Scanner data and quality adjustment

The French Experience*

Isabelle Léonard, Patrick Sillard and Gaëtan Varlet†

April 8, 2013

Abstract

Insee has launched a pilot experiment which aims at introducing scanner data in the French CPI. Started in 2010 with a small set of companies and a small number of industrial food products, the experiment has now reached a larger scale with a daily transmission of data covering 30% of the market. This experiment gives Insee the occasion to review the quality adjustments in the French CPI. Thus, Insee has chosen a strategy of analysis that is mainly based on the same principle as the one applied for the rest of the French CPI: the sample is drawn yearly in the universe of the products in order to reach a certain level of accuracy in the resulting CPI; a two-steps computation is made : the first step consists in computing micro-aggregate while dealing with possible substitutions that occur at the micro-level by the use of adequate price index formulae and the second step consists in the traditional Laspeyres aggregation; the product is followed until it disappears. It is then replaced by a new product after a quality adjustment.

The paper deals with the quality adjustment applied in scanner data. Besides prices and quantities associated to EAN (barcode), each day, and each shop, Insee has bought a database containing descriptive variables of each sold EAN. This information makes it possible to chose in a proper way, replacement products based on a kind of distance between products. It also makes it possible to estimate in an objective way, quality differences with respect to descriptive variables. We compare for yoghurts and chocolate bars the one year index variation with different strategies for quality adjustment. We show that quality adjustments are necessary, even for food products. We also show that the overlapping method which can be easily applied with scanner data may compete, in terms of results, with Hedonics.

JEL Codes : E31 ; C8 ; D1.

Key-words : Price indices, quality adjustment, scanner data.

*We wish to thank Eurostat for its support.

†Institut National de la Statistique et des Études Économiques – France ; corresponding author : Patrick Sillard (patrick.sillard@insee.fr)

1 Introduction

The main French retail chains have used, for quite a long time, centralised databases for management of stocks. In these databases, the products are identified with their EAN (European Article Number e.g. the barcode). In addition, since the beginning of 2000's, these data are collected by two private companies which, in agreement with the retailers, make market studies with these data. The interest of these data for CPI computation is clear since it gives a exhaustive picture of consumption (daily quantities and prices in all possible shops). A few European countries currently compute their CPI with the use of scanner data and many countries have started to study the introduction of these data in CPI. The French National statistical institute (INSEE) launched in 2010 a pilot project in order to get some insights on the suitability of these data for CPI purposes. In particular, there are two key issues Insee wants to examine:

1. the first one is the notion of product: in the CPI, the elementary product is identified physically from its characteristics observable for the consumer. In scanner data, the product is identified through its EAN code. And we know that the same product (in the classical CPI sense) might have different EANs. It is the case for some discount packagings. Generally, in that case, the consumer may still identify the difference. But there are cases where a perfectly unique product has two or more different EANs. For yoghurts for example, if the product is produced in one plant or another, it might have different EANs. And we found cases in our database where the same physical product is sold with different EANs at the same time in the same shop. This happens because this shop, for this product, is supplied by two different factories. Then we see that the concept of product in the CPI meaning is not confounded with concept implicitly described by the EAN.

2. The second issue is to adapt the usual data process which is more or less suitable for a certain volume of information to a volume of data that has nothing in common with the previous one. For example, in the French CPI, about 30 000 price observations are done each year for yoghurts, while there are 2 000 000 price observations for yoghurts in the scanner database we use for this paper, which only covers 25% of the main French retail chains.

Of course, the data processing we may think of is intimately related to the information we have. The data we use in this paper are presented in section 2. Without going into details now, we may say that in addition to daily sales (price and quantities), we have a full set of variables describing each EAN-identified product. This allows to make mass treatment of these data too.

This paper do not deal specifically with the first issue above: we use a sample of the data and we create series of products through a reasonable algorithm to select replacing when an article that was followed is missing in the shop where it was sold. The size of the sample is about 20 times larger than the sample followed in the current CPI for the same families of product. Insofar we make a sample selection, this approach is replicable on various samples and allows us to estimate error bars for our indices estimates and then to compare various methods of index computation.

This paper then mainly deals with the second issue presented above: insofar we may think of mass treatment of the main price information together with meta-data on the products, we may review the quality adjustment method. Indeed, for food goods, Insee use the so-called method of the bridged overlap with a real price increase to monitor quality adjustment (Armknecht, Moulton & Stewart 1994, Armknecht & Moulton 1995, Triplett 2006) when a followed product is missing. This is the best that can be done under the set of information we collect with the traditional price collection. In particular, the price of a possible replacing good is never observed together with the price of the replaced good at the same time period. And the main part of the observed characteristics aims at identifying the product in the shop rather than comparing the products. The situation with scanner data is different: when we think of replacing a product A by a product B, it is a very easy task to look back and measure the price of the two products when both were sold (if such a situation has existed) and as soon as we have a full set of EAN characteristics, we can *compare* the two products in terms of characteristics.

The paper then focuses on this issue: we compare with scanner data different strategies of quality adjustment and discuss the advantages and disadvantages of each of them.

The paper is divided into three parts: first we present the data and the basic algorithm for CPI computation used here; then we go into the quality adjustment method employed in the paper and we finally present the results we got.

2 From the data to a CPI

At the beginning of the project, at the end of 2010, Insee initiated a discussion with the main French retailers chains in order to get access to scanner data. Some of these retail chains have accepted to give access to their scanner data (about 30% of the potential "market"). Then Insee bought to the Symphony IRI Group (SIG) a test-sample of 3 years of scanner data.
This sample covers the weekly sales data (quantities, expenditures and prices of the sales) for 17 families of products for 1 000 hypermarkets and supermarkets during three years (2007 to 2009).
The sample contains 141 400 000 observations (45.6 million in 2007, 47.0 million in 2008 and 48.8 million in 2009). An observation corresponds to the sales (quantities and price) of a barcode in a store during one week. For the present paper we focus on the yoghurt family of products and the chocolate bars. In average, 224 different EANs of yoghurts were sold in a supermarket in December 2008 and 201 of Chocolate bars.
The basic information contained in the files for a given triplet [EAN, shop, week] is the number of products sold and the price (eventually averaged when different prices were applied for the elementary cell). In the monthly price index computation presented hereafter, prices and quantities are averaged over the month.

Two additional files were bought by Insee: the first one describes each barcode through variables

such as brand, perfume, packaging... (see table 1 for an overview of the file contents for yoghurts). The second one describes the shops (company, city, area...).

