NON-PARAMETRIC APPROACH TO THE COST-OF-LIVING INDEX

F. MAGNIEN et J. POUGNARD

INSEE, Division des prix à la consommation

Les débats récents sur une possible surestimation de l'inflation ont notamment porté sur l'ampleur du « biais de substitution » dans le calcul des indices de prix. Ce biais résulte de l'insuffisante prise en compte, avec un indice de Laspeyres, des transferts d'achats des consommateurs entre produits ou points de vente en fonction de l'évolution différenciée des prix.

Le biais de substitution peut être important au niveau « détaillé » des produits. Idéalement, il conviendrait de calculer un « indice à utilité constante » (IUC), qui mesure la variation de la dépense assurant au moindre coût le maintien du « niveau de vie » face à la variation des prix. Calculer un IUC est délicat : il est nécessaire de mettre en évidence une fonction d'utilité qui « rationalise » les données. Formellement, ce problème est résolu grâce à la théorie des « préférences révélées ». En pratique, il faut disposer de relevés très fins de prix et de quantités, ce que permettent aujourd'hui les données « scanner ».

Cette étude présente les résultats obtenus avec ce type de données : les choix des consommateurs, pris dans leur ensemble, sont effectivement rationnels. Il n'y a pas un, mais toute une « plage » d'IUC, dont les valeurs extrêmes coïncident de temps à autre avec les indices de Laspeyres et de Paasche ; cette plage contient presque toujours l'indice de Fisher.

1. Introduction

Consumer price indexes (CPIs) have recently been the focus of a polemic over a possible overstatement of inflation. The debate began in the United States around the Boskin Report (1996), then spread to Canada, the United Kingdom, and France. The issue at stake is crucial, because CPIs are used to revalue social benefits, maintain minimum-wage purchasing power, and adjust tax brackets (especially in the U.S.). The Boskin Report singles out new products and quality changes as the main causes of bias in the U.S., but other countries refuse to quantify either factor. Apart from these causes, experts attribute the bias mainly to the failure to measure substitutions, despite the use of chaining.[1] To be more precise, the *substitution bias* in the U.S. is most pronounced at the *detailed* level, i.e., a highly disaggregated level of product definition. Until recently (and sometimes still today), micro-indexes at that level were calculated with the aid of arithmetic means of price ratios, i.e., of Laspeyres indexes with equal weightings. The defect of this type of formula is that it overweights the goods with the steepest-rising prices. This generates a substitution bias that U.S. statisticians, as well as those of the European Union member States, have corrected by replacing the arithmetic mean of the price ratios by a geometric mean. The Boskin Report actually estimated the substitution bias at the detailed level by comparing the indexes obtained through the successive use of the two formulas. In fact, the adoption of the geometric formula and the resulting measure of the substitution bias are debatable—and have been challenged by the U.S. Bureau of Labor Statistics (BLS). The problem lies in the formula's assumption of a unit elasticity of substitution between products. Moulton (1996) suggested using utility functions with different elasticities of substitution for different products. The proposal was backed in France by Lequiller (1998). The elasticities could be estimated with the aid of *scanner data* supplied by marketing firms. Data scanned from bar codes give very detailed information on prices and quantities of staple consumer goods sold in mass-merchandise outlets.

Scanner data can be used not only to measure elasticities of substitution but also to calculate *cost-of-living indexes* (COLs). COLs were quickly perceived to be effective benchmarks, notably by the Boskin Commission—backed, on this point, by the BLS. An initial difficulty is that the computation of COLs requires data on quantities sold as well as on prices. Manser and McDonald (1988) used macro-data for which quantity series are available only at annual frequencies for the period 1959-85 in the U.S. Only micro-data offer quantity series at the same monthly frequency as prices in the CPI. In this study, we report the results obtained on such data. The series were supplied by the market research firm AC Nielsen, and concern three products sold in supermarkets and hypermarkets throughout metropolitan France : edible oils, detergents, and coffee. They cover the years 1994, 1995, and 1996.

---

[1] In the U.S., chaining is performed only once a decade.

Attempts to calculate a COL run into a major difficulty : the need to identify a utility function that rationalizes the data, i.e., a utility function such that, for each observation, the quantities acquired by consumers are optimal relative to the corresponding prices. Indeed, there is an even more fundamental question : does such a function exist? The tools for answering the question are supplied by the microeconomic theory of "revealed preferences," elaborated by Afriat (1967, 1977, 1981). Afriat defined the conditions in which a utility function that rationalizes the price and quantity data can exist. The quantities—and hence the revealed utility function—may apply to consumption by a single person or an entire population of consumers. Varian (1982, 1983) also contributed to the description of such conditions. Most important, he defined algorithms that allow the conditions to be verified by computer. Diewert and Parkan (1985) then applied these methods to macro-data.[2]

