SCANNER DATA: TOWARDS CONSTANT UTILITY INDEXES?

Patrick SILLARD(*)
(*)INSEE, Division des prix à la consommation

Introduction

Price indices are now well internationally defined economic indicators. In particular, the methods applied by European countries are made comparable. Price indices are constant-quality indices. In other words, the price index aims at reflecting the evolution of the prices of goods, the level of quality being fixed. From the economic point of view, the underlying concept is the utility constant price index (see for example [21, 27]). But, unless we adopt very restrictive assumptions, constant quality indexes are not coincident with constant utility indexes. Indeed, the constant utility framework supposes to observe the substitution that occurs between the consumption of goods during the successive states of the economy. In other words, if one wants to build constant utility indexes in the general case, one need to observe demand functions.

From the microeconomic point of view, substitutions concern goods of the same type, sold within a geographically limited area where the substitution is practically possible for the consumer. The question of how far are constant utility indexes from constant quality indices arises only at the level of micro-aggregates, that is the elementary computation of a price index based on the prices themselves. After this computation, the elementary aggregates are combined through a Laspeyres aggregation in a separate operation. This paper only deals with micro-aggregates.

Practically, the joint observation of selling prices and quantities supposes the availability of scanner data. In the project carried out by Insee, some retailors allowed Insee to have access to their scanner data. The whole data sets covers 3 years for 1000 shops and 10 families of products.

The bias of price indices has been a central issue of the Boskin Commission in the late 1990s [3, 23, 7] in the US. This work has shown the distance between classical indexes and constant utility indexes (Cost of Living Index). They also showed that the observation of demand functions makes it possible to build constant utility indexes. Some studies have been made on price index biases, especially those ones linked with introduction of new goods or loss of existing goods [14]. The empirical application of these ideas is stayed limited until the availability of scanner data. The papers on scanner data are mainly related to substitution phenomenons [9, 15, 17]. The present paper examines more specifically the application of these ideas to price indices.

1 The economic framework of consumer price indices

The breakdown of good consumption into volumes and prices is a fundamental issue for price indices. The empirical approach that prevails in most modern price indices, however, rests on a number of assumptions that can be reviewed within the theoretical framework of microeconomic theory of consumer behavior [5].

1.1 Economic modelling

We postulate, consistently with microeconomic theory [25], that the consumer makes choices among all the baskets of goods he may consume. He takes his decision according to a utility function maximised under a budget constraint.

Let us denote by $\mathbf{p}$ the price vector and $\mathbf{s}$ that of quantities, u the representative consumer utility function (the function is supposed to be time-independent). The consumer decides to consume a quantity vector $\mathbf{x}$ based on the following optimization program (whose argument is $\mathbf{x}$) :

$$e(\mathbf{p}, U) = \left| \begin{array}{l} \min\limits_{\mathbf{s}} \mathbf{p}.\mathbf{s} \\ \text{u.c. } u(\mathbf{s}) = U \end{array} \right. \tag{1}$$

Like this, $e(\mathbf{p}, U)$ is the expenditure function : it is the minimum amount of money the consumer must do to reach the utility level U conditionally to the exogenous price vector $\mathbf{p}$. In a dual manner, the maximum of utility that it is possible to reach for a given spending R conditionally to the exogenous price vector $\mathbf{p}$ is the indirect utility. We denote this function by $v(\mathbf{p}, R)$.

Let us consider now to different times: a reference one t and a second one t' ($t' > t$). Between the two periods of time, we want to build a price index that shows the evolution of prices. The observables are prices at both times and the initial spending R_t. At t, the consumer-optimiser reaches a utility level $v(\mathbf{p}_t, R_t)$. In order to reach the same utility level at t', he must spend $e(\mathbf{p}_{t'}, v(\mathbf{p}_t, R_t))$. By construction, the constant utility price index is :

$$I_{CU}^{t',t}(R_t) = \frac{e(\mathbf{p}_{t'}, v(\mathbf{p}_t, R_t))}{R_t} \tag{2}$$

This reflects changes in the budget that the consumer must accept to maintain its utility at t' at the level he reached at t with a spending R_t. It useful at this stage to define the first argument of the previous ratio ($\mathbf{p}$ and $\mathbf{q}$ are two price vectors ; R is any expense) :

$$\mu(\mathbf{p}; \mathbf{q}, R) = e(\mathbf{p}, v(\mathbf{q}, R)) \tag{3}$$

This function is the money metric indirect utility function. It corresponds to the spending the consumer must do, at price $\mathbf{p}$, to reach the utility level that he reaches with the price vector $\mathbf{q}$ and a spending R.

With the help of the money metric indirect utility function, the constant utility index may be written:

$$I_{CU}^{t',t}(R) = \frac{\mu(\mathbf{p}_{t'}; \mathbf{p}_t, R)}{R} \tag{4}$$

1.2 The constant utility index for common utility functions

It is relevant to see the form the price index (4) takes when a utility function is explicitly adopted to model consumer's choices. The classical utility functions can be looked at: Cobb-Douglas, Leontief and CES (Constant elasticity of substitution). All of these functions are based on rather different assumptions concerning the substitutability of the goods contained in the basket. The CES function was introduced by [1]. The main aspect of this function is to suppose that the elasticity of substitution between any couple of goods is the same whatever the couple (i, j). If x_i is the demand for good i and p_i is the price, elasticity of substitution $d\ln(x_i/x_j)/d\ln(p_i/p_j)$ is $-\varepsilon$ for the CES utility presented at table 1. This function is concave when $\varepsilon > 0$. We will suppose hereafter that this inequality is true.

Cobb-Douglas and Leontief utilities are limit cases of CES. The first one corresponds to the case $\varepsilon \to 1$ and the second to the case $\varepsilon \to 0$. The expression of utility, demand and index functions are given in table 1.

Table 1: Demand and price index for CES, Cobb-Douglas and Leontief utilities

		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 :	$I^{t',t} =$	$\frac{\left(\sum_k \alpha_k \left(\frac{p_k^{t'}}{\alpha_k}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}{\left(\sum_j \alpha_j \left(\frac{p_j^t}{\alpha_j}\right)^{1-\varepsilon}\right)^{1/(1-\varepsilon)}}$	$\prod_{i=1}^n \left(\frac{p_i^{t'}}{p_i^t}\right)^{\alpha_i}$	$\left(\sum_i \frac{p_i^{t'}}{\alpha_i}\right) / \left(\sum_j \frac{p_j^t}{\alpha_j}\right)$

Notes : The consumer consumes a basket of n goods, denoted $\{1, \ldots, n\}$, in quantities $\mathbf{s}$ and prices $\mathbf{p}$; the CES utility function is defined (concave) for $\varepsilon > 0$; the coefficients α_i for $i \in \{1, \ldots, n\}$ are strictly positive weights that reflects the consumer preferences ; R is the budget (exogenous) that the consumer uses to purchase all goods $\{1, \ldots, n\}$; p_i^t is the price of good i at time t; $I^{t',t}$ is the price index associated to the basket of good at time t' with respect to time t according to relation (2). Formally, the index depends on the budget at time t. But since the utilities are homothetic, le budget disappears in the index expression after some algebra.

One can notice that the Laspeyres price index derives from a Leontief utility function, while the geometric Laspeyres index (Jevons) derives from the Cobb-Douglas utility function. Both of these formulae are used in the French consumer price index micro-aggregate computation. This computation is made at the geographic scale of a town. For homogeneous products, the Laspeyres formula (with unitary weights) is used, while for heterogeneous products, the Jevons index is used.

The choice of utility function has obviously some important consequences on the constant utility index trajectory. In particular, the more the products are substitutes, the more the index is sensitive to relative price variations. Indeed, when the elasticity of substitution is large, when the relative price of a good increases, the consumption of this product decreases in favour of that of other goods, in such a way that the overall spending increase is smaller than the one which would result of the increase of the spending for that good, the quantities being equal. This last case corresponds actually to what occurs when the elasticity is null. The underlying utility is then Leontief and the price index is Laspeyres (see table 1): an increase of relative price of a good results in an increase of the index proportional to the increase of the price weighted by the weight of the good in the total spending. On the contrary, if the relative price decreases, the substitution amplifies the downward compared to what occurs when goods are not substitutable. In summary, the more the products are substitutable, the higher relative price increases will be mitigated and the more cuts will be accentuated.