In this data, the statistical unit is the EAN barcode crossed with the shop. It may be followed in time, exactly as the goods are followed in the current CPI. When the product disappears, it must be replaced by another good, close to the first one, exactly as is done currently in the CPI. The algorithm we chose is based on a kind of distance between EANs computed from the characteristics of the products, as they can be seen in the file of family characteristics (see table 1 fro an example for yoghurts). When a product disappears, we select another one according to proximity. The proximity is built on a subset of characteristics that we consider as the most important ones (for example, variety[1] of product, outlet, brand, volume range) : the idea is to find a replacing product that has the same characteristics. If there is a single candidate, it is chosen. If there is more than one candidate, we select the one whose price during the last month is the closest to the price of the missing product. If there are still several candidates, we select the one whose expenditures during the last month are the closest to replaced product. Finally, if there is no replacement with respect to these criteria, we relax a proximity criterion to find a candidate. The different criteria are presented in table 2. The quality rank of replacement is classified into a 6-levels scale: from 1 (high quality) to 6 (low quality).

At this stage, we have continuous series of products with replacement events when necessary. A sample is drawn in these series for products available in November and December 2008. This sample corresponds to a sampling rate of about 2% with a probability of inclusion proportional to the sales of these two months.

And then, a price index is computed for the family: first an elementary index is computed for the cell defined by the considered outlet and the considered variety of products[2]. Then, this elementary index is aggregated up to the family level through a Laspeyres-type aggregation process.

In order to computed error bars over the resulting indexes, a certain number of simulations are done. Table 3 shows the number of samples drawn for yoghurts and chocolate bars, as well as the size of each sample in terms of considered series of products.

[1]This notion of variety already exists in the current CPI. It corresponds to the ultimate division of products, without being a partition of the consumption. It is defined in a very detailed manner and is viewed as representative (in the statistical meaning) of the sub-part of consumption that it is supposed to represent. For example, for yoghurts, there are 6 varieties of products in the current CPI and they cover about half of the family they are supposed to represent. It means that in the CPI, we suppose that these 6 varieties taken as a whole have the same price dynamics as the whole family of yoghurts. Therefore, this varieties cover in general well sold and well followed products. The 6 current CPI varieties of yoghurts have been identified in the scanner dataset through adequate filters applied on the variables describing EANs characteristics. The same has been done for chocolate bars. Additional pseudo-varieties have also been developed to complement the coverage up to almost the whole families of products.

[2]This approach is a bit different from that applied in the current CPI where the elementary cell is the city, and not the shop. This choice is made in the context of this paper. It is still under study.

Table 1: Variables describing the yoghurt family

Variables	Categories	Main values
IRI Family	1	5701
Family description	1	Yoghurts
Brand	112	Activia (17%), Panier (9%), U (9%), Taillefine (8%), Velouté (6%), Auchan (5%)
Bonus 1	13	12/4 (30%), 8CO (27%), 8/4B (16%), 4/2B (7%)
Bonus 2	0	-
Location	1	Refrigerated cabinet
Type of product	2	Yoghurts (99,97%), Batch of yoghurts (0,03%)
Packing	6	Plastic pot (89%), Glass pot (8%), Cardboard pot (2%), Stoneware (1%), Bucket (0%), Bi-comp pot
Perfume variety	202	Natural (22%), Assorted fruit (12%), Vanilla (8%), Strawberry (6%), Red fruit (4%)
Fair Trade Information	0	-
Active ingredient	3	Bifidus (94%), Anti ch (6%), Omega 6 (0,2%)
Ethnic info	2	Standard (99,99%), Halal (0,01%)
Promoting information	7	Shock pricec (71%), Special price (14%), Special Offer (8%), Eco Pack (5%), Offers eco (2%)
Biological information	3	Non bio (94%), Bio (5%), Bio AB (1%)
Content of milk	4	Full-cream milk (69%), Skimmed milk (25%), Plant (6%), Semi-skimmed milk (0,5%)
Additives	32	Pieces of fruits (66%), Pulp (17%), Fruits in the background (8%), Fruit bed (2%)
Sugar content	3	Sweetened (78%), Unsweetened (20%), Sugar of cane (3%)
Process	10	Standard (39%), Farm (32%), Stirred (17%), Bilayer (6%), Creamy (3%)
Fat content in %	7	0% MG (98,5%), 6,5% MG (0,7%), 2,9% MG (0,5%)
Fat content	2	Regular (75%), Reduced (25%)
Type of yoghurt	5	Fruit Yoghurt (56%), Natural Yoghurt (22%), Flavoured Yoghurt (22%)
Total volume	42	500gr (46%), 1000gr (14%), 1500gr (10%), 2000gr (8%), 400gr (6%), 300gr (5%)
Number of units in the pack	8	4ct (54%), 8ct (14%), 2ct (10%), 12ct (10%), 16ct (8%), 1ct (2%), 6ct (2%), 3ct (0%)
Number of bags per pack	1	1ct
Volume per unit	23	125gr (79%), 150gr (7%), 100gr (7%), 115gr (2%), 140gr (1%), 120gr (1%), 135gr (1%)

Note: all the variable are discrete ones. Column "Categories" indicates the number of different possible values and the column "Main terms" the most frequent values with the corresponding percentage frequency (for available values only) in parenthesis.

Table 2: Statistics about quality rank of replacements

		Yoghurts	Chocolate bars
1	Same variety, same outlet, same brand	73%	55,7%
2	Same variety, same outlet	27%	44,3%
3	Same variety, same city, same brand	0%	0%
4	Same variety, same city	0%	0%
5	Same variety, same brand	0%	0%
6	Same variety	0%	0%
Total	*100%*	*100%*	

Table 3: Number of simulations and size of elementary samples

	Yoghurts	Chocoloate bars
Sample size	3592	2064
Number of samples	100	200

3 Tested quality adjustment method

Some papers have already identified the potential of scanner data in improving the practical implementation of quality adjustment methods (e.g. Ahnert & Kenny (2004),Silver & Heravi (2000),Silver & Heravi (2005)). Insee will, most probably, introduce scanner data in the field of food products first (except fruits and vegetables). Even if quality adjustment are known to be mainly significant in other sectors of consumption, the availability of scanner data together with precise description of the products invites us to have a look at this issue for food products. This paper aims at reviewing the possible techniques and their practical implementation in the French CPI in the case of food products. This section presents the various algorithms of quality adjustment used in this paper. The terminology used here is the one from the (Center of Excellence Network - CENEX 2008). For an comprehensive view of quality adjustment techniques, see (ILO, IMF, OECD, UNECE, Eurostat & The World Bank 2004).