The problem of the existence of a data-rationalizing utility function having been solved, the next step—the actual calculation of COLs with such utility functions—was performed by Manser and McDonald (1988). They exploited the findings of previous authors to determine the set of COLs relating to the different *homothetic* data-rationalizing utility functions. These functions, when they exist, are not unique. It is not just one index but an entire "range" of COLs that needs to be computed. Both the bounds and the points inside the range are COLs. We felt it useful to offer a proof of the second point, which Manser and McDonald regard as self-evident.[3]

The Laspeyres index widens the COL range, while the Paasche index narrows it. Manser and McDonald were able to provide boundaries for the substitution bias in the U.S. index (which is of the Laspeyres type) over the period 1959-85. Our study, based on micro-data, finds that the substitution bias may diminish in some periods and increase in others. Moreover, the respective biases of the Paasche and Laspeyres indexes do not move in step : one may rise while the other falls. On numerous occasions, the biases become negligible (measured relative to the upper bound of the COL range for the Laspeyres index, to the lower bound for the Paasche index). The Laspeyres (Paasche) substitution bias is null when product substitutability is imperfect in the base period (current period) and when the relative change in prices is sufficiently small by comparison with this substitution imperfection.

Manser and McDonald have shown by empirical means that they were able to calculate COLs by using macro-data more disaggregated than those used previously. The micro-data we have used are even more disaggregated, and lead us to the same conclusions. At a highly aggregated level, there is no evidence of a homothetic data-rationalizing utility function, and the Laspeyres index fails to produce a wider COL

[2]They also contributed to the "theory": see Diewert (1973); Diewert and Parkan (1978).

[3]In all fairness, Manser and McDonald's main goal was simply to measure the substitution bias in the United States CPI.

range than the Paasche index. This phenomenon may, however, occur on strongly disaggregated data : we have observed it for coffee beans and espresso coffee.

The plan of our paper is as follows. Section 2 briefly outlines the theory of cost-of-living indexes. Section 3 describes the methodology for determining—from price and quantity data— the set of COLs relating to homothetic utility functions for a given current period and a given base period. Section 4 reports the results obtained on micro-data. Section 5 examines two issues concerning the existence of the COL : the need for adequate data disaggregation, and the "Paasche ≤ Laspeyres" inequality. Section 6 sets out the proofs of the results in Section 3.

2. The cost-of-living index

The aim of this section is to provide a brief survey of the *cost-of-living index* (COL) theory. We recall the strict definition of the COL, show that the COL is bounded by the Paasche and Laspeyres indexes, and supply the COLs associated with the main homothetic utility functions. All these results are discussed in greater detail in Diewert (1981).

For a *finite* set E of periods $t = 1, \dots, T$ and for a set S of n s varieties, we have quotations for prices p_t^s and quantities sold q_t^s. It will be recalled that a variety is defined as a datum on a product in a sales outlet. Let $p_t = (p_t^s)_{s\in S} \in R_{++}^n$ and $q_t = (q_t^s)_{s\in S} \in R_+^n$ be the price and quantity vectors for the set of varieties.[4]

2.1. COL associated with homothetic utility function

To calculate a cost-of-living index, we need a utility function $U : R_+^n \to R_+$ that *rationalizes* the data $(p_t, q_t)_{s\in E}$:

$$U(q_t) = \text{Max}\,\{U(q), q \in R_+^n \text{ and } p_t\, q \le p_t\, q_t\} \quad \text{for all } t \in E$$

In other words, in each period $t \in E$, the bundle q_t must be optimal (for U) with respect to the price vector p_t and under the expenditure constraint $p_t q_t$. The *indifference curve* $\{q \in R_+^n, U(q) = U(q_t)\}$ of a utility function of this kind is shown in figure 1. The COL in a period t' by comparison with period t, for a given utility level u, is defined as the ratio of two expenditures : the minimal expenditure that

[4] R denotes the set of real numbers, R_+ the set of non-negative real numbers, and R_{++} the set of positive numbers.

enables the consumer to reach the utility level u in period t' (t) when prices are $p_{t'}$ (p_t).

To express the COL in mathematical terms, we therefore need to introduce the minimal cost $C_U(u, p)$ that enables the consumer to reach a given utility level u under an exogenous price system $p \in R^n_{++}$:

$$C_U(u, p) = \text{Min}\ \{pq,\ q \in R^n_+ \ \text{ and } \ U(q) \geq u\}$$

The COL between two periods t and t' is defined as :

$$\text{COL}_{t'/t}\,(U, u) = C_U(u, p_{t'}) / C_U(u, p_t)$$

The quantities Q_t and $Q_{t'}$ implicit in the calculation of $\text{COL}_{t'/t}\,(U, u)$:

$$p_t Q_t = C_U(u, p_t) \quad \text{and} \quad p_{t'} Q_{t'} = C_U(u, p_{t'})$$

are shown in figure 2.

