as χ^2_{DF}. To test for the validity of the error component specification the LM test statistic can be easily computed under various assumptions on null variances in the alternative hypothesis (Boumahdi et al., 2004).

5. RESULTS

Using the Lagrange Multiplier test, as described in the previous paragraph, we check for the significance of chain and type of store unobservables on food inflation. The null hypothesis $\bar{\boldsymbol{\vartheta}} = (0,0,0,\sigma_{OLS}^2)'$ has been checked vs. various alternatives: the two one-way models for chain and type of store ($\vartheta = (\sigma_\mu^2, 0,0, \sigma_u^2)'$ and $\vartheta = (0,0,\sigma_\lambda^2,\sigma_u^2)'$), the two-way model for chain and type of store ($\boldsymbol{\vartheta} = (\sigma_\mu^2, 0, \sigma_\lambda^2, \sigma_u^2)'$), and the three-way model ($\vartheta = (\sigma_\mu^2, \sigma_\nu^2, \sigma_\lambda^2, \sigma_u^2)'$). We do not reject H_0 only for the one-way model for type of store effect in yogurt but we reject the null for both the two-way and the three-way models for the same product category, thus suggesting the existence of significant differences in unobservable strategies among chains and types of stores.

Table 6 shows fixed effect and random effect estimates for each of the seven product categories. As expected we found that intensity on sales (*Promotion*) significantly reduces the consumer price index for all products categories. However, sales on PL are relatively less intense than sales on national brands as shown from the significant and positive sign of the coefficient on *PL promotion* (refrigerated milk is the only category where this coefficient is not significantly different from zero, suggesting that the effect of PL promotion does not have a statistically different impact compared with the market average).

Table 6: Estimated parameters under three-way ECM

	Butter		Cheese		Milk		Mozzarella		Yogurt		UHT		Cream	
	FE	**RE**	**FE**	**RE**	**FE**	**RE**	**FE**	**RE**	**FE**	**RE**	**FE**	**RE**	**FE**	**RE**
Constant		7.1759***		9.7064***		1.3174***		8.5021***		4.7864***		1.0224***		4.8963***
Assortment	0.0056**	0.0057**	0.0063	0.0079*	0.0073***	0.0075***	0.0084***	0.0058**	-0.0022***	-0.0023***	-0.0002	-0.0002	-0.0133***	-0.0101***
Share PL	-0.4587***	-0.4193***	-0.3252	-0.2020	-0.2690***	-0.2729***	-2.1910***	-2.1114***	-2.3033***	-2.2219***	-0.0392***	-0.0331**	-0.4655***	-0.4369***
PL line extension	0.9635***	0.9302***	0.5905***	0.5074***	0.1039***	0.1071***	0.5814***	0.5952***	0.2311***	0.2596***	0.0714***	0.0671***	0.1594***	0.1873***
Promotion	-1.9547***	-1.9555***	-3.0148***	-3.0106***	-0.1751***	-0.1885***	-2.9341***	-2.8873***	-1.5281***	-1.5096***	-0.3281***	-0.3281***	-1.5510***	-1.5544***
PL Promotion	0.1063***	0.0948**	0.2192***	0.2094***	0.0007	0.0019	0.5934***	0.5767***	0.2149***	0.2056***	0.0275***	0.0251***	0.2640***	0.2462***
Chain A	7.9909		10.5869		1.3796		8.4764		4.7314		1.1048		5.5624	
Chain B	7.1281		9.3941		1.3594		9.6150		4.7268		1.0840		5.0050	
Chain C	6.6180		9.4239		1.3753		8.1941		4.9133		0.9543		4.5319	
Chain D	8.0334		10.1577		1.3458		8.5473		5.0813		1.0732		5.1977	
Chain E	6.6382		9.5528		1.1932		8.0914		4.7628		0.9020		4.8535	
Chain F	7.3105		8.6622		1.1944		7.5408		4.5730		1.0172		4.7723	
Chain G	6.4124		8.8967		1.3262		8.1640		4.5933		1.0204		4.8084	
Chain H	7.2565		9.5373		1.3063		8.7916		4.8004		1.0506		5.1923	
Chain I	6.9429		8.8134		1.3833		8.3064		4.8378		0.9689		5.1526	
Chain L	7.4062		10.4329		1.3077		8.0243		5.1277		1.0422		3.9636	
Chain M	7.3280		9.7043		1.3929		9.0008		5.0744		0.9514		4.8365	
Chain N	7.3670		9.8862		1.2141		8.2528		4.7713		1.0322		5.0473	
Chain O	7.2812		9.8655		1.4122		8.7428		4.8600		1.0075		5.0103	
Chain P	7.2001		11.7558		1.3863		8.9290		4.6290		1.0946		5.1639	
Hyper	6.7733		9.1916		1.2637		8.4021		4.7327		0.9930		5.0913	
Super	7.1692		9.9550		1.3211		8.4218		4.8533		1.0227		4.9252	
Superette	7.5622		9.9086		1.3620		8.5552		4.8623		1.0325		4.8486	
σ_{μ}^2		0.2146		0.5531		0.0057		0.2314		0.0284		0.0037		0.1269
σ_{v}^2		0.1137		0.1049		0.0007		0.0240		0.0015		0.0002		0.0025
σ_{λ}^2		0.1282		0.1375		0.0019		0.0064		0.0025		0.0002		0.0075
σ_{u}^2	0.1967	0.1967	0.4627	0.4627	0.0031	0.0031	0.3525	0.3525	0.0476	0.0476	0.0035	0.0035	0.1250	0.1250