In this section, we will consider a series of product[3] i and we assume that (new) product N replaces the product O (old) at month m for this series. $P^O_{i,m-1}$ is the price of good O for series i at month $m-1$ and $P^N_{i,m}$ is the price of good N for series i at month m. Prices $P^O_{i,m-1}$ and $P^N_{i,m}$ are both observed. The quality adjustment, in the following formulas, is assumed to be made on the price of product O month $m-1$. Thus we have a price $\tilde{P}^O_{m-1}$ that is the supposed to be the price of the series at month $m-1$ such as the level of quality is the same as the new good one. Therefore, the monthly variation of the series i contributing to the index of month m

[3]coincident with a specific barcode.

will be:

$$I_{i,m} = \frac{P^N_{i,m}}{\tilde{P}^O_{i,m-1}} \tag{1}$$

We will also define the J coefficient as the inverse ratio of $I_{i,m}$ to the value of the index if the replacement was ignored ($I^0_{i,m} = P^N_{i,m}/P^O_{i,m}$):

$$J_{i,m-1} = \frac{I^0_{i,m}}{I_{i,m}} = \frac{\tilde{P}^O_{i,m-1}}{P^O_{i,m-1}} \tag{2}$$

Written like that, $J_{i,m-1}$ appears as the ratio of the computed price of good O with the quality level of good N and the price of good O with its own level of quality. In other words, $J_{i,m-1}$ is the measure of the value that the consumer makes of differences in characteristics of the products (see appendix A for an economic approach of this value). If $J_{i,m-1} > 1$, the quality of the new good is higher, and vice-versa (formally computed at month $m-1$).
There are five different ways of computing $\tilde{P}^O_{m-1}$ (sections 3.1 to 3.5).

3.1 Equivalent products or no quality adjustment[4] (1)

This is the basic case where the products are supposed to be directly comparable in terms of quality. Then

$$\tilde{P}^O_{m-1} = P^O_{m-1} \tag{3}$$

In this case,

$$J_{i,m-1} = 1 \tag{4}$$

3.2 The link-to-show-no-price change (2)

The technique consists in setting that the full price difference between the old product at month $m-1$ and the new one at month m is due to quality difference. Therefore

$$\tilde{P}^O_{i,m-1} = P^N_{i,m} \tag{5}$$

In this case,

$$J_{i,m-1} = \frac{P^N_{i,m}}{P^O_{i,m-1}} \tag{6}$$

3.3 Bridged overlap with real price increase (3)

A classical criticism of the last link-to-show-no-price change method is that the contribution of the series to the index is always null, even when there is a significant price change for observed products. In order to reduce this criticism, the Bridged overlap with real price increase method has been developed. It ensures that the contribution of series i is identical to the one obtained for the products really observed for the couple of months $m-1$ and m.
Let us define $\frac{1}{n}\sum_{j=1}^{n} \frac{P_{j,m}}{P_{j,m-1}}$ the average monthly evolution for the n observed products (same categories of products, for example, in the current CPI, same variety and same city). The idea is

[4]The numbers in parenthesis indicate the method. They are recalled in the tables of results.

almost the same as that of previous section (all the difference in price is a quality effect) but we assume that since the quality is identical, the monthly evolution of series i should be identical to the monthly real price increase. Then

$$\tilde{P}^O_{i,m-1} = \frac{P^N_{i,m}}{\frac{1}{n}\sum_{j=1}^{n}\frac{P_{j,m}}{P_{j,m-1}}} \tag{7}$$

and

$$J_{i,m-1} = \frac{P^N_{i,m}}{P^O_{i,m-1}} \times \frac{1}{\frac{1}{n}\sum_{j=1}^{n}\frac{P_{j,m}}{P_{j,m-1}}} \tag{8}$$

3.4 1 or 2 months overlapping

In this case we assume that the new product N and the old one O are sold together at the same time, during month $m-1$ (1 month overlapping) or month $m-2$ (2 months overlapping). And we assume that the value of the difference in price observed during month $m-1$ or $m-2$ is the value of the difference in characteristics the consumer makes.
Therefore, we have:

- **1-month overlapping (4)**:

$$\tilde{P}^O_{i,m-1} = P^N_{i,m-1} \tag{9}$$

and

$$J_{i,m-1} = \frac{P^N_{i,m-1}}{P^O_{i,m-1}} \tag{10}$$

- **Penultimate month overlapping (5)**:

$$\tilde{P}^O_{i,m-1} = \frac{P^O_{i,m-1}}{P^O_{i,m-2}} P^N_{i,m-2} \tag{11}$$

and

$$J_{i,m-1} = \frac{P^N_{i,m-2}}{P^O_{i,m-2}} \tag{12}$$

3.5 Hedonic models (6)

In this case, a Hedonic model is estimated through the econometric estimation of a model linking the logarithm of the price to the characteristics of the products. Let us write $\hat{P}^O_{i,m-1}$ the estimated Hedonic price of the old good and $\hat{P}^N_{i,m-1}$ for the new one. Then

$$\tilde{P}^O_{i,m-1} = \frac{P^O_{i,m-1}}{\hat{P}^O_{i,m-1}} \hat{P}^N_{i,m-1} \tag{13}$$

and

$$J_{i,m-1} = \frac{\hat{P}^N_{i,m-1}}{\hat{P}^O_{i,m-1}} \tag{14}$$

Overlapping and Hedonic techniques are both compatible with the usual consumer theory. We can show (see appendix A) that it is possible, in order to realize a constant utility price index

with goods of different levels of quality, to correct the prices for the difference in characteristics perceived by the consumer. In that case, the prices themselves might reveal the value the consumer makes on differences in characteristics, provided that a model linking price and characteristics is estimable or that the compared products are both present on the market at the same time. Of course the key issue is that in one way or another, an equilibrium is reached on the market at each period and the difference in price equilibrates the difference in tastes. This last assumption is obviously very restrictive but it is useful to keep this model in mind since all constant quality price corrections rely more or less on this idea.

4 Results

In the CPI, the treatment of quality effect depends on the type of replacement: equivalent (no quality correction) or dissimilar (quality adjusted).
All methods presented above have been tested on two families of products: yoghurts (section 4.1) and chocolate bars (section 4.2). The reference method is the Hedonic one (estimation based on product characteristics). A Hedonic model has been built for each family. Results of estimations are presented in appendix B.
100 samples are selected in order to simulate the results for yoghurt and 200 for chocolate bars (see table 3). The price change is computed between December 2008 and December 2009. Elementary indices are Jevons indices at [variety] x [outlet] level. The index for a family is the result of a Laspeyres aggregation of micro-indices (weighted by the expenditure on the base period being here November and December 2008).
Two approaches are possible concerning "Equivalent replacements": the first one consists in systematically applying a quality adjustment, even for quality-rank 1 replacements (sections 4 and 4.2.1); the second one consists in applying no quality adjustment when the quality-rank of the replacement is equal to 1 (see table 2), assuming that the price of the two products are directly comparable, or assuming that the products are equivalent (sections 4.1.2 and 4.2.2).