The COL's major drawback is its dependence on the benchmark "standard of living" u. At best, we can assume that the latter must lie somewhere between the base-period utility and the current-period utility. However, the COL does not depend on the utility level u when[5] the utility function U is *homothetic,* i.e.

$$U(q) = U(q') \;\Rightarrow\; U(\lambda q) = U(\lambda q') \quad \text{for all } \lambda > 0 \text{ and all } q, q' \in R^n_+ .$$

This fact is self-evident in the special case where U is *homogeneous* (of degree 1)

$$U(\lambda q) = \lambda U(q) \quad \text{for all } \lambda > 0 \text{ and all } q \in R^n_+$$

as, in that case, $C_U(u, p) = uC_U(1, p)$. Hence :

$$\text{COL}_{t'/t}\,(U, u) = u\, C_U(1, p_{t'}) \,/\, u\, C_U(1, p_t) = C_U(1, p_{t'}) \,/\, C_U(1, p_t)$$

which is written succinctly as $\text{COL}_{t'/t}\,(U)$ or $\text{COL}_{t'/t}$ when there is no ambiguity about the homothetic utility function underlying the COL calculation. Among the most important examples of homothetic (actually, homogeneous) utility functions are the *Cobb-Douglas* functions and *complementary-factors* functions (two special cases of a more general class of homothetic functions, the CES functions) and *quadratic* functions.

[5]...and only when: see Diewert (1981).

It is easy to compare the COL with the Laspeyres and Paasche indexes. As U rationalizes the data, we have[6] :

$$p_t\, q_t \;=\; C_U(U(q_t), p_t) \quad \text{for all } t \in E$$

Hence, if U is also homothetic :

$$\mathrm{COL}_{t'/t}(U) \;= C_U(U(q_t), p_{t'}) \,/\, p_t\, q_t \;= p_{t'}\, q_{t'} \,/\, C_U(U(q_{t'}), p\, p_t\, q_t).$$

Since, under the definition of the cost function C_U, we have :

$$C_U(U(q_t), p_{t'}) \;\leq\; p_{t'}\, q_t \quad \text{and} \quad C_U(U(q_{t'}), p_t) \;\leq p_t\, q_{t'}$$

it follows that the Paasche and Laspeyres indexes bound the COL :

$$P_{t'/t} \leq \mathrm{COL}_{t'/t}(U) \leq L_{t'/t}$$

where $P_{t'/t} = p_{t'}\, q_{t'} / p_t\, q_{t'}$ and $L_{t'/t} = p_{t'}\, q_t / p_t\, q_t$.

[6] Under some additive hypotheses on U such as its local nonsatiatedness: for every $q \in R^n_+$ and every neighborhood W of q there exists $q' \in W$ such that $U(q') > U(q)$.

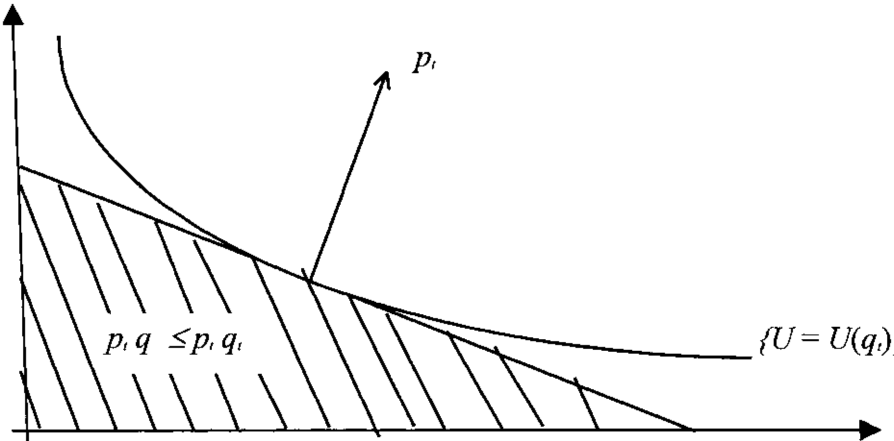


Figure 1: Indifference curve of a data-rationalizing utility function

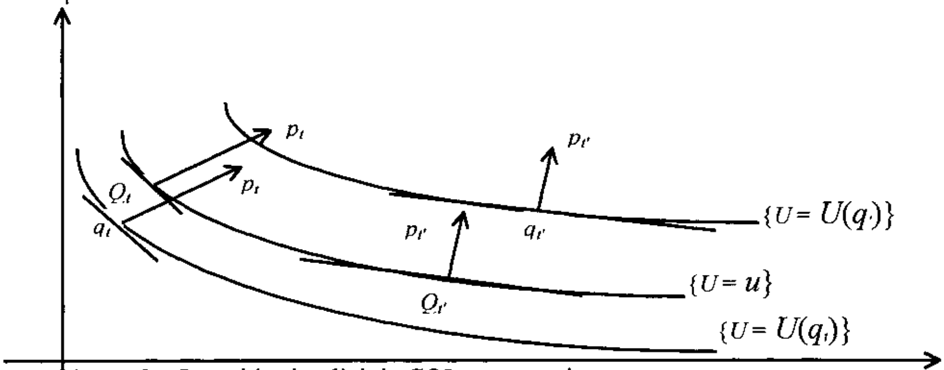


Figure 2 : Quantities implicit in COL computation

2.2. A few basic examples[7]

A basic example of a homogeneous utility function is the *Cobb-Douglas* function :

$$U(q) = a\prod_{s\in S}(q^s)^{\alpha^s} \quad \text{with} \quad a>0,\ \alpha^s>0 \text{ and } \sum_{s\in S}\alpha^s = 1$$