For example, in a behaviour model "à la Leontief", no new good should appear since in this model, the consumer is supposed to consume all the goods (see the Leontief demand function – table 1). Technically, this inability to describe zero consumption of some goods results, in particular, in the nullity of utility when the consumed quantity of any good is equal to zero. Whereby, the use of such a model leads to abandon the fundamental hypothesis of time stability of the utility function. By the way, it is what is done through the practise of replacement. The possibility to substitute one good to another is then a necessary condition for a model to allow null quantities for some consumed product at some period of time. In the case of CES, it is furthermore necessary to have $\varepsilon > 1$. Indeed, consider the case of a Cobb-Douglas model. To ensure that quantity consumed is zero, the price must be infinite (see the demand function – table 1). And in that case, the index, like utility function, degenerate and are equal to zero. It is the same for a CES utility when $0 < \varepsilon \leqslant 1$.

On the other hand, for a CES utility, if $\varepsilon > 1$, then the good i consumed quantity is equal to 0 when the price p_i^t of the good is infinite at t. In this case, the good do not contribute anymore to the index, except during a transition between two periods, one during which it is consumed and the other when it is not. In such a model, the emergence of a new product results in an increase in consumer utility and lower prices. Conversely, a loss results in a price increase.

The possibility to compute price indexes on time varying baskets has been studied by Balk [2]. Is was successfully applied by Mesler [22] with a CES utility, one of the main difficulties lies in estimating ε which may lead to values below 1 and thus not allowing to use the method.

1.3 From demand function observation to constant utility indices

Without making any additional economic assumption besides the utility maximisation, it is possible to derive utility function from the observation of empirical demand functions of some more or less substitutable dwellings. We consider that these goods are bought by a representative consumer who buys, during successive periods of time, the goods in a shop where it is physically feasible for him to make substitutions. The theoretical framework applied here is that of the integrability of demand functions [19].

The derivation of utility functions or money metric indirect utility functions has been studied by Varian [25, 26] and Hausman [12, 13].

First, all the demand functions do not derive from a utility function. One shows [19, 25] that the demand functions of any goods i and j must verify the following symmetry condition, in order to be solutions of a utility maximisation under budget constraint :

$$\forall (i,j) \in \{1,\dots,n\}\ ,\ \frac{\partial x_i(\mathbf{p},R)}{\partial p_j}+\frac{\partial x_i(\mathbf{p},R)}{\partial R}x_j(\mathbf{p},R) = \frac{\partial x_j(\mathbf{p},R)}{\partial p_i}+\frac{\partial x_j(\mathbf{p},R)}{\partial R}x_i(\mathbf{p},R) \quad (5)$$

where x_i and x_j are demand functions[1] of goods i and j.

[1]These functions depend on $\mathbf{p}$ and R. They are called, in literature, Marshalian demand. For the

If this hypothesis is true, one shows that a relationship exists between the demand conditional to prices $\mathbf{p}$ and for a given budget R, denoted $\mathbf{x}(\mathbf{p}, R)$, and the money metric indirect utility function μ. This relation derives from the Shephard lemma and is a partial differential equation :

$$\forall i \in \{1, \ldots, n\} \ , \ x_i(\mathbf{p}, \mu(\mathbf{p}; \mathbf{q}, R)) = \frac{\partial \mu(\mathbf{p}; \mathbf{q}, R)}{\partial p_i} \tag{6}$$

In this set of equations, $\mathbf{q}$ et R are parameters, the function μ depending on $\mathbf{p}$. If x_i is observed (i.e. its dependency with respect to the price vector $\mathbf{p}$ and the budget R are established empirically), then the set (6) is a partial differential equation set in μ, depending on $\mathbf{p}$ and on the parameters $\mathbf{q}$ and R. A limit condition is added to the set to fully solve the system:

$$\mu(\mathbf{q}; \mathbf{q}, R) = R \tag{7}$$

Whereby, with the help of relation(5), (6) and (7), it is possible to derive a constant utility index from observable demand functions of a set of dwellings $\{1, \ldots, n\}$.

In practise, two solutions could be used. The first one consists in specifying an empirical demand function verifying symmetry conditions (5), estimating this function with econometric methods, and computing a price index after having solved the set of differential equations (6). This is the method we use in this paper. A second method could be used. It is based on non parametric demand functions, and solve for these functions the set (6). This method has been studied by Varian [24] making the link with revealed preference. Nevertheless, practical applications seem to be rather difficult. Hausman and Newey [16] applied non parametric method to measure the effect of a price variation on welfare.

Let us come back to the parametric approach and treat an example of a demand specification and its econometric estimation. There are plenty of possible demand functions. The log-linear specification is often used to characterize demand functions for a given type of product. We then adopt a demand of good i of the form :

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

This specification is log-linear. The application of integrability conditions (5) leads to (M_{ij} is the left member of equation 5) :

$$M_{ij} = \frac{\beta_i}{R} x_i(\mathbf{p}, R) x_j(\mathbf{p}, R) \quad \text{and} \quad M_{ij} = M_{ji} \Leftrightarrow \beta_i \equiv \beta$$

Subject to this, the specification

$$x_i(\mathbf{p}, R) = p_i^{\alpha_i} R^{\beta} e^{\gamma_i} \tag{8}$$

satisfies integrability conditions.
The solution is derived in Appendix. One can find finally that the money metric indirect utility function associated to the set of log-linear demands is :

$$\mu(\mathbf{p}; \mathbf{q}, R) = \left\{ (1-\beta) \left[\frac{R^{1-\beta}}{1-\beta} + \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_i^{1+\alpha_i} - q_i^{1+\alpha_i}}{1+\alpha_i} \right] \right\}^{\frac{1}{1-\beta}} \tag{9}$$

whole basket, the vector of demand is denoted $\mathbf{x}(\mathbf{p}, R)$.

And the constant utility index between t et t' (where $t' \geqslant t$ and R_t is the budget of the representative consumer at time t) is :

$$I^{t',t}_{LogLin} = \left\{ 1 + (1-\beta) R_t^{\beta-1} \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_{it'}^{1+\alpha_i} - p_{it}^{1+\alpha_i}}{1+\alpha_i} \right\}^{\frac{1}{1-\beta}} \tag{10}$$

Unlike the indices presented in table 1, this index involves the budget R_t. The underlying utility is not homothetic which in itself is consistent with intuition. Indeed, there is no reason that a budget increase results in a proportional increase of the consumed quantities. On the contrary, it is rather likely that when the budget increases, consumption shifts to higher quality products, as the works on Engel curves show at the macroeconomic level (see e.g. [4, 18]).