*** 1% significance, ** 5% significance, * 10% significance

Higher PL shares (*Share PL*) are related to a decrease in the average category price, with the only exception of cheese where the PL share does not seem to lower the food inflation rate, perhaps suggesting that PL products have less competitive advantage in the cheese category. In fact, cheese production is usually characterized by high know how and high differentiation, and these characteristics might increase the contracting power of processors vis-à-vis retailers not allowing high price differential between PL and NB products.

Furthermore, results show that the development of PL in different market segments (*PL line extension*) for a given product category is related to an upward trend of the price inflation rate. This result might be explained with the development of the PL product lines in "premium" segments of the market. The introduction of a PL might cause a reduction of the average price of the premium segment where the introduction takes place.

This reduction in average prices might induce consumers to shift their consumption from relatively cheaper to more expensive segments. Even a possible reduction of the average prices in the segment where the PL entry takes place does not seem to completely balance the cannibalization of relatively cheaper market segments.

We can think of two different effects on prices related with the intensity of the assortment strategy. First, more BU can lead to a higher price competition among brands pushing downward prices. On the other side, a more intense assortment is an extra service the retailers offer to consumers leading to higher costs, thus higher prices. In fact, for different product categories we found mixed effects of the assortment strategy (*Assortment*). A downward impact on the food inflation rate from a higher degree of assortment, as in the yogurt and cream categories, is likely due to an increase in price competition among manufactures. The assortment strategy does not seem to have any effect on the food inflation rate in the UHT milk category. In all other segments the coefficient is positive and significant meaning that a higher assortment intensity causes an upward trend of the food inflation rate. The positive coefficient of the assortment strategy might suggest that consumers pay the service of having a richer assortment when going to the supermarket.

6. Discussion and Conclusion

The European Union has been affected by an increasing rate of food inflation, starting from the last decade with a sharp increase during 2008. Many explanations for this sharp increase in food prices along the food supply chain can be advanced. While the increase in input costs, such as energy, may be one of the factors contributing to food price increase, other phenomena can determine the upward food inflation rate. For instance, retailers using particular market strategies might be able to accelerate or slow down the inflation trend. Chevalier and Kashyap (2011) showed how retailers use promotion strategies to charge consumers different prices influencing their reservation price: "Thus, how consumers update reservation prices for individual goods becomes a critical factor affecting inflation" (Chevalier and Kashyap, 2011).

Broda et al. (2009) using household scanner data showed that "poor households systematically pay less than rich households for identical goods".

In this paper we use high frequency scanner data (weekly SymphonyIRI data from different point of sales in Italy from January 2009 to January 2012) to empirically explore the contribution of some observed retailers' strategies on seven dairy product categories. Moreover we test if the unobserved heterogeneity among chains and types of stores (hypermarket, supermarket and superette) gives a significantly different contribution to food inflation rates.