When such a utility function rationalizes the data, the associated COL is the *weighted geometric mean* :

$$G_{t'/t} = \prod_{s\in S}\left(\frac{p_{t'}^s}{p_t^s}\right)^{w_t^s} \quad \text{where} \quad w_t^s = \frac{q_t^s p_t^s}{\sum_{s'\in S} q_t^{s'} p_t^{s'}} = \alpha^s$$

Another important example is the *quadratic* utility function :

$$U(q) = \left[\sum_{s,s'\in S} a_{s,s'} q^s q^{s'}\right]^{1/2}$$

where $(a_{s,s'})$ is a symmetric matrix (whose coefficients are properly chosen). If a function of this type rationalizes the data, the corresponding COL is the *Fisher* index :

$$F_{t'/t} = \sqrt{L_{t'/t} P_{t'/t}} \,.$$

But the simplest example is still the utility function with *complementary factors :*

$$U(q) = \text{Min}\{\frac{q^s}{a^s},\ s\in S\} \quad \text{where } a^s > 0 \text{ for all } s\in S$$

(figure 3). The associated COL is, quite simply, the *Laspeyres* index—but also the *Paasche* index, since in this case the two coincide.

[7] See Diewert (1981)

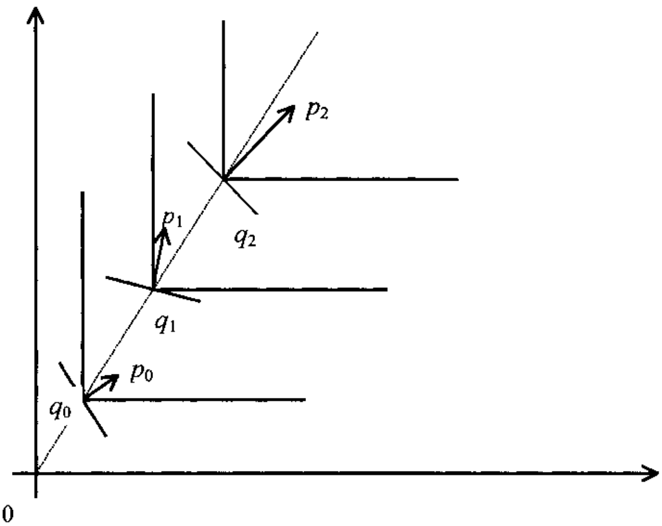


Figure 3: Indifference curve of a utility function with complementary factors

3. Calculating the COL : the non-parametric approach

The cost-of-living index implies the existence of a "rationalizing" homothetic utility function : the calculation of such an index from price and quantity quotations (§4) therefore raises the following issues :

(1) Does such a utility function exist?

(2) If it does, find for each pair of periods $(t, t') \in E^2$ the set of cost-of-living indexes $COL_{t'/t}(U)$ associated with the different U solutions to question (1).

3.1. The theory of revealed preferences

Question (1) was first studied by Afriat, founder of the "revealed preferences" theory.[8] The theory has also been refined by Diewert (1973) and Varian (1982, 1983). These authors have developed a fairly simple condition—simple at least in its formulation—that provides an answer to the question "does a *homothetic* data-rationalizing utility function exist?" :

[8] Afriat (1967,1977,1981).

Theorem 1. *There exists a homogeneous, continuous, concave, and non-satiated utility function that rationalizes the data $(p_t, q_t)_{t \in E}$ if and only if the following condition is met :*

$$\frac{p_i q_j}{p_i q_i} \frac{p_j q_k}{p_j q_j} \cdots \frac{p_m q_i}{p_m q_m} \geq 1 \qquad (1)$$

for all "cycles" of periods $i, j, k, \ldots, m, i \in E$.

This is known as the HARP (Homothetic Axiom of Revealed Preference) condition. The cycle of periods $i, j, k, \ldots, m, i$ is not necessarily arranged in chronological order; moreover, it is of arbitrary length. The left-hand member of inequality (1) is a chained Laspeyres index of *quantities*. Like the price indexes of its kind, this index has a "circularity defect," i.e., it does not return to 1 if the quantities return to their initial values. The HARP condition implies that the circularity defect is always in the same direction.