4.1 Simulations on Yoghurt family

The first results concern the yoghurt family (tables 4 to 10). First of all, the monthly price evolution for all the couple [EAN] x [month] between December 2008 and December 2009 is negative: on the period, the price decrease by 0.013% with a large standard deviation equal to 8.2% (table 4).

Table 4: *[Yoghurts]* Price evolution between month m and m-1 for still present products

Number of units (1)	Average	Std-Dev	Percentile 5	Percentile 50	Percentile 95
4 267 697	-0.013%	8.2%	-12.7%	0.0%	12.7%

Note: (1) the number of units corresponds to all month $m-1$ – month m price evolutions that can be observed (really)at the level of the barcode in the union of all selected samples over 12 months.

4.1.1 Overall quality adjustment regardless the type of replacement

In this section, we apply a quality adjustment in all situations of replacement. Table 5 shows the price evolution measured in the cases of replacement depending on the adopted technique for quality adjustment. Table 6 shows the quality correction coefficient (coefficient J in previous equations (3) to (14).

The price evolution for replaced products at time of replacement is different depending on the adjustment quality method (table 5). The average price evolution is increasing with the level of refinement of the quality adjustment method. It shows that the price evolution is improperly estimated, especially when no quality correction is applied. Nevertheless, standard deviations are not small enough to reject the equality of all the estimations.

Table 5: *[Yoghurts]* Price evolution between month m and m-1 for replaced products

Type of quality adjustment	Number of units (1)	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) No quality correction	42 703	-3.95%	19.8%	-35.8%	-1.6%	23.6%
(2) Link-to-show-no-price-change	42 703	0.00%	0.0%	0.0%	0.0%	0.0%
(3) Bridged overlap with real price increase	42 703	-0.40%	1.4%	-2.7%	-0.3%	1.8%
(4) Last month overlapping	42 703	-1.08%	8.5%	-15.2%	0.0%	11.7%
(5) Penultimate month overlapping	42 703	0.41%	13.7%	-18.0%	0.0%	22.0%
(6) Hedonic model	42 703	1.24%	14.1%	-16.7%	-0.1%	24.5%

Note: (*) the number of unit corresponds to the set of all replaced products in the union of all selected samples. Numbers in parenthesis correspond to the method used for quality adjustment (see sections 3.1 to 3.5)

We can see also a similar property on the estimated quality coefficients (see table 6): the level of quality is decreasing (not significantly) and all the methods conclude in the same direction. But refined methods conclude to a higher degree of correction than raw methods. Again, we cannot reject the equality of the estimated coefficients. This means that, statistically speaking, the different methods applied do not lead to distinct results.

To conclude this part, we show in table 7 the results for the distribution of the yearly index evolution for yoghurts. This distribution is computed from the elementary computation of 100 indices based on the same number of selected samples. This time, there is a significant difference (at the 95% level) between the index computed with no quality correction and the one computed with the hedonic model for quality adjustment. Consistently with the results obtained on the values of quality adjustments, the highest yearly evolution is obtained for Hedonic model. Nevertheless it is not possible to distinguish statistically the yearly evolution seen by one or the other indices when a quality correction is applied .

Table 6: *[Yoghurts]* Estimation of quality effect between month m and m-1 for replaced products

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) No quality correction	42 703	1.00	0.00	1.00	1.00	1.00
(2) Link-to-show-no-price-change	42 703	0.96	0.20	0.64	0.98	1.24
(3) Bridged overlap with real price increase	42 703	0.96	0.20	0.65	0.98	1.24
(4) Last month overlapping	42 703	0.97	0.18	0.70	1.00	1.17
(5) Penultimate month overlapping	42 703	0.96	0.19	0.64	0.98	1.21
(6) Hedonic model	42 703	0.95	0.19	0.65	0.98	1.19

Note: A coefficient greater (lower) than 1 means that the quality of the replacing product is higher (lower) than the quality of the replaced product.

Table 7: *[Yoghurts]* Price evolution between December 2008 and December 2009

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) No quality correction	100	-4.14%	0.19%	-4.5%	-4.1%	-3.8%
(2) Link-to-show-no-price-change	100	-3.55%	0.18%	-3.9%	-3.5%	-3.3%
(3) Bridged overlap with real price increase	100	-3.59%	0.18%	-3.9%	-3.6%	-3.3%
(4) Last month overlapping	100	-3.71%	0.17%	-4.0%	-3.7%	-3.4%
(5) Penultimate month overlapping	100	-3.60%	0.17%	-3.9%	-3.6%	-3.3%
(6) Hedonic model	100	-3.52%	0.17%	-3.8%	-3.5%	-3.2%

4.1.2 Quality adjustment for dissimilar products only

In this section, no quality correction is applied for quality-ranked 1 replacements. We remind that among all the situations of replacements, 73% are quality-ranked 1 (see table 2). Then for all these replacements, the J coefficient is set to 1. In addition, quality corrections are still applied (with the same level of correction) for all the rest of replacements, as in section 4.1.1.
The quality is rather decreasing (the average value of J is lower than 1 – see table 6). Setting the value of J equal to 1 for 73% of the replacements leads to increase the new average value of J (table 9). At the same time, the price of the replacing product is, on average, lower than the price of the replaced product (tables 5 and 8). When the price is decreasing, as well as the quality, the constant quality price decrease is smaller (in absolute value) than the decrease observed with no quality correction. Tables 5 and 8 exactly show this result and in this second computation, since the quality decrease is supposed to be smaller, then mechanically, the price decrease is amplified (in absolute value).
The difference in terms of price evolution seems to be high: the maximum is obtained for the quality adjustment based on Hedonic models. For replaced products only, the prices are increasing by +1.24% (table 5) while they seem to be decreasing by -2.95% in this new computation (table 8). Nevertheless, we cannot reject the equality of the two estimates of price variations.

Table 8: *[Yoghurts]* Price evolution between month m and m-1 for replaced products only

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) Equivalence and (2) link-to-show-no-price-change	42 703	-3.03%	17.3%	-32.2%	0.0%	18.8%
(1) Equivalence and (3) bridged overlap with real price increase	42 703	-3.21%	17.3%	-32.2%	-0.6%	18.8%
(1) Equivalence and (4) last month overlapping	42 703	-3.43%	17.6%	-32.2%	0.0%	19.3%
(1) Equivalence and (5) penultimate month overlapping	42 703	-2.90%	18.6%	-32.8%	-0.2%	21.9%
(1) Equivalence and (6) hedonic model	42 703	-2.95%	18.7%	-32.8%	-0.7%	22.9%

For indices (table 10), the consequence of a smaller quality adjustment is the same as for the replaced products: the yearly decrease of the indices is larger (in absolute value) in this new computation (table 10) than it was in the first (table 7). The largest difference is obtained in the case of Hedonic models where the decrease is -4.03% in the new computation; it was -3.52% in the first computation. Again, taking into account the distribution of the mean, we cannot reject the equality of the two indices.

