SCANNER DATA: TOWARDS CONSTANT UTILITY INDICES?

Patrick Sillard

Division des prix à la consommation

INSEE
INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

01/06/2012

Introduction

THIS PAPER…

- Is a tentative paper

  -to illustrate the use that can be done of scanner data to **realize** in a better way the underlying economic concepts of price indexes

  -it is then for us a mean to look at these concepts in order to analyze in detail the use of scanner data we must do to stay consistent with the current CPI framework (the French CPI is **not** intended to realize a constant utility index concept)

  -the scale of interest is the one where substitutions can occur => micro-aggregates computation

- Is not what we intend to do in our CPI
  - Since the French index (or even the European HICP) is not intended to realize a constant utility index, we do not plan to apply these ideas to our CPI
  - But it helps us to understand better what we can do with scanner data

Organisation of the presentation

1. Constant utility indexes
2. Data used
3. Application and results

Constant Utility Indexes

Expenditure at constant utility

- Expenditure function: $e(\mathbf{p}, U) = \left| \begin{array}{l} \min_{\mathbf{s}} \mathbf{p}.\mathbf{s} \\ \text{s.c. } u(\mathbf{s}) = U \end{array} \right.$

  Vector of basket quantities

- Indirect Utility : $v(\mathbf{p}, R)$. It is the utility level reached for a certain amount of money R (expenditure)

- Constant utility index (t': current period; t: reference period)

$$I_{UC}^{t',t}(R_t) = \frac{e(\mathbf{p}_{t'}, v(\mathbf{p}_t, R_t))}{R_t}$$

Price vector in t'

Utility level reached in t when the prices are $\mathbf{p}_t$ and the Expenditure is R_t

DERIVATING AN INDEX FROM A GIVEN UTILITY FUNCTION

- Utility maximation subject to a budgetary constraint

	CES	Cobb-Douglas	Leontief
Utility: $u(\mathbf{s}) =$	$\left(\sum_k \alpha_k s_k^{(\varepsilon-1)/\varepsilon}\right)^{\varepsilon/(\varepsilon-1)}$	$s_1^{\alpha_1} \times \ldots \times s_n^{\alpha_n}$	$\min\{\alpha_1 s_1, \ldots, \alpha_n s_n\}$
Demand: $x_i(\mathbf{p}, R) =$	$\frac{R}{p_i} \frac{\alpha_i \left(\frac{p_i}{\alpha_i}\right)^{1-\varepsilon}}{\sum_k \alpha_k \left(\frac{p_k}{\alpha_k}\right)^{1-\varepsilon}}$	$\alpha_i \frac{R}{p_i}$	$R/\left(\alpha_i \sum_k \frac{p_k}{\alpha_k}\right)$
Index: $I^{t',t} =$	$\frac{\left(\sum_k \alpha_k \left(\frac{p_k^{t'}}{\alpha_k}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}{\left(\sum_j \alpha_j \left(\frac{p_j^t}{\alpha_j}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}$	$\prod_{i=1}^n \left(\frac{p_i^{t'}}{p_i^t}\right)^{\alpha_i}$	$\left(\sum_i \frac{p_i^{t'}}{\alpha_i}\right) / \left(\sum_j \frac{p_j^t}{\alpha_j}\right)$

DERIVATING AN INDEX FROM A GIVEN DEMAND FUNCTION

- We define the money metric indirect utility function (numerator of the CUI): $\mu(\mathbf{p};\mathbf{q},R) = e(\mathbf{p}, v(\mathbf{q},R))$
- The money metric indirect utility function [μ] and the demand function [$\mathbf{x}(\mathbf{p},R)$] are linked by a differential equation in μ whose variables are $\mathbf{p}$

$$\forall i \in \{1,\ldots,n\}\ ,\ x_i(\mathbf{p},\mu(\mathbf{p};\underbrace{\mathbf{q},R}_{\text{parameters}})) = \frac{\partial \mu(\mathbf{p};\mathbf{q},R)}{\partial p_i}$$