The first problem, therefore, has found a theoretical solution. In practice, the verification of this condition is extremely cumbersome when the set of periods E exceeds a few units. Varian (1982, 1983) has proposed an algorithm that allows such a verification (§3.3).

3.2. Expressing the range of COLs

Having established the condition for the existence of a homothetic data-rationalizing utility function, we have to determine the associated COL between any two periods in E. In fact, there are usually several rationalizing utility functions (because the E set is *finite*). This means there is no unique COL between two given periods. We must therefore identify the complete set of these COLs. As we have seen (§2.1), the set is bounded from above—usually strictly—by the direct Laspeyres index and bounded from below by the Paasche index. Manser and McDonald (1988) have determined the exact bounds of the COL range between any two periods :

Theorem 2. Let us assume the HARP condition is satisfied. The set of COLs of period i calculated relative to period j is the closed interval $[1/\Delta_{ji} ; \Delta_{ij}]$ where

$$\Delta_{ij} = \frac{p_i q_i}{p_j q_j} Min_{i,k,l..,m,j} \left\{ \frac{p_i q_k}{p_i q_i} \frac{p_k q_l}{p_k q_k} \cdots \frac{p_m q_j}{p_m q_m} \right\}$$

for any sequence of periods $i, k, l, \ldots, m, j \in E$.

In this formulation, the COL relate to the homothetic data-rationalizing utility functions. The period i does not necessarily follow j in time. Manser and McDonald appear to have taken for granted the fact that all the points in the interval $[1/\Delta_{ji} ; \Delta_{ij}]$ are COLs :

Proposition 1. Let us assume the HARP condition is met. If U and U' are homothetic data-rationalizing utility functions, then for all periods i, j $\in$ E and all $\lambda \in]0, 1[$, there exists a homothetic data-rationalizing utility function U" such that

$$COL_{i/j} (U") = \lambda\, COL_{i/j} (U) + (1-\lambda) COL_{i/j} (U').$$

We give the proof of this result in §6. The utility function U" depends on the periods i and j (and on λ). It may be assumed to be linear in segments, as we obtain the following result (§6) :

Proposition 2. For any homothetic utility function U rationalizing the data $(p_t, q_t)_{t \in E}$ and for any pair of periods $(i, j) \in E^2$, there exists a utility function V of the form

$$V(q) = \mathrm{Min}\{ \lambda_t p_t q , t = 1, \ldots , T \}$$

where $\lambda_t > 0$ (t = 1, ... , T) such that V rationalize the data and $COL_{i/j}$ (U) = $COL_{i/j}$ (V).

Figure 4 plots the indifference curves for such a utility function. We observe that the tangents to the indifference curves along a radius issuing from the origin are parallel. This is a general property of homothetic utility functions. The finding will be very useful for analyzing the results on micro-data (§4.3).

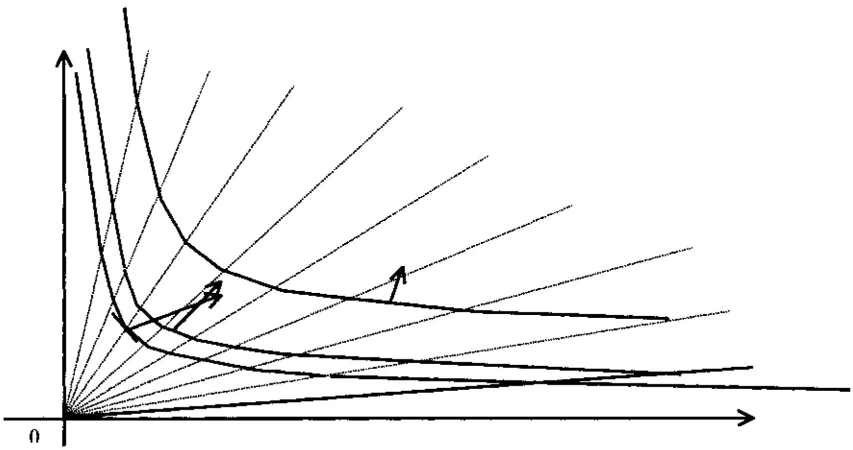


Figure 4: Indifference curve of a homothetic data-rationalizing utility function that is linear in segments

3.3. Algorithm

Given the large number of varieties, it is impossible to verify the HARP condition or to calculate the COL range (cf. Theorem 2) without a suitable algorithm. Like Varian (1983), we have used Warshall's algorithm (1962), which combines a matrix $M = (m_{ij})$ with the matrix $D = (d_{ij})$ such that

$$d_{ij} = \inf_{i,\ k, l, \dots, m,\ j} (m_{ik} + m_{kl} + \dots + m_{mj}).$$

In other words, if m_{ij} represents the "cost" of *direct* passage from i to j, then d_{ij} represents the *minimal* "cost" of passage from i to j for the set of paths leading from i to j. We can show that the matrix D is obtained as follows :

