Scanner data and quality adjustment

Isabelle Léonard, Patrick Sillard, Gaëtan Varlet and Jean-Paul Zoyem

INSEE – France

May 2015

Plan

Plan

Aim of the paper (1)

Scanner data in France

- daily files
- one line = (day, outlet, EAN, quantity sold, turnover) ; [*Rem* : EAN=barcode]
- additional file containing EAN characteristics
- at the moment, data received by INSEE cover 30% of the "potential market"

It allows us

to treat scanner data as a *simple* additional source of data, without changing the CPI "traditional" concepts

Aim of the paper (2)

Since Scanner data are an additional source of data...

- ... it is the occasion to review the current processes going back to the concepts
- Focus on quality adjustment
- Results obtained on 13 typical food products and manufactured good families : yogurts, chocolate bars, 4 different families of cheese, chicken eggs, frozen pizzas, toilet tissue, fruit juice, 3 different families of ground coffee.

Some (secondary) points... hold fixed throughout the paper

- Daily data are aggregated over a month : unit value prices and expense are computed for each cell (=EAN×outlet×month)
- Base weights are computed with expense of November and December 2008 ; indices are computed over 2009 with respect to December 2008
- Index computed from a sample selection (rate of sampling=2%=20 times the current CPI rate) with a probability of inclusion proportional to the expense in base period
- The micro indices are Jevons at the level of (outlet×variety of product) and Laspeyres-aggregated to a national level per family of products

Plan

The life cycle of a product

The equivalent class (EC) at a given date

☞ EAN is not equivalent to the traditional good followed in the CPI :

- $1 \leftrightarrow N$ relationship : more than one EAN in a given shop may refer to the same good in the CPI sense
- The equivalent class (EC) contains all the EAN referring to the same good (seen as so by the consumer) ; in particular it contains promotion articles
- Most of time, the EC contains a little bit more than 1 EAN : 1.3 on average

The product is a time series of EC's

- of coincident ($1 \leftrightarrow 1$ relationship at a given date) EC
- when the coincident EC disappears, a replacing one is used as the product representative at the considered date
- At time of replacement, a quality adjustment is done

The choice of the replacing good

The central algorithm

When an equivalent class (EC) disappears, we select the EC class according to the following criteria :

1. same outlet, variety, brand, range of volume
2. same outlet and variety
3. same city and variety

The alternative algorithm

When an EC disappears, we select the replacing EC **randomly** among the goods of the same variety in the same outlet

Quality adjustment methods : practice

Tested methods

- No adjustment : prices are assumed to be directly comparable
- Link-to-show-no-price-change
- Bridged overlap with real price increase
- Last month overlapping
- Penultimate month overlapping
- Hedonic model

Quality adjustment methods : theory

In a Cost of Living (COLI) approach of CPI

☞ All these methods are based on the same underlying idea :

- A difference in *utility* might be taken into account, in the constant-quality CPI, through a *price correction*
- In the Constant Utility CPI (i.e. COLI) framework, the price correction appears as a measure of a shift in the marginal utility w.r.t. the considered *need* (=product).
- This shift appends in case of replacement because the *same* need is covered by two *different* goods (EC) before and after replacement.

☞ The various quality adjustment methods illustrate various *conventional* ways to estimate this shift.

Reformulation of the goal of the paper

To see

- if the application of one or another method of quality adjustment leads to statistically different results
- if the choice of the replacing product (equivalence class) has an impact on the result

Plan

Monte-Carlo Simulation

We select several samples (30 to 200 depending on the family) in the universe of products

- this allows to compute the mean index
- and confidence intervals (95%)

A point to keep in mind

The replacements are not very frequent : they represent from 1% up to 8% of month-to-month price transitions (for a given outlet), depending on the family of products