Table 9: *[Yoghurts]* Estimation of quality effect between month m and m-1 for replaced products

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) Equivalence and (2) link-to-show-no-price-change	42 703	0.99	0.10	0.85	1.00	1.05
(1) Equivalence and (3) bridged overlap with real price increase	42 703	0.99	0.10	0.85	1.00	1.06
(1) Equivalence and (4) last month overlapping	42 703	0.99	0.10	0.88	1.00	1.06
(1) Equivalence and (5) penultimate month overlapping	42 703	0.99	0.11	0.84	1.00	1.06
(1) Equivalence and (6) hedonic model	42 703	0.99	0.11	0.84	1.00	1.06

Note: A coefficient greater (lower) than 1 means that the quality of the replacing product is higher (lower) than the quality of the replaced product.

Finally, to conclude this part, two remarks must be pointed out:

- From a statistical point of view, it is not possible to distinguish the results obtained for the yoghurt family with the various techniques of quality adjustment; nevertheless the direction of the possible bias is clear: whenever we go towards a more sophisticated technique of quality adjustment, the quality-corrected price variation is getting higher.
- There is clearly a quality bias that exists for what is called here "equivalent replacements": when a quality adjustment is done for these replacements, the price evolution is much higher (table A2) than when no adjustment is done (table A5). This is connected with the definition taken here for equivalent products: the algorithm selecting equivalent products is based on the proximity of products in terms of characteristics but the choice of these characteristics is, at the stage of the process, rather arbitrary. Therefore, one should not interpret the fact that there is a residual quality effect in an "equivalent replacement" as a problem. This rather shows that what is called here equivalent products are not purely equivalent and this should lead us to revise the way equivalent products are defined. In any case, this does not invalidate the relevance of the results obtained here, especially regarding the proximity of the various quality adjustment methods.

4.2 Simulations on Chocolate bars family

In this section, we test the methods of quality adjustment on the family of chocolate bars. Table 11 shows the average monthly price evolution observed over the whole sample of products

Table 10: *[Yoghurts]* Price evolution between December 2008 and December 2009

Method to estimate the difference of prices due to difference of quality	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) Equivalence and (2) link-to-show-no-price-change	100	-4.01%	0.18%	-4.3%	-4.0%	-3.7%
(1) Equivalence and (3) bridged overlap with real price increase	100	-4.04%	0.18%	-4.3%	-4.0%	-3.7%
(1) Equivalence and (4) last month overlapping	100	-4.07%	0.18%	-4.4%	-4.1%	-3.8%
(1) Equivalence and (5) penultimate month overlapping	100	-4.02%	0.19%	-4.3%	-4.0%	-3.7%
(1) Equivalence and (6) hedonic model	100	-4.03%	0.18%	-4.3%	-4.0%	-3.7%

between December 2008 and December 2009. Prices are increasing a bit and the distribution is much narrower than in the case of yoghurts: the 95% interval is [-5.5%; +5.3%] for chocolate bars; it was [-12.7%; +12.7%] for yoghurts.

Table 11: *[Chocolate bars]* Price evolution between month m and m-1 for non replaced products

Number of units (1)	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
4 807 744	0,054%	4.4%	-5.5%	0.0%	5.3%

Note: (1) the number of unit corresponds to the set of all non-replaced products in the union of all selected samples over 12 months.

4.2.1 Application of the same methods on all products replacements

In this section, we apply a quality adjustment in all cases of replacements. Table 12 shows the price evolution in case of replacement when various methods of quality adjustment are applied. This computation is made on the replaced products only. The correction is dominated by the noise: the 95% confidence intervals are quite large and consistent from the statistical point of view insofar their intersection is never empty.

The correction coefficient presented in table 13 is consistent with the previous result. One can notice that the result for quality effect estimates are very close, whatever the computation method is.

At the end (table 14), the quality-adjusted price indices are very consistent and look quite similar whatever the quality adjustment method is. The yearly evolution seen by the price index where no quality-adjustment is done is significantly different from the yearly evolution of corrected

Table 12: *[Chocolate bars]* Price evolution between month m and m-1 for replaced products only

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) No quality correction	145 856	7.78%	29.8%	-24.5%	1.5%	60.7%
(2) Link-to-show-no-price-change	145 856	0.00%	0.0%	0.0%	0.0%	0.0%
(3) Bridged overlap with real price increase	145 856	-0.04%	0.5%	-1.0%	0.0%	0.8%
(4) Last month overlapping	145 856	0.09%	5.9%	-7.6%	0.0%	6.3%
(5) Penultimate month overlapping	145 856	0.15%	11.6%	-15.3%	0.0%	14.0%
(6) Hedonic model	145 856	1.13%	13.5%	-17.8%	0.0%	22.8%

Note: (1) the number of unit corresponds to the set of all replaced products in the union of all selected samples.

Table 13: *[Chocolate bars]* Estimation of quality effect between month m and m-1 for replaced products

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) No quality correction	145 856	1.00	0.00	1.00	1.00	1.00
(2) Link-to-show-no-price-change	145 856	1.08	0.30	0.75	1.01	1.61
(3) Bridged overlap with real price increase	145 856	1.08	0.30	0.76	1.02	1.60
(4) Last month overlapping	145 856	1.08	0.29	0.77	1.01	1.60
(5) Penultimate month overlapping	145 856	1.08	0.30	0.78	1.02	1.61
(6) Hedonic model	145 856	1.07	0.29	0.78	1.02	1.58

Note: A coefficient greater (lower) than 1 means that the quality of the replacing product is higher (lower) than the quality of the replaced product.

indices. It shows that the quality adjustment is a more important issue for chocolate bars than for yoghurts. This is especially true when we realise that the part of corrected price transitions represents only[5] 3% of the number of price transitions used to compute the yearly evolution.