(1) let $k = 1$
(2) for all i, j : if $m_{ij} \geq m_{ik} + m_{kj}$ let $m_{ij} = m_{ik} + m_{kj}$
(3) if $k < T$ let $k = k + 1$ and go to (2); if not, let $d_{ij} = m_{ij}$ for all i, j. END

Warshall's algorithm enables us to calculate the quantities

$$\inf_{i,\ k, l, \dots, m,\ j} \ln \left\{ \frac{p_i q_k}{p_i q_i} \frac{p_k q_l}{p_k q_k} \dots \frac{p_m q_j}{p_m q_m} \right\}$$

by positing $m_{kl} = \ln \dfrac{p_k q_l}{p_k q_k}$ then, by applying the exponential function (which, like the log function, is increasing), the quantities

$$\inf_{i,\ k, l, \dots, m,\ j} \left\{ \frac{p_i q_k}{p_i q_i} \frac{p_k q_l}{p_k q_k} \dots \frac{p_m q_j}{p_m q_m} \right\}.$$

We can immediately deduce the bounds $1/\Delta_{ji}$ and Δ_{ij} of the COL range. Warshall's algorithm obviously allows us to determine if the HARP condition is met : if, in a step k, we have $m_{ii} < 0$ for an i, then the HARP condition is not met.

4. Application to micro-data

We will now apply the results of §3 to the calculation of the COLs associated with the different homothetic utility functions that rationalize *micro-data*. These consist of scanner data supplied by the AC Nielsen firm and covering three products : coffee, edible oils, and detergents.

4.1. Data description

The prices and quantities were collected weekly over the period 1994-96. Quotations were gathered in more than 400 supermarkets and hypermarkets. The varieties are accompanied by highly detailed product descriptions. These enabled us to define a concept of *elementary product* corresponding to a very fine level of disaggregation. At the end of the exercise, we had 523 distinct elementary products for edible oils, 383 for detergents, and 1,168 for coffees! The advantage of such detail is that we can fully track product substitutions. The downside is that the cross-tabulation of the elementary products and the sales outlets yields a very high number of varieties, a sizable proportion of which are often unobservable. This problem is solved in two ways : (1) by a monthly aggregation of the data (summing the quantities and calculating a weighted price average); (2) by reaggregating the outlets into four categories (summing the quantities and calculating a weighted price average) : small (large) supermarkets and hypermarkets. For each of these four types, we add up the elementary products sold in each month of the period studied (January 1994 – December 1996) in at least one outlet. This eliminates new products and products temporarily or permanently withdrawn from any one of the four types of outlets. We obtained 138 elementary products for oils, 133 for detergents, and 353 for coffees, accounting for 91%, 75%, and 92% respectively of sales before simplification. In the end, for each of the thirty-six months of the period examined, we obtained the average selling prices and average quantities sold for 333 *varieties* of oils, 424 varieties of detergents, and 895 varieties of coffee.[9]

4.2. Results

The results of the COLs obtained on these micro-data are reported in charts 1a, 1b, and 1c. The charts show, for each of the three products, the range of COLs in a given month relative to January 1994. For example, with January 1994 (month 1) as the base (=100), the COLs for edible oils in August 1995 (month 20) take on all the values between 107 and 107.75.

---

[9]Magnien and Pougnard (1998).

It should be remembered that these COLs are associated with the different *homothetic* utility functions that "rationalize" the data pertaining to the largest possible set of periods. They are not the utility functions that rationalize the data ranging only from the base month, January 1994 (month 1) to the month of study. And they are even further removed from the utility functions that rationalize only the data of the base month and the month of study. For coffee, we were able to rationalize all the data (month 1-36); for oils and detergents, we were only able to rationalize months 1-26 and 1-27 respectively.

The range does not widen with time. In fact, it can even narrow to practically a single point, as in the case of detergents in month 13. The Fisher index lies consistently inside the COL range for coffee (charts 2b). This does not mean, however, that a quadratic utility function rationalizes the data. Conversely, it is clear that there is no such rationalizing utility function for oils and detergents, since in several months the Fisher index departs slightly from the COL range (charts 2a and 2c). No Cobb-Douglas function rationalizes the data, whether for coffee, oils, or detergents. In all three cases, the index of the weighted geometric mean lies outside the COL interval.