The novelty of this paper is first on the research design. At our knowledge, no empirical study has previously analyzed how retailers' strategies influence the food inflation rate. From our data, for each of the observed dairy product categories, chains and types of store we compute a weekly price index free of drift chain bias, as proposed by Ivanic et al. (2011). After computing the GEKS index for each product and chain-format combination, we use a three-way ECM estimator (Davis, 2002) to capture unobservable effects due to chain, time and type-of-store heterogeneity. Moreover, for each product of our dataset, we estimate the effect on the index of observed retailers' strategies such as promotional activities, PL share, retailers' assortment and PL line extension, adapting the three-way ECM estimation developed by Davis (2002).

Results show that while higher PL shares help in slowing down an upward food inflation rate, on the contrary higher PL line extension tends to accelerate it. Sales activities, as expected, alleviate the burden of a general increase in prices; however, PL sales have an effect on reducing the price inflation rate which is proportionally smaller than the overall average. This means that sales on PL may contribute less intensively on reducing a generalized upward price trend. Finally, assortment strategies have a mixed effect depending on the competition environment of the market we refer to. In general, unobservable characteristics related to chains and types of store play a significant role in controlling the rise of prices.

The research structure applied in this study might be further developed and used to explore smaller segments of the market within the same products. Moreover, having a geographic identification of the points of sales, it would be interesting to explore how retailers' strategies differ from rich to poor neighborhoods and their influence on the price differentials.

Furthermore, we believe that the use of high frequency data along with this methodology can be developed and used by statistical and governmental agency to monitor and explore the contributions of food retailers on inflation rates.

ACKNOWLEDGMENTS

This research was carried out as part of the TRANSFOP (Transparency of Food Pricing) research project (Scientific coordinator: Steve McCorriston), funded by the European Commission under the 7th Framework programme.

REFERENCES

Betancourt R., Gautschi D.(1993). "The Outputs of Retailers activities: Concepts, Measurements and Evidence for the US Census Data." *The Review of Economics and Statistics* 75:294-301.

Betancourt R., Gautschi D.(1992). "The Outputs of Retailers activities: French evidence." *Applied Economics* 24:1043-1052.

Boskin, M. S., Delberger, E. R., Gordon, R. J., Griliches, Z. and Jorgenson, D.W. (1998). "Consumerprices in the consumer price index and the cost of living". *Journal of Economic Perspectives* 12: 3-26.

Boumahdi, R., Chaaban, J. and A. Thomas (2004). "Import Demand Estimation with Country and Products Effects:
Application of Multi-Way Unbalanced Panel Data Models to Libanese Imports". Cahier de Recherche INRA, 2004-17.

Broda C., Leibtag A., Weinstein D.E. (2009). "The Role of Prices in measuring the Poor's leaving standards." *Journal of Economic Perspectives* 23:77-97.

Chevalier J. A., Kashyap A.K. (2011). "Best Prices". *National Bureau of Economic Research Working Paper 16680*, available at: http://www.nber.org/papers/w16680

Davis (2002). "Estimating multi-way error components models with unbalanced panel data structures" *Journal of Econometrics* 106: 67-95.

de Haan J., van der Grient H.A., (2011). "Eliminating Chain Drift in Price Indexes Based on Scanner Data". *Journal of Econometrics* 161: 36–46.

Diewert, W. E. (1998). "Index Number Issues in the Consumer Price Index". *Journal of Economic Perspectives* 12: 47-58.

Ivanic L . et al. (2011). "Scanner data, time aggregation and the construction of price indexes." *Journal of econometrics* 161:24-35.

Nakamura A.O et al. (2011)."Price dynamics, retail chain and inflation measurement." *Journal of Econometrics* 161: 47-55.

Silver M., Heravi S. (2001). "Scanner data and the measurement of inflation". *The Economic Journal* 111: 384-404.

Szulc, B.J. (Schultz) (1983). Linking price index numbers. In: Diewert, W.E.,Montmarquette, C. (Eds.), *Price Level Measurement*. Statistics Canada, Ottawa, 537–566.

Walsh, C.M. (1901). *The Measurement of General Exchange Value*. Macmillan, NewYork.