Table 14: *[Chocolate bars]* Price evolution between December 2008 and December 2009

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) No quality correction	200	1.90%	0.34%	1.4%	1.9%	2.5%
(2) Link-to-show-no-price-change	200	-0.23%	0.18%	-0.5%	-0.2%	0.1%
(3) Bridged overlap with real price increase	200	-0.24%	0.19%	-0.6%	-0.2%	0.1%
(4) Last month overlapping	200	-0.23%	0.18%	-0.5%	-0.2%	0.1%
(5) Penultimate month overlapping	200	-0.35%	0.18%	-0.7%	-0.4%	0.0%
(6) Hedonic model	200	-0.11%	0.18%	-0.4%	-0.1%	0.2%

4.2.2 Application of different methods for equivalent and dissimilar replacements

In this section, no quality correction is applied for quality-ranked 1 replacements. We remind that among all the situations or replacements, 56% are quality-ranked 1 (see table 2). Then for all these replacements, the J coefficient is set to 1. In addition, quality corrections are still applied (with the same level of correction) for all the rest of replacements, as in section 4.2.1. With less quality correction, the prices are, for replaced products only, rather increasing (table 15 compared to table 12). This can also be seen from table 13 which shows that the quality is increasing, but with a less important magnitude when the quality-ranked 1 replacements are assumed to involve two products with the same level of quality (equivalent products). The consequence is that an increase of price that was seen as an increase in quality is now seen as a pure price increase. That is why we find that the average price evolution is higher now (table 15) than it was in section 4.2.1 (table 12). All the methods of quality adjustment lead to the same estimate of quality effect (table 16).

Table 17 shows the results on yearly prices indices: even if the average price evolution seems to be different from one method of quality adjustment to another, the 95% confidence intervals intersect each other. Interestingly, the price evolution is significantly different when quality-ranked 1 replacements are set to be equivalent in quality (table 17) or when they are quality adjusted (table 14). It seems that, like in the yoghurts case, the quality-ranked 1 replacements are not at all equivalent. All the methods used for quality adjustment lead to similar results.

[5]=145 856/(4 807 744+145 856), see tables 12 and 13.

Table 15: *[Chocolate bars]* Price evolution between month m and m-1 for replaced products only

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) Equivalence and (2) link-to-show-no-price-change	145 856	4.55%	20.2%	-21.6%	0.0%	33.5%
(1) Equivalence and (3) bridged overlap with real price increase	145 856	4.52%	20.2%	-21.6%	0.1%	33.5%
(1) Equivalence and (4) last month overlapping	145 856	4.57%	20.6%	-22.1%	0.0%	34.9%
(1) Equivalence and (5) penultimate month overlapping	145 856	4.97%	21.6%	-22.9%	0.0%	38.3%
(1) Equivalence and (6) hedonic model	145 856	5.19%	22.1%	-23.4%	1.7%	39.1%

Table 16: *[Chocolate bars]* Estimation of quality effect between month m and m-1 for replaced products

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) Equivalence and (2) link-to-show-no-price-change	145 856	1.03	0.23	0.90	1.00	1.14
(1) Equivalence and (3) bridged overlap with real price increase	145 856	1.03	0.23	0.90	1.00	1.14
(1) Equivalence and (4) last month overlapping	145 856	1.03	0.22	0.92	1.00	1.12
(1) Equivalence and (5) penultimate month overlapping	145 856	1.03	0.23	0.88	1.00	1.19
(1) Equivalence and (6) hedonic model	145 856	1.03	0.23	0.86	1.00	1.21

Note: A coefficient greater (lower) than 1 means that the quality of the replacing product is higher (lower) than the quality of the replaced product.

Table 17: *[Chocolate bars]* Price evolution between December 2008 and December 2009

Type of quality adjustment	Number of units	Average	Std-Dev	Perc. 5	Perc. 50	Perc. 95
(1) Equivalence and (2) link-to-show-no-price-change	200	1.09%	0.26%	0.7%	1.1%	1.6%
(1) Equivalence and (3) bridged overlap with real price increase	200	1.08%	0.26%	0.6%	1.1%	1.6%
(1) Equivalence and (4) last month overlapping	200	1.08%	0.25%	0.7%	1.1%	1.5%
(1) Equivalence and (5) penultimate month overlapping	200	1.14%	0.25%	0.7%	1.1%	1.6%
(1) Equivalence and (6) hedonic model	200	1.18%	0.25%	0.8%	1.2%	1.6%

5 Conclusion

We have shown the possible use that could be done of scanner data when an additional set of variables describing the EAN is available. Indeed, this additional set makes it possible to follow exactly the concepts of the traditional Laspeyres type index, while improving their realisation. For studied food products (yoghurts and chocolate bars), while a raw algorithm for selecting replacing goods in case of replacement is used, quality effects are significant, even when the number of replacements is small. Indeed, non quality-adjusted indices are significantly different from the quality-adjusted indices. In that case, Hedonic method still appears as the reference method, but overlapping or even bridged overlap with real price increase methods do not lead to significantly different price indices. Nevertheless, evidence suggests that the quality effect is slightly overestimated (and therefore the pure price increase is slightly underestimated) with the latter two methods. From a practical point of view, all the methods that rely on the traditional economic model of consumer theory might be considered in scanner data analysis, the overlapping being the simplest one.

References

Ahnert, H. & Kenny, G. (2004). Quality Adjustment of European Price Statistics and the Role for Hedonics, *Occasional paper 15*, European Central Bank.

Armknecht, P. & Moulton, B. (1995). Quality Adjustment in Prices Indices: Methods for imputing Price and Quality Change, *2nd Ottawa Group Meeting*.

Armknecht, P., Moulton, B. & Stewart, K. (1994). Improvements to the Food at home, Shelter and Prescription Drug indexes in the U. S. Consumer Price Index, *1st Ottawa Group Meeting*.

Center of Excellence Network - CENEX (2008). HICP Quality Adjustment Handbook, *Manual*, EUROSTAT.

Deaton, A. & Muelbauer, J. (1980). *Economics and consumer behavior*, Cambridge University Press.

ILO, IMF, OECD, UNECE, Eurostat & The World Bank (2004). *Consumer price index manual : Theory and Practise*, International Labour Office.

Silver, M. & Heravi, S. (2000). The Measurement of Quality-Adjusted Price Changes, *in* R. C. Feenstra & M. D. Shapiro (eds), *Scanner Data and Price Indexes*, National Bureau of Economic Research.

Silver, M. & Heravi, S. (2005). A failure in the measurement of inflation: Results from a hedonic and matched experiment using scanner data, *Journal of Business & Economic Statistics* **23**(3): 269–281.

Triplett, J. (2006). Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes, *Technical report*, OECD.