ALL DEMAND FUNCTION DOES NOT DERIVATE FROM A UTILITY OPTIMIZATION

- The demand function must follow some integrability conditions (Hurwicz et Uzawa, 1971)
- Example: let us consider a log-linear demand form

$$x_i(\mathbf{p}, R) = p_i^{\alpha_i} R^{\beta_i} e^{\gamma_i}$$

- In order to follow integrability conditions: $\beta_i \equiv \beta$

- The Constant Utility Index is then :

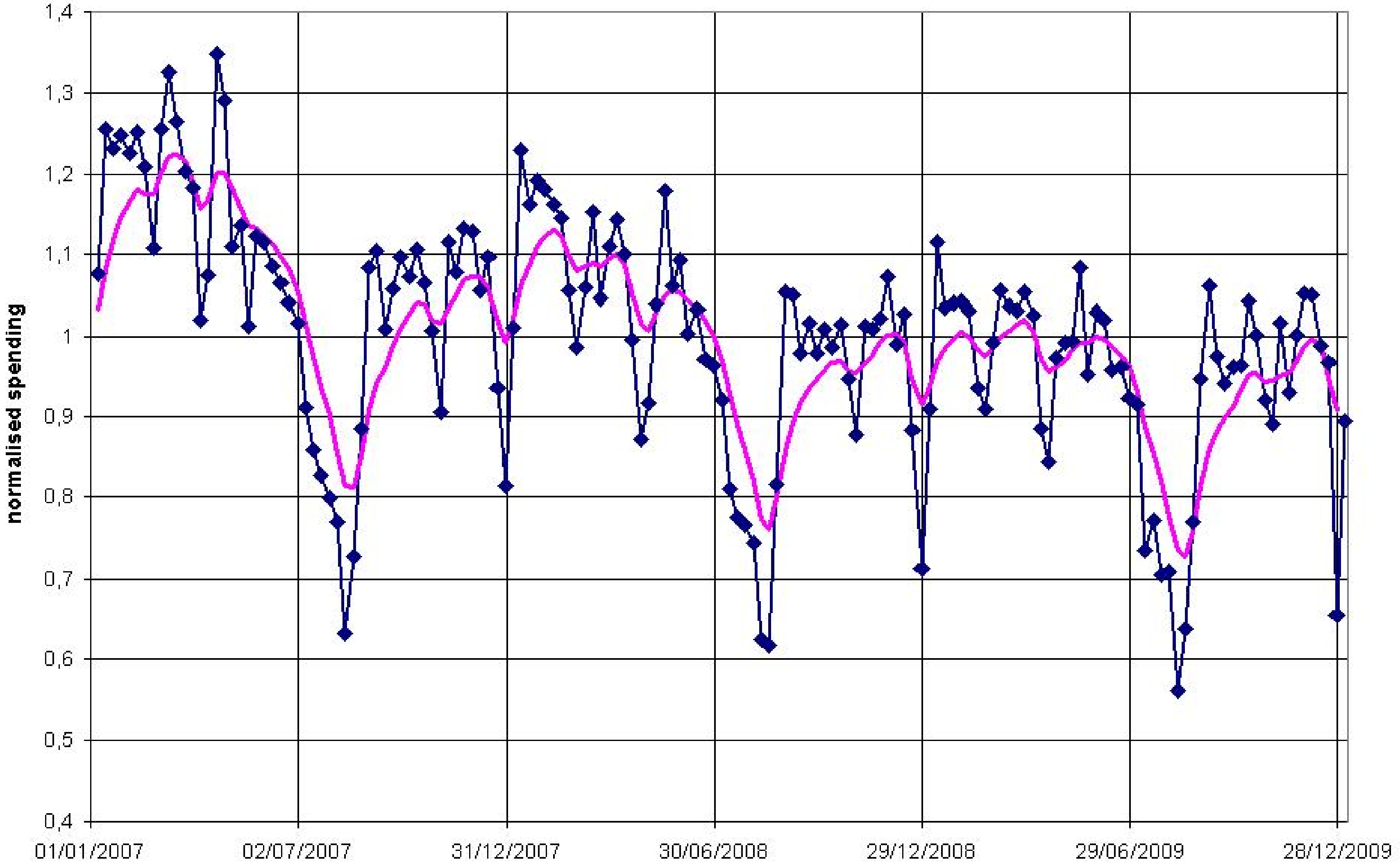
$$I_{LogLin}^{t',t} = \left\{ 1 + (1-\beta) R_t^{\beta-1} \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_{it'}^{1+\alpha_i} - p_{it}^{1+\alpha_i}}{1+\alpha_i} \right\}^{\frac{1}{1-\beta}}$$