2 The data

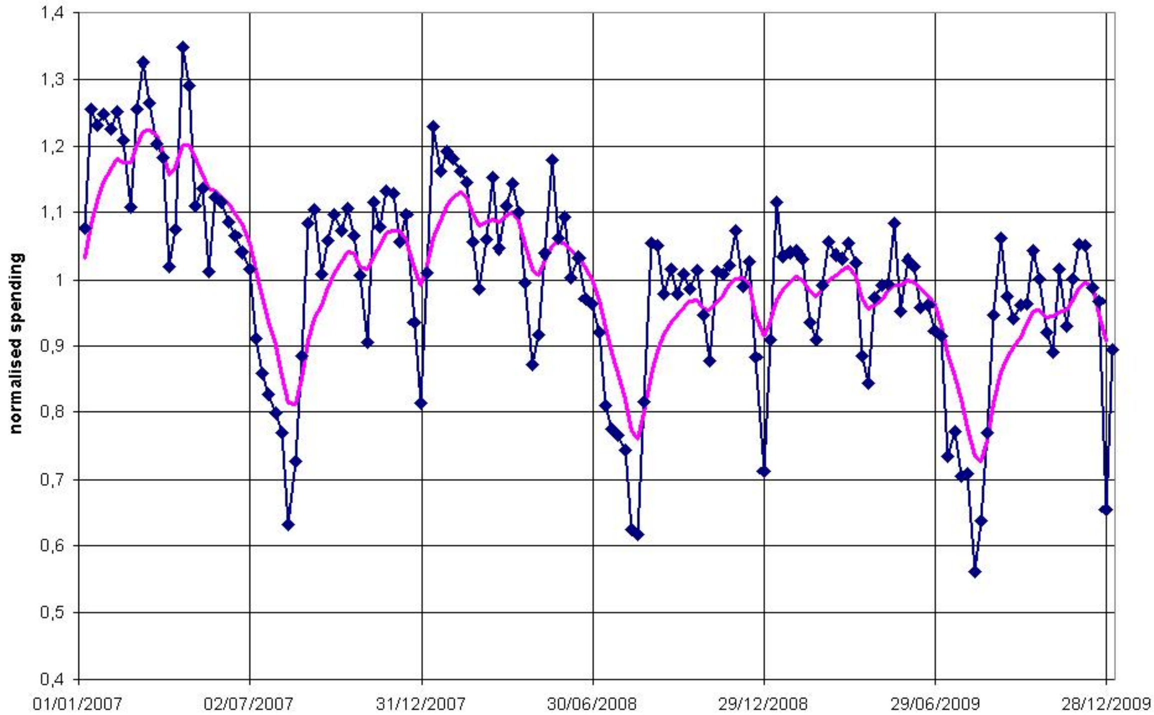
The model proposed in the previous section is based on assumptions of rational behaviour of a representative consumer. The idea of a representative consumer is not obvious. We also know that in general, aggregation of preferences is not based on the optimization of a welfare function. Therefore, it is not certain that the aggregate behaviour of consumers derives from the optimization of a utility function. However, if we wish to retain the theoretical foundation exposed in the previous section, we are led to rely on the notion of representative consumer. In other words, on this market, it is as if a single consumer buys all goods consumed in this market, substituting a good to another according to a rational decision resulting from the optimization of a utility function. This representative consumer is price taker on the market. Obviously, these hypothesis put together are rather unrealistic. However, while considering micro markets, operating conditions are probably not very far from these. In order to adopt a realistic framework, we have to work on a local market where consumers make a choice between really substitutable products. We therefore chose to work with the market for yoghurt sold in a specific retail store. The data cover three years. For this store, we have the quantities sold and weekly prices of all yoghurt sold at least once in the week.

On this market, the assumptions relative to the representative consumer are acceptable since it is possible, in this store, to substitute a good to another. Moreover, for the consumer who consumes a yoghurt in the store, it is easier to substitute yoghurt for another in this store rather than go to another store to buy a replacement product. Thus, an aggregate model of behaviour on the market of yoghurts sold in this store, if not the result of the aggregation of individual preferences, is nevertheless a plausible picture of how the micro-market works. Therefore, a constant utility index based on this model will probably properly reflect the evolution of prices in that market.

The data set includes about 35,000 observations spread over 157 weeks (from January 2007 to December 2009). Over this period, approximately 600 references (barcode) of yoghurt were sold at least once. In this study, the bar code is the identifier of the product. On average, each reference has been sold 950 times during the three years of observation ; median is 300 while 1^{st} and 9^{th} deciles are respectively 30 and 2300.

Figure 1 shows the weekly turnover (normalized by the average of three years) of the yoghurt market for the store. Over the period, the ratio between the highest and the lowest turnover is 2. The curve also shows very marked seasonal effects, the high

Figure 1: Turnover associated with the yoghurt market for the studied store



Note : The "normalized spending" are the ratio of the weekly turnover and the three year-mean weekly turnover. The plain curve is the low-pass Butterworth filtered ratio [11].

point being roughly the first half of the year, while the low point is in August. It is also characterized by a slight downward trend over the period.

3 Application

Three indices where computed according to the principles given in section 1: an annually chained Laspeyres index, a CES constant utility index (table 1 formula) and a constant utility index based on the inversion of demand functions (formula 10).

The Laspeyres index is based on a fixed basket of goods annually updated. The basket is based on the set of goods sold during the base period, here chosen as the first month of the year (i.e. in January). The index is computed according to the Laspeyres formula with weights equal to the base period quantities. For the case, when a product is not sold during a month, its price is estimated by applying the observed monthly mean variation to the last observed price. The algorithm is applied iteratively, also in the case of product loss. Except from replacements, this process corresponds to that of the French CPI. In food products, the quality corrections are rather small – estimated at less than 0,1% in annual increase [10] – therefore, the quality bias associated with the mechanics used in this computation is probably small. In contrast, the index computed like this does not take into account the manufacturers' promotions (such as the sale of two yoghurts for the price of one). Indeed, the barcode of this products is different. The relation between the two products (the one in promotion and the original one) is not done in this study. This limit implies that the present Laspeyres index is not taking into account manufacturers' promotions, and is then biased. Note that in the CPI, such bias does not exist since manufacturers' promotions are followed. The collector makes the connection between products in promotion and the elementary product concerned. The price taken into account in the CPI is a unit price (per unit volume, weight, or otherwise as may be relevant for the product concerned) and then the manufacturers'promotion are naturally taken in the CPI.

The computation of a CES index presupposes that the weighting parameters of individual products and elasticity of substitution are fixed. They are estimated with econometric methods based upon the whole 3-year set of data (prices and quantities). This estimation makes it possible to fix, for each product i, an elementary preference parameter α_i (see table 1 formula). This parameter describes, at each time period, the consumer's preference for the considered good, this good being present in the shop or not. Built like this, the utility which drives consumer choices is stable over time: the new products can be integrated over time and the disappearing products are not any longer taken into account in the index. Manufacturers'promotions fit naturally in the calculation.

The computation of a constant utility index based on a log-linear demand (formula 10) needs, like for CES utility, the estimation of some parameters. Again, the resulting utility is stable over time: depending on the estimated values of parameters, new products can be integrated naturally in the index by considering that their price is infinite if they are not available for sale.

3.1 Econometric estimation of model parameters

The estimation of CES utility parameters is based on the equation of demand (see table 1). Taking the logarithm of the demand function, we get the following structural equation (good i, time t) :

$$\ln(x_{it}) = \underbrace{\varepsilon \ln(\alpha_i)}_{c_i} - \varepsilon \ln(p_{it}) + \underbrace{\ln R_t - \ln\left[\sum_k \alpha_k \left(\frac{p_{kt}}{\alpha_k}\right)^{1-\varepsilon}\right]}_{\nu_t} \tag{11}$$

The unknown coefficients c_i and ε might be estimated through the regression of $\ln(x_{it})$ on $\ln(p_{it})$, ν_t being here an error term. A few methods on panel data can be used (first differences or fixed effect). The main issue is that the price p_{it} is very likely endogenous in the equation. Indeed, we can imagine that the seller fixes the price according to reaction of the demand he expects. This generates a simultaneity bias in the ordinary least square (OLS) estimation. We then need an instrument to circumvent this problem. The structural equation we are interested in is a demand equation. To identify this equation, we look for a instrument of the price that is exogenous in a demand equation. We chose to take a variable related to production costs that is the index of wholesale prices for milk food consumption published by Insee every month. This variable is resampled to make a weekly index. Elasticity and product fixed effect are estimated by two stage least squares (2SLS). The results are presented in table 2. They show in a very clear way that price is an endogenous variable in the demand equation and that the instrument is valid: the elasticity estimate goes from 0.373 in the case of OLS without fixed effects to 1.83 for 2SLS on differences. Regressions on levels and on differences give the same results concerning 2SLS. We keep as the reference regression that on levels: besides the fact that the estimate is more accurate, estimation of fixed effects products (c_i in equation 11) allows us to estimate the coefficients α_i. Indeed, in that case, $\hat{\alpha}_i = \exp(\hat{c}_i/\hat{\varepsilon})$ is an unbiased estimate.