Plan

The index : Yogurt Mean annual increase over 2009

Family of yogurts

Type of quality adjustment	Central algorithm	Alternative algorithm
(1) No quality correction	−4.14% [−4.5% , −3.8%]	−3.17% [−3.6% , −2.7%]
(2) Link-to-show-no-price-change	−3.55% [−3.9% , −3.3%]	−3.51% [−3.8% , −3.2%]
(3) Bridged overlap with real price increase	−3.59% [−3.9% , −3.3%]	−3.56% [−3.8% , −3.2%]
(4) Last month overlapping	−3.71% [−4.0% , −3.4%]	−3.60% [−3.9% , −3.3%]
(5) Penultimate month overlapping	−3.60% [−3.9% , −3.3%]	−3.51% [−3.8% , −3.2%]
(6) Hedonic model	−3.52% [−3.8% , −3.2%]	−3.52% [−3.8% , −3.2%]

The index : Chocolate bars mean annual increase over 2009

Family of chocolate bars

Type of quality adjustment	Central algorithm	Alternative algorithm
(1) No quality correction	1.90% [1.4% , 2.5%]	9.10% [7.5% , 10.6%]
(2) Link-to-show-no-price-change	−0.23% [−0.5% , 0.1%]	−0.03% [−0.3% , 0.3%]
(3) Bridged overlap with real price increase	−0.24% [−0.6% , 0.1%]	−0.04% [−0.3% , 0.3%]
(4) Last month overlapping	−0.23% [−0.5% , 0.1%]	−0.03% [−0.3% , 0.3%]
(5) Penultimate month overlapping	−0.35% [−0.7% , 0.0%]	0.01% [−0.3% , 0.3%]
(6) Hedonic model	−0.11% [−0.4% , 0.2%]	0.13% [−0.1% , 0.4%]

Bias or not bias ? (1)

[yogurts] Difference of estimated quality factors by Hedonic's and the other techniques

Type of quality adjustment	Difference in Mean with respect to Hedonic's
(2) Link-to-show-no-price-change	−0.006 [−0.017 , 0.003]
(3) Bridged overlap with real price increase	−0.010 [−0.020 , −0.001]
(4) Last month overlapping	−0.016 [−0.024 , −0.009]
(5) Penultimate month overlapping	−0.008 [−0.016 , −0.001]

Quality factor : the correction, proportional to the price, that should be applied to the replaced good to reach the level of quality of the replacing.

Bias or not bias ? (2)

For chocolate bars : same results

- Results consistent throughout the different quality adjustment methods
- Alternative algorithm of replacement selection do not lead to different results

[Chocolate bars] Difference of estimated quality factors by Hedonic's and the other techniques

Type of quality adjustment	difference in Mean with respect to Hedonic's
(2) Link-to-show-no-price-change	−0.004 [−0.012 , 0.003]
(3) Bridged overlap with real price increase	−0.005 [−0.013 , 0.002]
(4) Last month overlapping	−0.004 [−0.011 , 0.003]
(5) Penultimate month overlapping	−0.009 [−0.015 , −0.004]

Same situation for the 11 other families of products

- The choice of replacing product has no consequence provided that the quality adjustment is made properly
- A quality adjustment is necessary : indices not corrected for quality are significantly different, even for the studied families where we could imagine that quality effects are small
- At the level of precision reached here (the samples are here 20 times larger than in the current French CPI), it is not possible to identify a bias of a quality correction method at the level of the index ; nevertheless when we look at the quality corrected price transitions (representing 1 to 8% of the price transitions), there are evidence that a (small) bias may exist with respect to Hedonic's.

Plan

Conclusion

The resulting index is not sensitive to the algorithm of replacement

Whatever the choice of replacing product is (at least to the extent of variety), if quality adjustment is well done, then the resulting index is a quality constant price index

But...

Quality adjustment methods, at the product level, are not fully consistent, even in mean. So there are (small) biases.

Finally,

If the number of replacements is small and if the goals in terms of index accuracy are not too high (sample of reasonable size), the signal to noise ratio is favorable and the quality adjustment biases are diluted into the pure price noise (coming from the sampling). **In this case, the differences that may be observed at the level of the computed indices are not statistically significant (i.e. no bias).**