The data

THE DATA

- The consumer theory is applicable on a field where the frame of a representative consumer is meaningful (especially with respect to the question of substitutable products)
- =>1 type of product in a given shop
- We consider here in the family of yoghurts sold in a given shop between 2007 and 2009:
  - About 600 barcodes
  - 35 000 weekly price and quantity observations
  - Each barcode is sold, in mean, 950 times on the period (median=300, D1=30, D9=2300)

SHOP TURNOVER FOR YOGHURT PRODUCT FAMILY



Results

THREE TYPES OF MONTHLY INDICES COMPUTED

- Chained Laspeyres
- CES index
- Index based on Log-linear demand (called after Log-linear demand index)

THE CHAINED LASPEYRES INDEX

- Formula (t': current; t: base):

$$I^{t',t} = \left(\sum_i \frac{p_i^{t'}}{\alpha_i}\right) / \left(\sum_j \frac{p_j^{t}}{\alpha_j}\right)$$

- Annually chained (like in the French CPI)
- Base: mean over January (t in previous formula)
- Fixed basket without replacement (missing product prices are filled by applying the evolution computed over present products)
- Weighting: weight in the base period expenditure

THE CES INDEX

- Formula (t': current; t: base):

$$I^{t',t} = \frac{\left(\sum_k \alpha_k \left(\frac{p_k^{t'}}{\alpha_k}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}{\left(\sum_j \alpha_j \left(\frac{p_j^{t}}{\alpha_j}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}$$

- Elasticity of substitution (ε) and product coefficients (α_i) are estimated by regressing **quantities on prices** by 2SLS (instrument for the price variable: wholesale – production- price of milk)
- All the products are involved (600) in the index: missing products have an infinite price ($\hat{\varepsilon}>1$)