Table 2: CES utility parameter estimation

Regression	$\hat{\varepsilon}$	product fixed effect	AUX
OLS	0.373^{***} (0.014)	no	
Fixed effect est.	1.21^{***} (0.039)	yes	
OLS on the $\Delta's$	1.27^{***} (0.057)	no	
2SLS on the levels	1.82^{***} (0.090)	yes	0.36^{***} (0.005)
2SLS on the $\Delta's$	1.83^{***} (0.222)	no	

Note : Δ means first differences. 3 stars denote a 1% significant parameter. Standard deviations are given in parenthesis. For 2SLS, the auxiliary regression is the log of the product price on the log of the index of wholesale prices for milk food consumption and product fixed effects when they are included in the main equation (column "product fixed effects"). When differences are involved, auxiliary regression is also treated in differences. The computation is based on about 34000 observations. AUX denotes the coefficient associated with the log of the milk price in the auxiliary regression (2SLS).

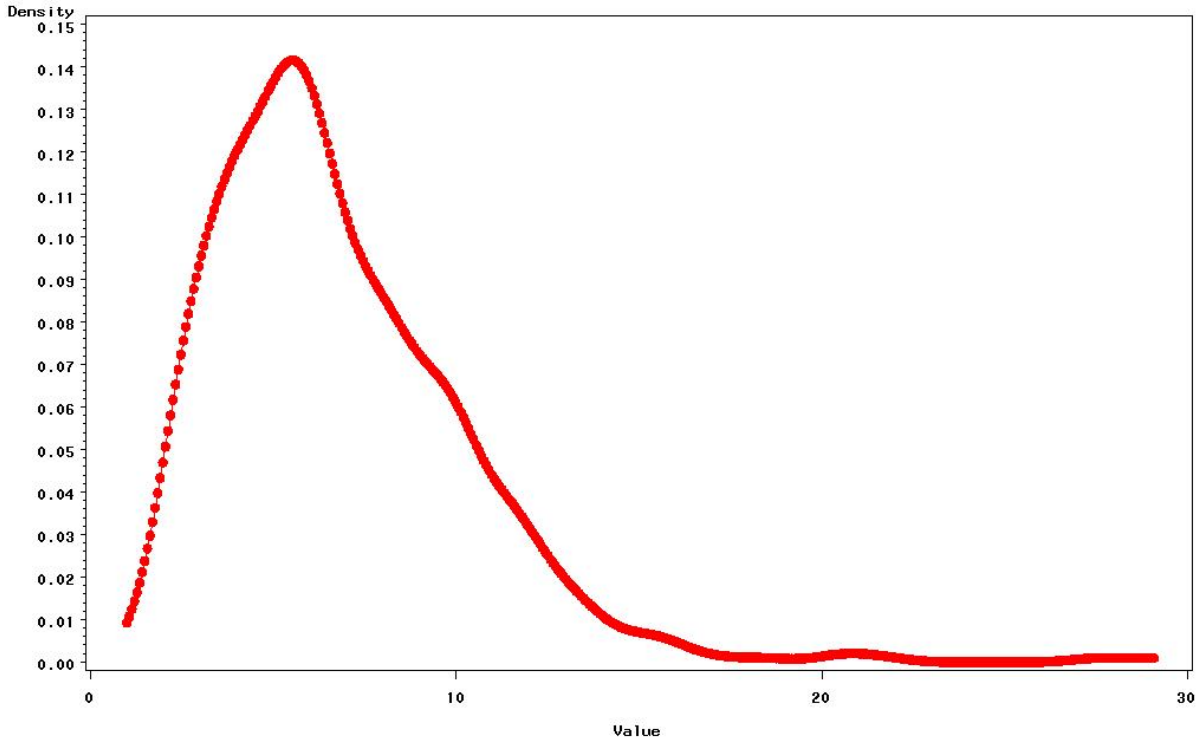
Figure 2 shows the density of estimated α_i. This curve shows the magnitude of consumer preferences towards the various yoghurts of the basket. The differences are large: the contribution to utility goes from 1 to 20 depending on the considered yoghurt.

Like for CES utility parameter estimation, the estimates of the log-linear demand parameters are based on demand equation (8). Taking the logarithm, we get a structural equation:

$$\ln(x_{it}) = \gamma_i + \alpha_i \ln(p_{it}) + \beta \ln(R_t) \quad (12)$$

We estimate the coefficient by regressing $\ln(x_{it})$ on $\ln(p_{it})$ and $\ln(R_t)$.

Figure 2: Density of α_i coefficients of equation 11 (product weight in CES utility)



Note : coefficients α_i are based on fixed effects coefficients c_i in the 2SLS regression on the levels and come from the relation: $\hat{\alpha_i} = \exp(\hat{c}_i/\hat{\varepsilon})$. Smoothing by kernel estimation.

However, as before, the logarithm of the price p_{it} is endogenous in the regression for the same reasons as those mentioned in connection with the CES utility. We only have one instrument for the whole set of product prices, therefore we are not able to identify the whole set of α_i coefficients. But if we assume that all α_i are identical, then the above equation simplifies to:

$$\ln(x_{it}) = \gamma_i + \alpha \ln(p_{it}) + \beta \ln(R_t) \quad (13)$$

This equation is identifiable if the expenditure R_t is exogenous in the equation or if we have an instrument for it. We could imagine some cause for endogeneity: for example, if the consumer is price taker, then the seller is likely to fix its price in order to maximise its profit (that is also the consumer expenditure), knowing the consumer reaction to price change. In that case, quantities result from the simultaneous determination of prices, quantities and expenditure.We try to test the endogeneity of the expenditure variable with the help of an instrument of this variable in a demand equation, that is to say with a variable that is correlated with the expenditure but that do not play any role in the demand equation.

For example, a wage variable would not be appropriate, because if it is actually correlated with expenditure, it is likely that it plays a role in the demand. The variable we use is a binary variable indicating the end of the month. Indeed, in terms of demand determinants, there is no indication that the end of the month is a special time to buy yoghurt. However, since the wages are usually paid in the last week of the month, the last part of the month is probably a period xhen the budget constraint is heavier for consumers. A binary variable indicating the end of the month (wee took the period from the 23^{th} to the 30^{th}) is then a good candidate for instrument: empirically, this variable is negatively correlated with expenditure (see column "AUX2" in table 3). However, we can see the the instrument is not very powerful since the standard deviations of the expenditure coefficient in the 2SLS regression are much larger than the ones we get with OLS. Finally, a Hausman test shows that expenditure is not endogenous in the demand equation. We select the 2SLS regression on the levels with an instrument for prices only (regression "2SLS on the levels (^{2})" in table 3).