A Quality and Constant Utility Framework

This section follows mainly the classical approach of constant quality indices presented in the framework of the Constant Utility price index proposed by Deaton & Muelbauer (1980).
We assume that the price index consists in following a set of products which cover a set of needs. These products are consumed by a representative consumer who decides on the quantities of products he buys after having optimised a utility function. This utility function operates on the vector of quantities referring to the previous set of products. Let us write u the utility function[6] and $\mathbf{q}$ the vector[7] of quantities of products bought at any period[8] of time.
The problem of consumer is the following : its Marshallian demand is

$$\mathbf{x}_u(\mathbf{p}; R) = \underset{\mathbf{q}}{\operatorname{argmax}} \{u(\mathbf{q}) | \mathbf{p}.\mathbf{q} = R\}$$

and the expense function, dual of the previous one, is :

$$e_u(\mathbf{p}, \bar{u}) = \min_{\mathbf{q}} \{\mathbf{p}.\mathbf{q} | u(\mathbf{q}) = \bar{u}\}$$

Since the two problems are dual of each other, we have the usual relationship between the two : the expense associated to the optimised basket given by the Marshalian demand is equal to the expense function taken at this optimised basket utility point, the same price vector being taken for the two problems. Formally, this is written :

$$\mathbf{p}.\mathbf{x}_u(\mathbf{p}, R) = e_u(\mathbf{p}, u(\mathbf{x}_u(\mathbf{p}, R)))$$

In this model, the Utility constant price index, by definition, follows the evolution of the budget the consumer needs to assume in order to keep its utility constant between the two periods of comparison 0 and t :

$$I_t^0 = \frac{e_u(\mathbf{p}_t, u(\mathbf{x}_u(\mathbf{p}, R_0)))}{R_0}$$

while the corresponding level of utility is the one reached for the Marshallian demand at period 0, equal to $u(\mathbf{x}_u(\mathbf{p_0}; R_0))$. We note $\bar{u}_0$ this level of utility.

Consider now that for a reason or another, the need 1 is no longer covered by one good but by another one. This old good is replaced by the consumer by a new one in order to cover the same need. Since the two goods are not the same, we cannot assume that, from the consumer point of view, consuming 1 unit of the old good brings the same amount of utility as consuming 1 unit of the new good. Therefore, it is reasonable, in order to model this situation, to consider that for the need number 1, 1 unit of the new good generates an amount of utility equal to α units of the old one. This might be translated in the following way : let us assume that after the replacement, the utility function from which the consumer takes the decision to consume is

[6] We assume that the function is increasing with respect to all its components.

[7] In this appendix, vectors are in bold letters.

[8] We then assume that the utility function does not change from time to time. But prices change, then quantities as well.

v. This function operates on the same set of products covering the same needs, except form the first good which is now the new one. The relationship between u and v is therefore :

$$v(q_1, \mathbf{q}_{(1)}) = u(\alpha q_1, \mathbf{q}_{(1)})$$

where $\mathbf{q}_{(1)}$ stands the vector $\mathbf{q}$ where the first component has been removed. In the previous expression, everything else being equal, if $\alpha > 1$ then the quality of the new good is greater than the quality of the old good, and vice-versa when $\alpha < 1$. While the consumer optimises from utility function v, its expense function becomes:

$$\begin{aligned} e_v(\mathbf{p}, \bar{v}) &= \min_{\mathbf{q}} \{\mathbf{p}.\mathbf{q} | v(\mathbf{q}) = \bar{v}\} \\ &= \min_{\mathbf{q}} \left\{\mathbf{p}.\mathbf{q} | u(\alpha q_1, \mathbf{q}_{(1)}) = \bar{v}\right\} \end{aligned}$$

The first order conditions of the previous problem can be written (there are n needs covered by the basket ; λ is the Lagrange multiplier)

$$\left\{ \begin{array}{rcl} p_1 &=& \lambda\alpha\partial_1 u(\alpha q_1, \mathbf{q}_{(1)}) \\ p_2 &=& \lambda\partial_2 u(\alpha q_1, \mathbf{q}_{(1)}) \\ \vdots & & \vdots \\ p_n &=& \lambda\partial_n u(\alpha q_1, \mathbf{q}_{(1)}) \\ \bar{v} &=& u(\alpha q_1, \mathbf{q}_{(1)}) \end{array} \right.$$

Let us define $\mathbf{r} = (\alpha q_1, \mathbf{q}_{(1)})$. Then, by construction,

$$\left\{ \begin{array}{rcl} p_1 &=& \lambda\alpha\partial_1 u(\mathbf{r}) \\ p_2 &=& \lambda\partial_2 u(\mathbf{r}) \\ \vdots & & \vdots \\ p_n &=& \lambda\partial_n u(\mathbf{r}) \\ \bar{v} &=& u(\mathbf{r}) \end{array} \right.$$

$\mathbf{r}$ appears as being equal to $\min_{\mathbf{r}} \{\boldsymbol{\pi}.\mathbf{r} | u(\mathbf{r}) = \bar{v}\}$ where $\boldsymbol{\pi}$ is a vector of prices defined by $\boldsymbol{\pi} = \left(\frac{p_1}{\alpha}, \mathbf{p}_{(1)}\right)$. Finally,

$$e_v(\mathbf{p}, \bar{v}) = e_u\left(\left(\frac{p_1}{\alpha}, \mathbf{p}_{(1)}\right), \bar{v}\right) \tag{15}$$

If we consider a certain level of utility, $\bar{u}$ and if we assume that the perceived quality of the new good is greater than that of the old good ($\alpha > 1$), then we find that if the price vector is unchanged, the expense necessary to reach $\bar{u}$ is lower in the case of the new good than in the case of the old good :

$$\alpha < 1 \Rightarrow e_u\left(\left(\frac{p_1}{\alpha}, \mathbf{p}_{(1)}\right), \bar{u}\right) \leqslant e_u\left(\mathbf{p}, \bar{u}\right)$$

because e_u is an increasing function with respect to each price components.

The main consequence of equation (15) for our problem is that it is possible to correct prices in order to stay in a constant utility framework even when we allow the goods of the basket to change : even if we don't know the utility function nor the coefficient (α) scaling the quantities

in this utility function in case of a quality change, it is possible to correct the vector of price for new goods in order to stay on the same curve of utility. The correction that should be adopted on the price of new goods is exactly the α-coefficient that scales the quantities in the utility function.

For a market in equilibrium, the scale of prices reflects the price differences that make the consumer indifferent to consume a product or another one. If we can assume that the equilibrium is respected at each period of time, and if the old good and the new one are sold at the same time, then the price ratio reflects the α coefficient. In other words, the ratio of prices is exactly the value that the consumer gives to the difference in product characteristics.

One might argue about the robustness of the idea of market equilibrium. But from the economic point of view, the overlapping technique or the hedonic models both rely on this idea. The only difference between the two is that with respect to this idea, the hedonic approach should be more robust in the sense that it is less sensitive to stochastic perturbation that could occur at the level of the product price. Indeed, the hedonic technique aims at determining the average link that exists between the prices and the characteristics of the products and not the link between two specific products.