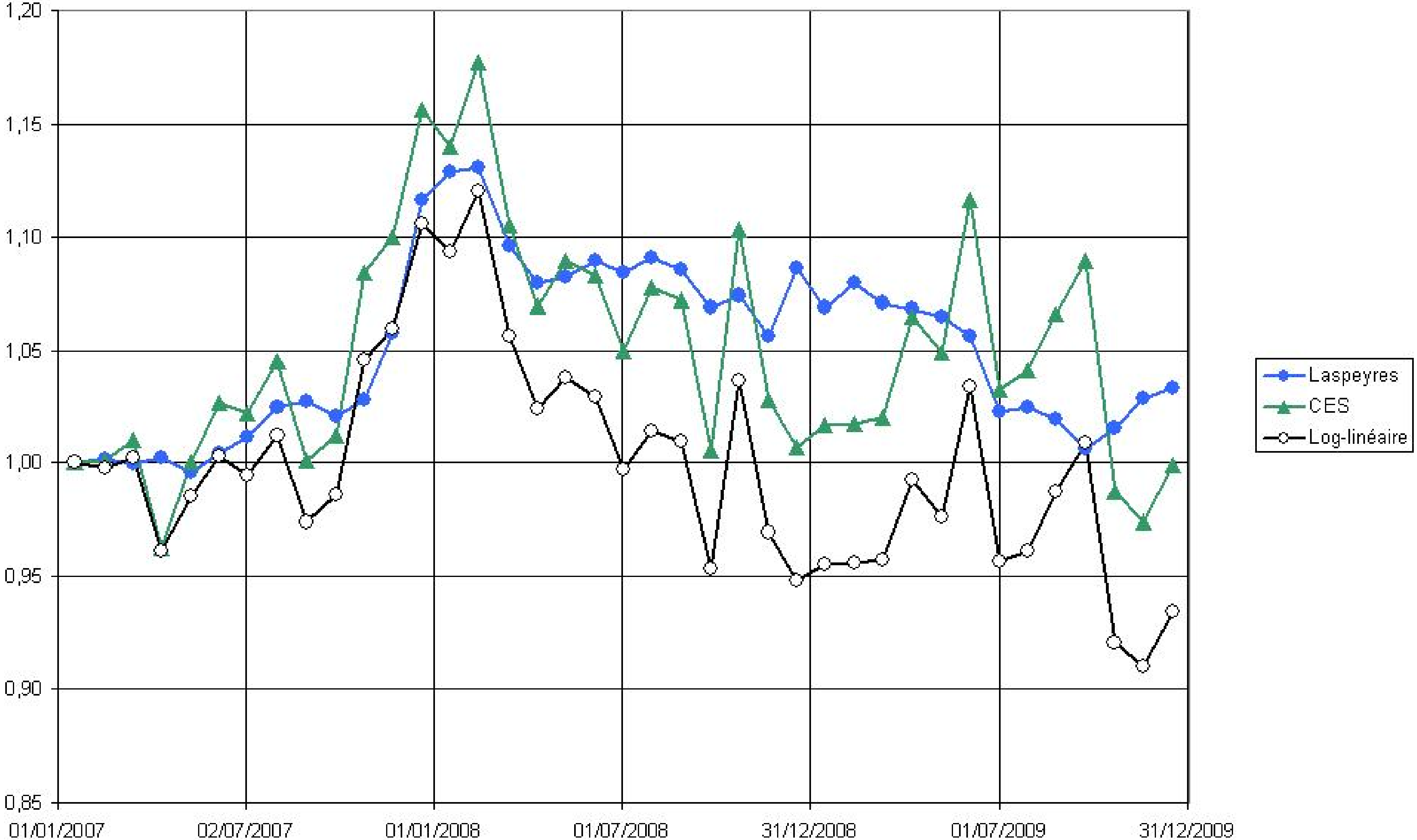
THE LOG-LINEAR DEMAND INDEX

- Formula (t': current; t: base):