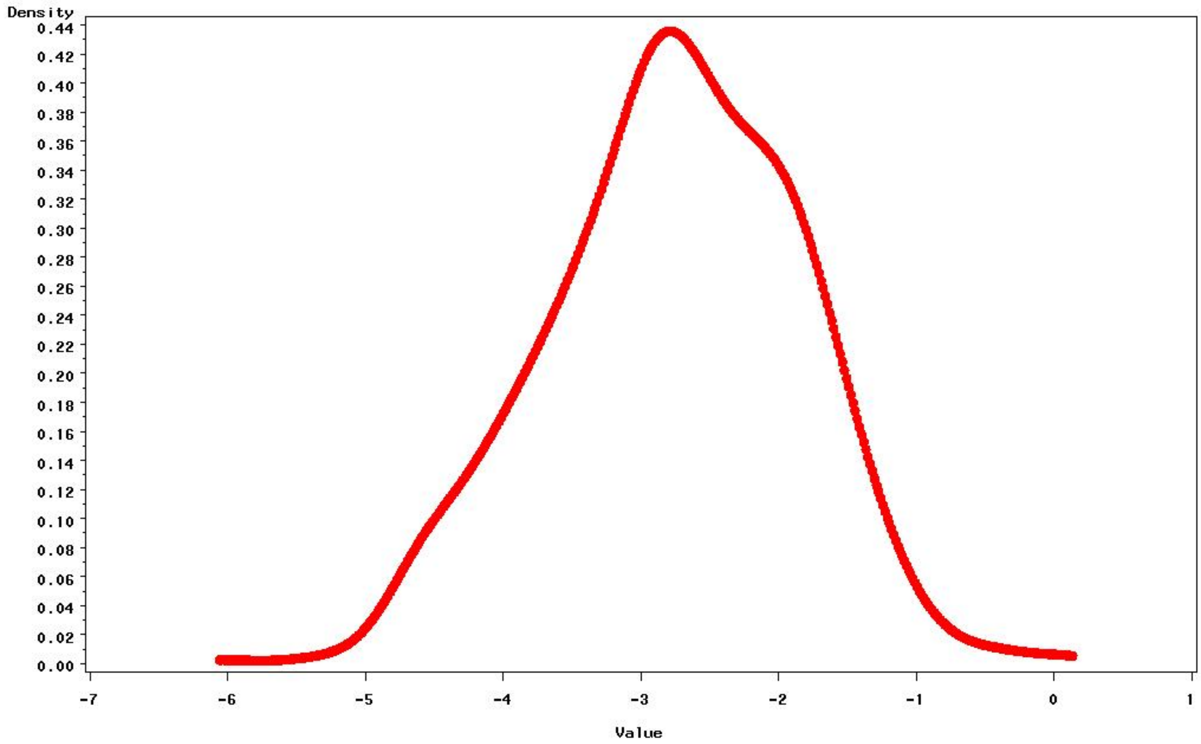
Table 3: Estimation of the log-linear demand parameters

Regression	$\hat{\alpha}$	$\hat{\beta}$	product fixed effect	AUX1	AUX2
OLS on the levels	-1.13^{***} (0.04)	1.09^{***} (0.02)	yes		
OLS on $\Delta's$	-1.19^{***} (0.06)	1.00^{***} (0.03)	no		
2SLS on the levels	-2.00^{***} (0.18)	0.67^{*} (0.42)	yes	0.26^{***} (0.02)	-0.017^{***} (0.002)
2SLS on $\Delta's$	-1.21 (1.26)	2.51^{*} (1.63)	no	0.24^{***} (0.03)	-0.005^{***} (0.001)
2SLS on the levels (^{2})	-1.86^{***} (0.10)	1.08^{***} (0.02)	yes	0.26^{***} (0.02)	

Note : Δ means first differences. 3 stars denote a 1% significant parameter; 1 star denotes a 15% significant parameter. Standard deviations are given in parenthesis. For 2SLS, the auxiliary regressions are: 1) the log of the product price on the log of the index of wholesale prices for milk food consumption; 2) the log of expenditure on the binary variable indicating the end of the month. When differences are involved, auxiliary regression is also treated in differences. The computation is based on about 34000 observations. AUX1 is the coefficient associated with the log of the milk price in the auxiliary regression ; AUX2 is the coefficient of the binary variable indicating the end of the month in the auxiliary regression of the log of expenditure. (^{2}) : regression without expenditure instrumentation.

As for the CES index, the index based on log-linear demand makes it possible to treat incoming and outgoing products through a natural way, considering that the price of missing products is infinite. Figure 3 gives de density of γ_i coefficients based on specification (13) for the reference regression. These coefficients are all negative and their spread goes from 0 to -6. This means that the weight of products in terms of contribution to the index (see relation 10) is 1 to 1/400.

Figure 3: Density of Eq. 13 γ_i coefficients (constant multiplier of log-linear demand)

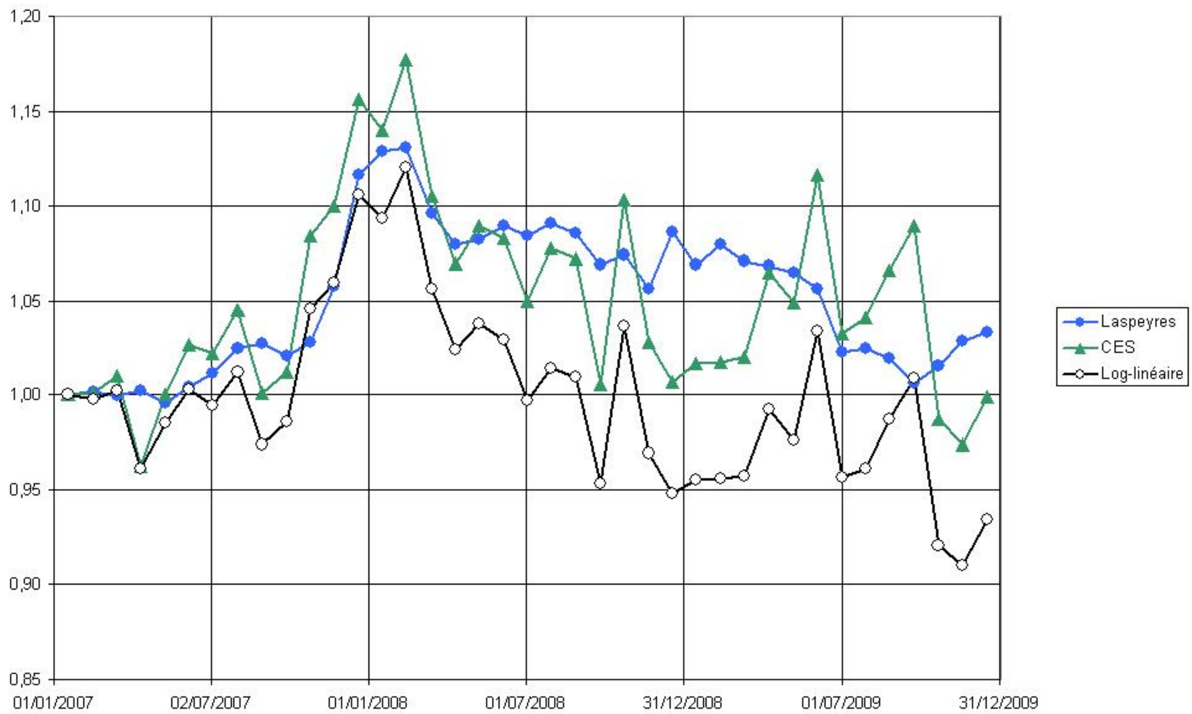


Note : The coefficients correspond to the fixed effects of the 2SLS regression with auxiliary regression for prices, the expenditure being taken as exogenous. Smoothing by kernel estimation.

3.2 Results on Indices

Figure 4 shows the curves of the three studied indices. The indices are computed each month. Quantities are the sum of weekly quantities and prices are average prices over the month for each item. It may be noted that the Laspeyres index is less noisy than the other two. This is related to the fact that the log-linear and CES indices are, by construction, sensitive to input-output of goods in the store. This applies to products whose presence is almost continuous but may occasionally disappear. This also applies to manufacturers' promotions whose duration in scanner data is relatively short.