Chart 1a : COL range for edible oils

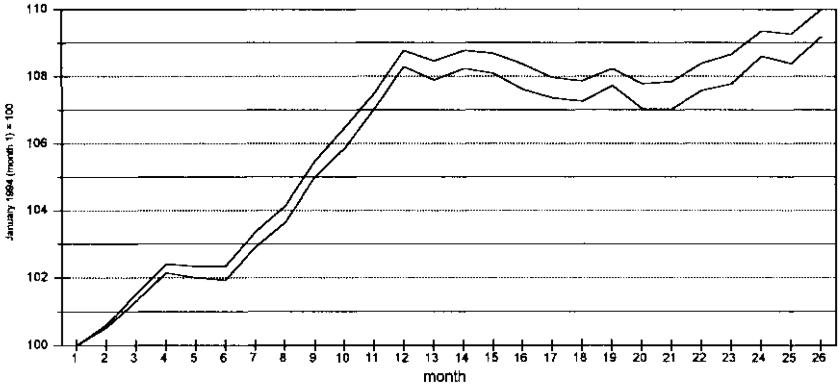


Chart 1b : COL range for coffee

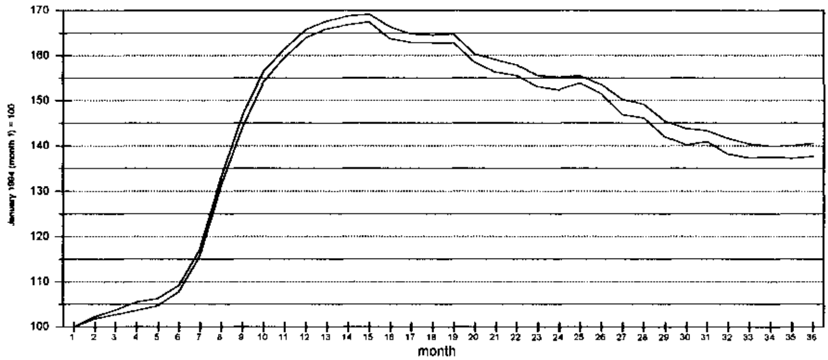


Chart 1c : COL range for detergents

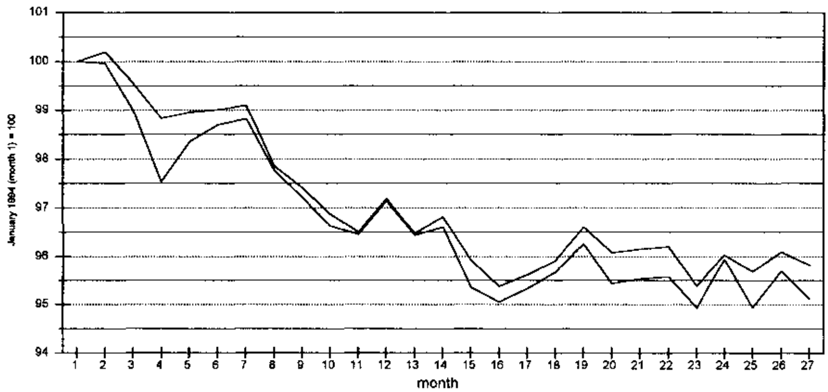


Chart 2a: Fisher, geometric mean, and
COL range for edible oils

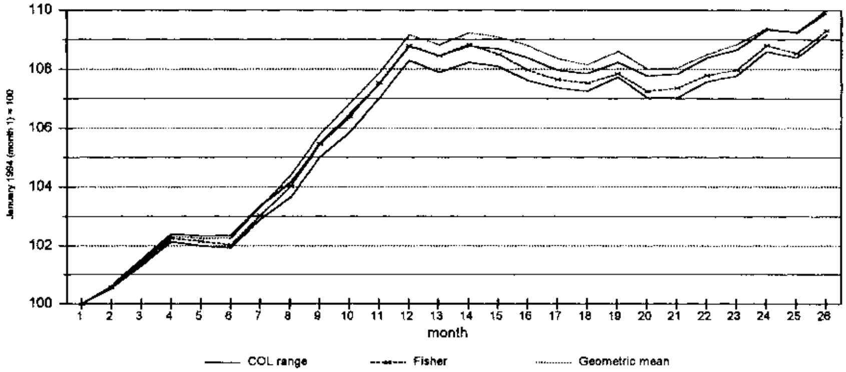


Chart 2b: Fisher, geometric mean, and
COL range for coffee

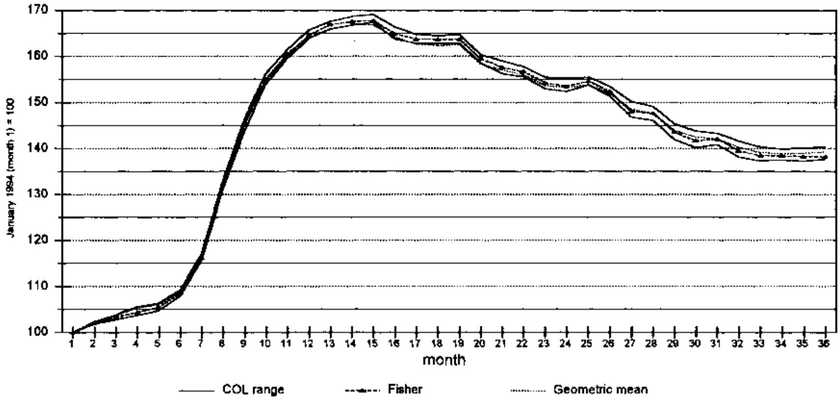
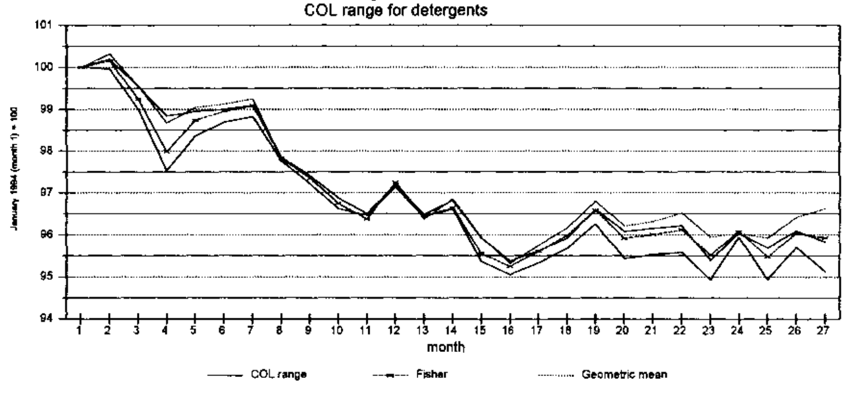


Chart 2c: Fisher, geometric mean, and
COL range for detergents



4.3. Relative variation in prices and product substitutability

The L-P interval is strictly larger than the COL bounds for edible oils (chart 3a) over most of 1995 (months 15-24). However, in months 9-14, the Paasche index coincides (or nearly) with the smallest COL whereas the Laspeyres index lies well above the largest. This phenomenon occurs during a period of sharp price increases. We observe the same phenomenon with detergents in months 16-24 (chart 3c). Symmetrically, in month 6 for edible oils, months 4 and 6 for detergents, and months 32-36 for coffee (chart 3b), we observe a situation in which the Laspeyres index coincides with the largest COL, whereas the Paasche index is strictly inferior to the smallest.

In sum, the respective substitution biases of the Paasche and Laspeyres indexes, measured relative to the COL range, move rather differently from each other but also over time. They can become null or nearly null in different periods or simultaneously. Diewert (1990, p. 87) has indicated that the weakness of the Laspeyres (Paasche) index bias relative to the COL was due to an appropriate relationship, in the base period (current period) between the *relative* price variation and the *substitutability* of the varieties (products or outlets). We will show that the rationalization of data by the *linear-in-segments* utility functions (cf. Proposition 2)—whose indifference curves consequently exhibit *sharp inflections*—lead to an extreme application of this analysis.