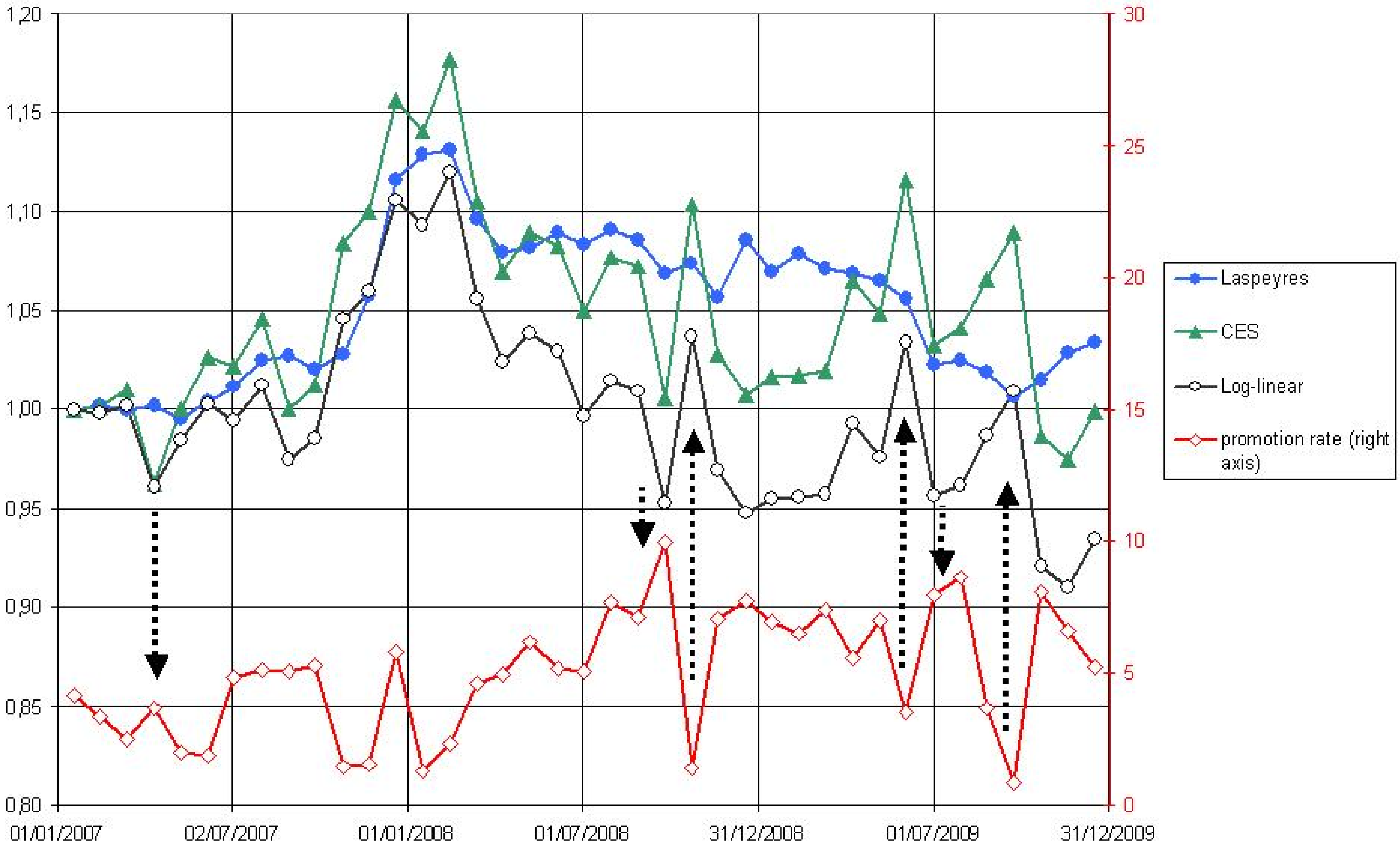
$$I_{LogLin}^{t',t} = \left\{ 1 + (1-\beta) R_t^{\beta-1} \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_{it'}^{1+\alpha} - p_{it}^{1+\alpha}}{1+\alpha} \right\}^{\frac{1}{1-\beta}}$$

- The coefficients (β, α and γ_i) are estimated by regressing **quantities on prices** by 2SLS (instrument for the price variable: wholesale –production- price of milk). The expenditure is shown to be exogenous.
- All the products are involved (600) in the index: missing products have an infinite price ($\hat{\alpha}$ <-1)

THE WEEKLY INDICES



UNDERSTANDING DIFFERENCES WITH THE NUMBER OF DISCOUNT PRODUCTS



Conclusion

GOING FURTHER

- Generally speaking, Constant Utility Indexes can only be computed if both prices and quantities are available
- The results might be quite different from “classical indices” since these last indices generally underestimate product substitutions
- It is also a possible way to treat automatically promotions which are numerous in scanner data
- **BUT**
- This leads us to think about the right perimeter of the product definition: the “right” one should be defined with the observation of real substitutions
- In the French case, we are not ready to introduce these ideas in the CPI but it clearly constitutes a field of research in price indices