Figure 4: Laspeyres, CES and log-linear monthly price indices between 2007 and 2010

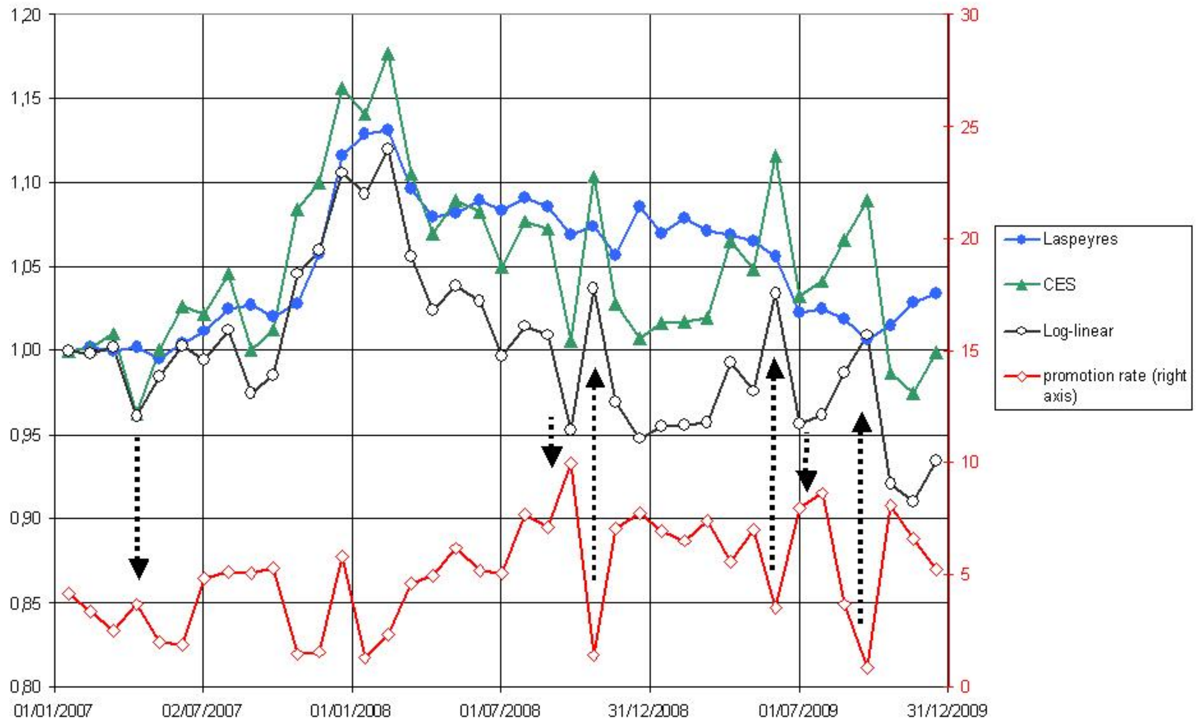


Note : Laspeyres index is annually chained with a fixed basket updated annually.

In order to illustrate the differences between the indices, figure 5 shows the indices and the rate of discount products foreach month (the rate is computed with respect to the number of article references available for sale). It can be noticed that when the promotion rate is low, the CES and log-linear indices are higher, while when the promotion rate is high, the indices decrease. This comparison confirms that the spread between the Laspeyres index and the CES index is primarily related to promotions that are not included in the first, while they are in the second.

Apart from promotional effects, one can note the good consistency of the chained Laspeyres and that based on CES utility. However, the index based on log-linear demand reflects a price growth lower. If we examine further the index formula (10), we can show, under various assumptions of smallness of some terms (see Appendix), that the index is

Figure 5: Laspeyres, CES and log-linear monthly price indices and rate of discount products



Note : Arrows indicate the months with higher promotion rate (resp. low discount rate), depending on the case, which in turn leads to low (resp. high) prices.

approximately equal to:

$$
\begin{aligned}
I^{t',t}_{LogLin} \simeq\ & 1 + \sum_{i \in \overset{\circ}{I}} \omega_{it} \frac{p_{it'} - p_{it}}{p_{it}} + \\
& \frac{1}{1+\alpha} \left(\sum_{i \in \overset{\circ}{I}} \omega_{it} - \sum_{i \in \overset{\circ}{I}} \omega_{it'} \right) + \frac{1-\beta}{1+\alpha} \frac{R_{t'} - R_t}{R_t} \left(1 - \sum_{i \in \overset{\circ}{I}} \omega_{it} \right)
\end{aligned}
\tag{14}
$$

where $\overset{\circ}{I}$ is the set of all the products present at all observation dates, ω_{it} is the product i weight [2] in the expenditure at the time t, the latter being denoted R_t. The first term in this expression is a Laspeyres index, restricted solely to the goods present at time t (basis) and t' (in practice, we define $\overset{\circ}{I}$, with the help of all goods present at all dates). Empirically, the corresponding price index, shown in Figure 6, reflects a positive change in prices over the period of observation. The first term of (14) is then positive. For the products always present, the trajectory of the price index is similar to that of the chained Laspeyres index shown in Figure 4, this one being related to all goods.

The two other terms of (14) are related, on the one hand to the variation of weights in the contemporary expenditure, of the products always present and, on the second hand, to total expenditure variations. Empirically, the weights of always present products is relatively stable over the period 2007-2009 (see figure 7). A larger variability can be seen in 2009 due, inter alia, to a greater sensitivity of weights to promotion rates, probably reflecting more frequent substitutions. The decrease in the weight of the products on all dates occurs mainly in the last quarter of 2009. The second term of (14) is then empirically negative.

Finally, we observe (see figure 1) a downward trend of expenditure throughout the reporting period. Since $\alpha < -1$ and $\beta > 1$, the last terme of (14) is negative.

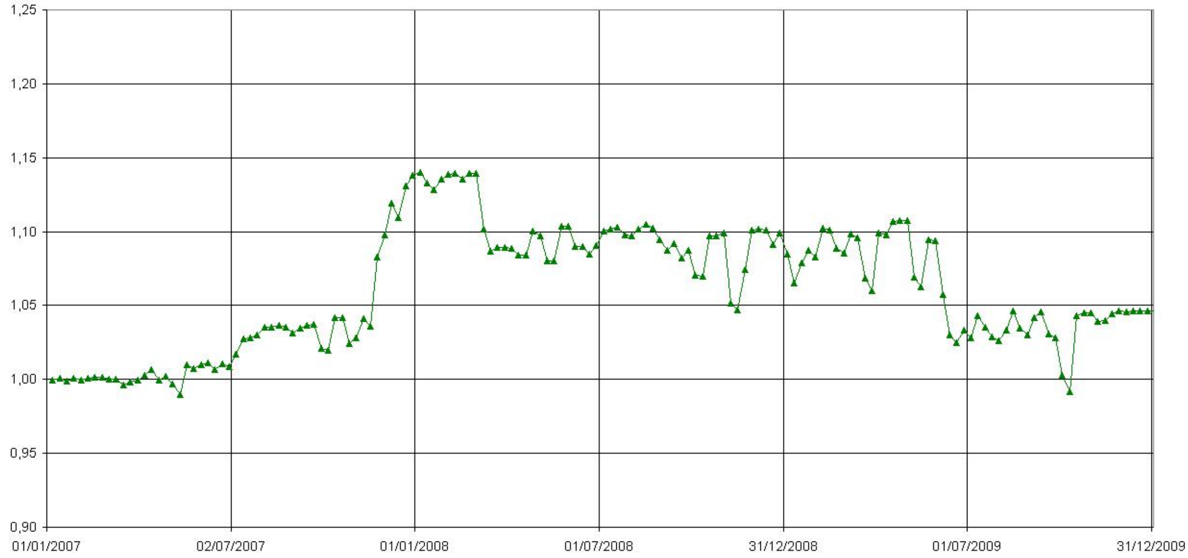
Economically, the second term of (14) represents the effect of substitutions, while the last term is sensitive to changes in overall spending. When utility is homothetic, this term vanishes. In this case, given the estimated values for α and β, a decline in spending is associated with a utility gain ; in this context, the yoghurts are akin to an inferior good. The numerical computation presented in appendix shows that the magnitude of substitution effect is twice larger than that related to the decreasing trend of expenditure.