B Econometrics of Hedonic models

The dependent variable in models is the logarithm of the price per unit (the unit is the gram). The calculations are performed over 14 months of observations (from November 2008 to December 2009). A dummy variable for each month is added to the model to take into account price changes during the year. Most of variables characterizing the products are discrete (categorical) ones. The values of some discrete explanatory variables are highly correlated so an additive model is not adequate. For example, only the brand Danacol contains anti-cholesterol. This defect was corrected by introducing into the model, instead of the dummy variables of characteristics, all possible crossings of these variables (interactions). We may introduce in the model all the interactions, even if they are numerous (1 642 for yoghurts, 1 149 for chocolate bars), because the number of observations is huge (1.8 million for yoghurts, 1.1 millions for chocolate bars). The structure of models is common to both product families (i is a considered series and t is the time):

$$\ln(p_{it}) = c + \sum_{k=1}^{K} \beta_k . \mathbf{1}_{i \in k} + \sum_{m=1}^{13} \gamma_m . \mathbf{1}_{t \in m} + \varepsilon_{it} \tag{16}$$

where β_k is the k^{th} interaction coefficient (there are K possible values), $\mathbf{1}$ is equal to 1 when the condition is true and 0 otherwise. γ_m is the month m fixed effect. 14 months of data are used (1 month is taken as reference).

An interaction corresponds, for example in the yoghurt case, to a natural non-organic unsweetened yoghurt, with bifidus, full milk and a normal fat content, farm, without additives, which brand is Activia, sold in the chain "Auchan" and whose total net weight is 500g. With these

notations, all possible crossings of characteristics are introduced. p_{it} is the price per gram of the product sold (i) at week t.
For some discrete variables, some values are grouped because there is a very small number of corresponding series. For example, the overall volume for yoghurt takes 42 different values (500g, 1000g...). We treat this variable as discrete variable limited to 5 different values (<500g, 500g,> 500g and <= 1000 g,> 1000 g and <= 1500 g,> 1500 g). Missing values were recoded. For example, the missing values of the variable "Additives" were recoded into "No additives".

B.1 Model for yoghurts

The model is computed ones for the whole family of products. The variables used in the computation are given in table 18.

Table 18: *[Yoghurts]* Variables used in the Hedonic model

Name of the variable	Number of categories
Store Chain	7
Brand	19
Type of pack	5
Perfume variety	19
Active ingredient	4
Biological information	3
Content of milk	4
Additives	8
Sugar content	3
Process	6
Fat content	2
Type of yoghurt	3
Overall volume	5
13 variables	88 categories

These 13 variables are then crossed to obtain all possible interactions (1 642). The regressions have been done on these 1 642 interactions. The results are as follows are given in table 19. The last part of the table shows the estimated values for the coefficient γ_m (month m fixed effect). We see that this coefficient captures essentially the time trend (decrease) of prices: the mean level of prices in November 2009 is about 5.2% lower than in December 2008 (reference month).

B.2 Model for chocolate bars

As for the yoghurt model, discrete variables are sometimes reshaped (modalities grouped, missing values coded). For example, the brand variable contains 274 different values. Most of them

Table 19: *[Yoghurts]* Hedonic model results

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1 654	202 722,190	122,565	12 416,5	<.0001
Error	1 800 000	17 776,947	0,010		
Corrected Total	1 800 000	220 499,136			

R-square	Coeff Var	Root MSE	Distance Mean	Number of obs
0,919379	-7,441059	0,099354	-1,335207	1 802 558

Source	DF	Type III SS	Mean Square	F Value	Pr > F
crossings	1 641	201 844,2	123,00	12 460,7	<.0001
months	13	449,0	34,54	3 499,2	<.0001

Parameter	Estimated value	Standard-dev	T-test value	Pr > |t|
month 200811	0,006	0,00039	15,89	<.0001
month 200812	0,000	.	.	.
month 200901	-0,003	0,00039	-8,36	<.0001
month 200902	-0,004	0,00039	-10,21	<.0001
month 200903	-0,013	0,00039	-32,54	<.0001
month 200904	-0,001	0,00039	-2,8	0,005
month 200905	-0,015	0,00039	-38,71	<.0001
month 200906	-0,025	0,00039	-64,57	<.0001
month 200907	-0,024	0,00039	-60,63	<.0001
month 200908	-0,023	0,00040	-56,88	<.0001
month 200909	-0,021	0,00040	-54,05	<.0001
month 200910	-0,041	0,00040	-104,02	<.0001
month 200911	-0,052	0,00039	-131,76	<.0001
month 200912	-0,036	0,00040	-91,18	<.0001

correspond to sub-brand. For example, we grouped all sub-brand Lindt into one Lindt brand. The variables used in the computation are given in table 20.

Table 20: *[Chocolate bars]* Variables used in the Hedonic model

Name of the variable	Number of categories
Store chain	7
Brand	11
Type of product	6
Type of pack	6
Perfume variety	6
Biological information	3
Fair Trade Information	2
Number of parts	7
Size	5
Additives	12
Overall Volume	5
11 variables	70 categories

These 11 variables are then crossed to obtain all possible interactions (1 137). The regressions have been done on these 1 137 interactions. The results are as follows are given in table 21. The last part of the table shows the estimated values for the coefficient γ_m (month m fixed effect). It seems that there are very small seasonal variations: at the end, the mean level of prices in December 2009 is 0.2% lower than in December 2008. No time trend is observable in this case, by opposition to the yoghurt case.

Table 21: *[Chocolate bars]* Hedonic model results

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1 149	240 154	209,011	20 905,7	<,0001
Error	1,15E6	11 510	0,010		
Corrected Total	1,15E	251 664			

R-square	Coeff Var	Root MSE	Distance Mean	Number of obs
0,954263	54,97	0,0999	0,18189	1 152 427

Source	DF	Type III SS	Mean Square	F Value	Pr > F
crossings	1136	240 142,6	211,39	21 144,0	<.0001
months	13	11,3	0,87	87,03	<.0001

Parameter	Estimated value	Standard -dev	T-test value	Pr > |t|
month 200811	-0,0011	0,00050	-2,28	0,0224
month 200812	0,000	.	.	.
month 200901	-0,0012	0,00050	-2,31	0,0210
month 200902	0,0023	0,00050	4,58	<,0001
month 200903	-0,0005	0,00050	-1,03	0,3026
month 200904	0,0005	0,00050	1,02	0,3067
month 200905	-0,0074	0,00050	-14,88	<,0001
month 200906	-0,0001	0,00050	-0,22	0,8248
month 200907	0,0031	0,00050	6,08	<.0001
month 200908	0,0029	0,00050	5,82	<.0001
month 200909	0,0020	0,00050	3,97	<.0001
month 200910	-0,0019	0,00050	-3,81	0.0001
month 200911	-0,0073	0,00050	-14,67	<.0001
month 200912	-0,0020	0,00050	-3,94	<.0001