Scanner data and quality adjustment

Isabelle Léonard, Patrick Sillard and Gaëtan Varlet

INSEE – France

September 2013

Plan

1. Introduction
2. Replacements
3. Method
4. Results
5. Conclusion

Plan

1. Introduction
2. Replacements
3. Method
4. Results
5. Conclusion

Aim of the paper (1)

Scanner data in France

- daily files
- one line = (shop,EAN,quantities sold, turnover, date)
- additional file containing EAN characteristics
- at the moment, 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, as an illustration, on two families of products : yoghurts and chocolate bars

Some (secondary) points...

- daily data are aggregated over a week : quantities and turnover are added over a week
- base weights are computed with spendings 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 spendings in base period
- the micro indices are Jevons at the level of (shop×variety of product) and Laspeyres-aggregated to a national level per family

Plan

1 Introduction

2 Replacements

3 Method

4 Results

5 Conclusion

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 he consumer) ; in particular it contains promotion articles
- Most of time, the EC contains only 1 good (1.3 on average)

The product is a time series

- 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

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 variety, city, outlet

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

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

All these methods

Are based on the same underlying idea that a difference in *utility* might be taken into account, in the constant-quality CPI, through a price correction ; the chosen methods illustrate different *conventional* ways of estimating this correction.

Goal

To see

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

Plan

1 Introduction

2 Replacements

3 Method

4 Results

5 Conclusion

Simulation

We select several samples (100 to 200) in the universe of products

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

Some additional issues

- Two different algorithms for the choice of replacing products : a central one and an alternative one
- the replacement are not frequent : 1% for yoghurts and 3% for chocolate bars

Plan

1 Introduction

2 Replacements

3 Method

4 Results

5 Conclusion

The index : Mean annual increase over 2009

Family of yoghurts

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%]

So, bias or not bias ? (1)

[Yoghurts] 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.

So, 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]

Plan

1 Introduction

2 Replacements

3 Method

4 Results

5 Conclusion

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 biases)
- If the number of replacement 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 favourable and the quality adjustment biases are diluted into the pure price noise (coming from the sampling).