4 Conclusion

Three different approaches are possible to construct a price index. The first, the classical approach, is based mainly on the *ad hoc* properties a price index should follow, these properties being inherited from the properties satisfied by a sequence of successive price of a product [8]. This approach, called axiomatic approach [20], is that which prevails today in the construction of "official" CPI. The second approach is to derive the index formula after the adoption of a utility function which is supposed to drive the choices made by a representative consumer. The parameters of the utility function are estimated from the observation of both prices and quantities. The third approach is based on observation and parameterization of the demand functions from which we infer a constant utility index. The last two approaches are possible only with scanner data. For many years economists have examined the differences between the indices based on the axiomatic approach, easy

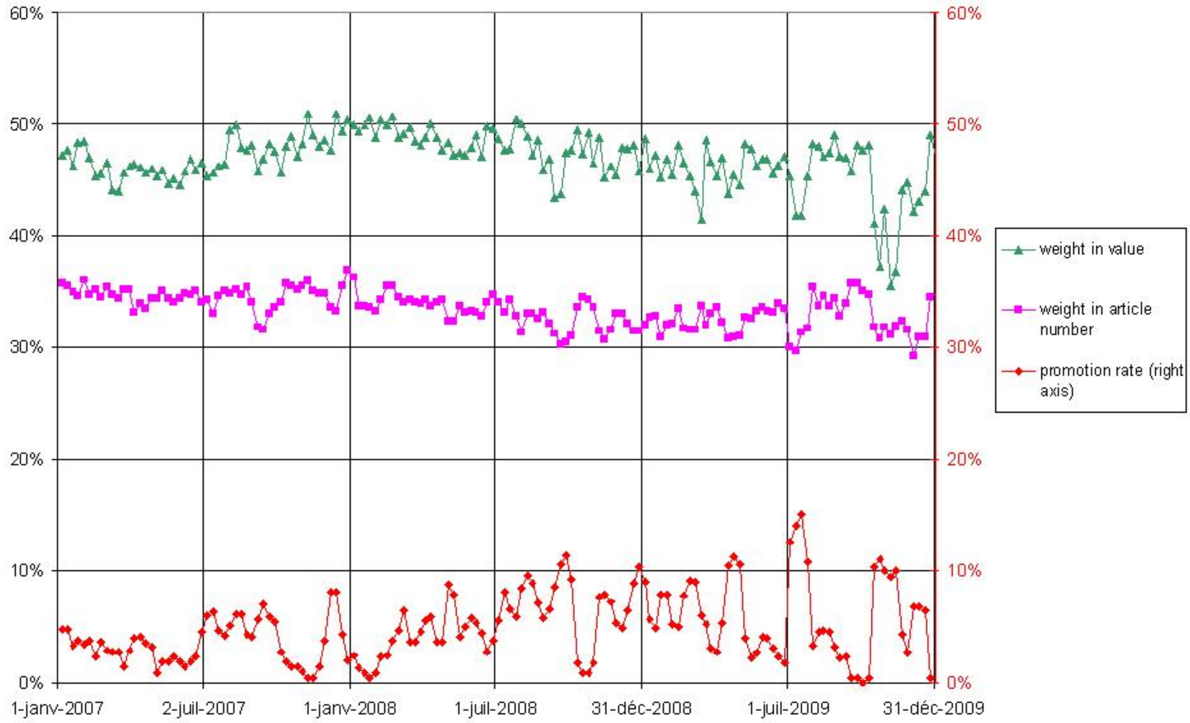
[2] At any time t, $\sum_i \omega_{it} = 1$.

Figure 6: Laspeyres Index for product always present



Note : there are about 200 references of articles always present. The products that are always present over the 2007-2009 period represent approximatively 35% of the products (see figure 7).

Figure 7: Weight in total expenditure of the products that are always present



to use and easy to explain, and the constant utility index, satisfactory from an economic point of view, but which requires rarely available elements, on demand functions, to be computed. If the link between the two types of index is known for a long time [6] , the recent availability of scanner data makes it possible to reconsider putting into practice the concept of constant utility index. The first results show that the computation might be simplified, particularly with regard to replacement of products. Beyond this, much remains to be done to measure in detail the implications of the econometric estimation of parameters on the computed index. Ideally, it would be good also to overcome some assumptions which are not always justified, for example on a specific utility function.The non-parametric approach could therefore constitute a goal in the use of scanner data.

References

[1] Arrow, K. J., Chenery, H. B., Minhas, B. S., and Solow, R. M. Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics 43*, 3 (1961), 225–250.

[2] Balk, B. On Curing the CPI's Substitution and New Goods Bias. Discussion paper, The Ottawa Group, 1995, 1999.

[3] Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z., and W.Jorgenson, D. Consumer Prices, the Consumer Price index, and the Cost of Living. *Journal of Economic Perspectives 12*, 1 (1998), 3–26.

[4] Clerc, E., and Coudin, E. L'IPC, miroir de l'évolution du coût de la vie en France ? Ce qu'apporte l'analyse des courbes d'Engel. *Économie et Statistique 433–434* (2010), 77–105.

[5] Deaton, A., and Muelbauer, J. *Economics and consumer behavior.* Cambridge University Press, 1980.

[6] Diewert, E. Aximoatic and economic approaches to elementary price indexes. Working paper 5104, NBER, 1995.

[7] Diewert, E. Index Number Issues in the Consumer Price Index. *Journal of Economic Perspectives 12*, 1 (1998), 47–58.

[8] Eichhorn, W., and Voeller, J. *Theory of the Price Index.* Springer-Verlag, 1976.

[9] Feenstra, R. C., and Shapiro, M. D. *High-Frequency Substitution and the Measurement of Price Indexes.* University of Chicago Press, 2003, pp. 123–146.

[10] Guédès, D. Impact des ajustement de qualité dans le calcul de l'indice des prix à la consommation. Document de travail F0404, INSEE, 2004.

[11] Hamming, R. *Digital Filters*, 3 ed. Dover, 1997.

[12] Hausman, J. Exact consumer's surplus and deadweight loss. *The American Economic Review 71*, 4 (1981), 662–676.

[13] Hausman, J. Exact Consumer's Surplus and Deadweight Loss. *American Economic Review 71*, 4 (1981), 662–676.

[14] Hausman, J. Sources of Bias and Solutions to Bias ind the Consumer Price Index. *Journal of Economic Perspectives 17*, 1 (2003), 23–44.

[15] Hausman, J., and Leibtag, E. Consumer benefits from increased competition in shopping outlets : measuring the effect of WAL-MART. *Journal of Applied Econometrics 22* (2007), 1157–1177.

[16] Hausman, J. A., and Newey, W. K. Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss. *Econometrica 63*, 6 (1995), 1445–1476.

[17] Hausman, J. A., Pakes, A., and Rosston, G. L. Valuing the Effect of Regulation on New Services in Telecommunications. *Brookings Papers on Economic Activity, Microeconomics* (1997), 1–54.

[18] Herpin, N., and Verger, D. *Consommation et modes de vie en France*, 3 ed. La Découverte, 2008.

[19] Hurwicz, L., and Uzawa, H. On the integrability of demand functions. In *Preferences, Utility, and Demand* (1971), J. S. Chipman, L. Hurwicz, M. K. Richter, and H. F. Sonnenschein, Eds., Harcourt, pp. 114–148.

[20] ILO, IMF, OECD, UNECE, Eurostat, and The World Bank. *Consumer price index manual : Theory and Practise.* International Labour Office, 2004.

[21] Magnien, F., and Pougeard, J. Les indices à utilité constante : une référence pour mesurer l'évolution des prix. *Économie et Statistique 335* (2000), 81 – 94.

[22] Mesler, D. Accounting for the effects of new and disappearing goods using scanner data. *Review of Income and Wealth 52*, 4 (2006), 547–568.

[23] Moulton, B. R., and Stewart, K. J. An Overview of Experimental U.S. Consumer Price Indexes. *Journal of Business & Economic Statistics 17*, 2 (1999), 141–151.

[24] Varian, H. The Nonparametric Approach to Demand Analysis. *Econometrica 50*, 4 (1982), 945–973.

[25] Varian, H. R. *Microeconomic Analysis.* The MIT Press, 1975.

[26] Varian, H. R. Trois évaluation de l'impact "social" d'un changement de prix. Cahiers du séminaire d'économétrie, 1982.

[27] Viglino, L. Le concept unificateur des indices de prix et proposition d'un nouvel indice. In *Actes des journées de méthodologie statistique* (2000), INSEE.

Appendix: computational developments

Solving of the partial differential equation system (6) for demand functions (8)

We must solve in μ the following set of differential equations :

$$1 \leqslant i \leqslant n \ , \ \frac{\partial \mu}{\partial p_i} = p_i^{\alpha_i} \mu^\beta e^{\gamma_i}$$

For i, we have in particular $\frac{\partial \mu}{\partial p_i} = p_i^{\alpha_i} \mu^\beta e^{\gamma_i}$ which can be solved, partially, by integrating on p_i :

$$\frac{\mu^{1-\beta}}{1-\beta} = \frac{p_i^{1+\alpha_i}}{1+\alpha_i} e^{\gamma_i} + C_i(\mathbf{p}_{(i)}; \mathbf{q}, R)$$

where C_i is a function depending on $\mathbf{p}_{(i)}$ (vector of prices $\mathbf{p}$ without the i^{th} component), on $\mathbf{q}$ and on R. We can therefore seek a solution of the form

$$\mu(\mathbf{p}; \mathbf{q}, R) = \left\{ (1-\beta) \left[\sum_{i=1}^{n} \frac{p_i^{1+\alpha_i}}{1+\alpha_i} e^{\gamma_i} + C(\mathbf{q}, R) \right] \right\}^{\frac{1}{1-\beta}}$$

where C is a function depending on $\mathbf{q}$ and R. It can be found using the limit condition (see relation 7) $\mu(\mathbf{q}; \mathbf{q}, R) = R$, so:

$$C(\mathbf{q}, R) = \frac{R^{1-\beta}}{1-\beta} - \sum_{i=1}^{n} \frac{q_i^{1+\alpha_i}}{1+\alpha_i} e^{\gamma_i}$$

Finally,

$$\mu(\mathbf{p}; \mathbf{q}, R) = \left\{ (1-\beta) \left[\frac{R^{1-\beta}}{1-\beta} + \sum_{i=1}^{n} e^{\gamma_i} . \frac{p_i^{1+\alpha_i} - q_i^{1+\alpha_i}}{1+\alpha_i} \right] \right\}^{\frac{1}{1-\beta}}$$

Approximation of the log-linear index (14)

The proof is based on the assumption that $p_{it} - p_{it'}$, $R_t - R'_t$, $\omega_{it} - \omega_{it'}$ and that the complement to 1 of the index $I_{LogLin}^{t',t}$ are infinitesimals of first order. We also suppose that all the α_i are equal to a unique α (framework of econometrics). Starting from (10), to the first order, we have :

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

If we note I^t the set of products present at time t, it is possible to define the intersection $\mathring{I}$ of these sets for all the dates. Thus defined, $\mathring{I}$ corresponds to the set of the products that are always present. For all t, one note $\bar{I}^t$ the complement to $\mathring{I}$ in I^t. The previous relation can then be written:

$$I_{LogLin}^{t',t} \simeq 1 + R_t^{\beta-1} \sum_{i \in \mathring{I}} e^{\gamma_i} \frac{p_{it'}^{1+\alpha} - p_{it}^{1+\alpha}}{1+\alpha} + R_t^{\beta-1} \sum_{i \in \bar{I}^{t'}} e^{\gamma_i} \frac{p_{it'}^{1+\alpha}}{1+\alpha} - R_t^{\beta-1} \sum_{i \in \bar{I}^{t}} e^{\gamma_i} \frac{p_{it}^{1+\alpha}}{1+\alpha}$$

PThen, at any time t, the budget constraint implies $\sum p_{it}x_{it} = R_t$. Taking into account the expression of demand function (relation 8), it follows that $\omega_{it} = R_t^{\beta-1} p_{it}^{1+\alpha} e^{\gamma_i}$ is the good i weight in the expenditure R_t at time t.

For $i \in \overset{\circ}{I}$, by a Taylor expansion to first order in $p_{it'} - p_{it}$, we have :

$$p_{it}^{1+\alpha} - p_{it}^{1+\alpha} \simeq (1+\alpha) p_{it}^{\alpha} (p_{it'} - p_{it})$$

It follows that:

$$R_t^{\beta-1} \sum_{i \in \overset{\circ}{I}} e^{\gamma_i} \frac{p_{it'}^{1+\alpha} - p_{it}^{1+\alpha}}{1+\alpha} \simeq \sum_{i \in \overset{\circ}{I}} \omega_{it} \frac{p_{it'} - p_{it}}{p_{it}}$$

Then, the other terms are simplified by using the budget constraint and the fact that the sum of weights is 1 for a given time t. Thus:

$$R_t^{\beta-1} \sum_{i \in \bar{I}^t} e^{\gamma_i} \frac{p_{it}^{1+\alpha}}{1+\alpha} = \frac{1}{1+\alpha} \left(1 - \sum_{i \in \overset{\circ}{I}} \omega_{it} \right)$$

Similarly,

$$R_t^{\beta-1} \sum_{i \in \bar{I}^t} e^{\gamma_i} \frac{p_{it'}^{1+\alpha}}{1+\alpha} = \frac{1}{1+\alpha} \left(\frac{R_t}{R_{t'}} \right)^{\beta-1} \left(1 - \sum_{i \in \overset{\circ}{I}} \omega_{it'} \right)$$

The last expression can be simplified on the basis of the assumptions that the differences $R_{t'} - R_t$ and $\omega_{it} - \omega_{it'}$ are small. Indeed,

$$\frac{1}{1+\alpha} \left(\frac{R_t}{R_{t'}} \right)^{\beta-1} \left(1 - \sum_{i \in \overset{\circ}{I}} \omega_{it'} \right) = \frac{1}{1+\alpha} \left(1 + \frac{R_{t'} - R_t}{R_t} \right)^{1-\beta} \left(1 - \sum_{i \in \overset{\circ}{I}} \omega_{it} + \sum_{i \in \overset{\circ}{I}} (\omega_{it} - \omega_{it'}) \right)$$

By developing to the first order and summing the result with the other two terms already explained, we have:

$$\begin{aligned} I_{LogLin}^{t',t} \simeq\ & 1 + \sum_{i \in \overset{\circ}{I}} \omega_{it} \frac{p_{it'} - p_{it}}{p_{it}} + \\ & \frac{1}{1+\alpha} \left(\sum_{i \in \overset{\circ}{I}} \omega_{it} - \sum_{i \in \overset{\circ}{I}} \omega_{it'} \right) + \frac{1-\beta}{1+\alpha} \frac{R_{t'} - R_t}{R_t} \left(1 - \sum_{i \in \overset{\circ}{I}} \omega_{it} \right) \end{aligned}$$

numerical computation : we adopt the following numerical values

- $\sum_{i \in \overset{\circ}{I}} \omega_{it} = 0.45$
- $\frac{1}{\sum_{i \in \overset{\circ}{I}} \omega_{it}} \sum_{i \in \overset{\circ}{I}} \omega_{it} \frac{p_{it'} - p_{it}}{p_{it}} = 0.05$
- $\alpha = -1.86$
- $\beta = 1.08$
- $\sum_{i \in \overset{\circ}{I}} (\omega_{it} - \omega_{it'}) = 0.46 - 0.44 = 0.02$ (mean difference of the first and last semesters of observation)

With these numbers,

$$\begin{aligned} I_{LogLin}^{2009S2,2007S1} &\simeq 1 + \underbrace{0.45 \times 0.05}_{0.02} - \underbrace{\frac{1}{0.86} \times 0.02}_{0.02} - \underbrace{\frac{0.08}{0.86} \times 0.2 \times (1 - 0.45)}_{0.01} \\ &\simeq 1 - 0.01 \end{aligned}$$

As expected, the index $I_{LogLin}^{t',2007}$ is negative for the last semester of 2009. This computation allows us to derive the contributions: the Laspeyres index of the goods that are always present contributes positively to 2 points; the decrease of the total expenditure contributes negatively to 1 point; and substitutions contribute negatively to 2 points. Finally, according to this index, prices have decreased over the period;