For this, let us consider the COL pertaining to a homothetic utility function U of the form described in Proposition 2. We have :

$$\mathrm{COL}_{t'/t}(U) = C_U(U(q_t), p_{t'}) / p_t q_t = p_{t'} q_{t'} / C_U(U(q_{t'}), p_t).$$

(cf. § 2.1). Let $Q_{t'}(Q_t)$ be a commodity bundle such that $p_{t'} Q_{t'} = C_U(U(q_t), p_{t'})$ ($p_t Q_t = C_U(U(q_{t'}), p_t)$. These bundles are shown in figure 5a. We have :

$$\mathrm{COL}_{t'/t}(U) = p_{t'} Q_{t'} / p_t q_t = p_{t'} q_{t'} / p_t Q_t .$$

Even if the relative variation of prices is small (in other words, if the direction of vector $p_{t'}$ differs little from that of p_t), the composition of the bundle Q_t (which procures the same utility as $q_{t'}$ but under the price system p_t) is distinct from that of $q_{t'}$. This is due to the strong substitutability between varieties in the neighborhood of $q_{t'}$, where the utility function is locally linear. The value $p_t Q_t$ of bundle Q_t is thus below $p_t q_{t'}$. The Paasche index $P_{t'/t}$ (equal to $p_{t'} q_{t'} / p_t q_{t'}$) is therefore below $\mathrm{COL}_{t'/t}(U)$ (equal to $p_{t'} q_{t'} / p_t Q_t$). By contrast, owing to the strong bend in the utility function in q_t (where the indifference curve displays a sharp inflection[10]), i.e., owing to the low substitutability between varieties from this bundle, the same bundle ($Q_{t'} = q_t$) maintains the utility level $U(q_t)$ under the price system $p_{t'}$ despite the fact that the system is moving in a different direction from p_t. The Laspeyres index, therefore, is equal to the upper bound of the COL range :

$$L_{t'/t} = p_{t'} q_t / p_t q_t = p_{t'} Q_{t'} / p_t q_t = C_U(p_{t'}, U(q_t)) / p_t q_t = COL_{t'/t}(U).$$

[10]So that, locally, the utility function resembles a "complementary-factors" function (figure 3).

Chart 3a: COL range and L-P bounds
for edible oils

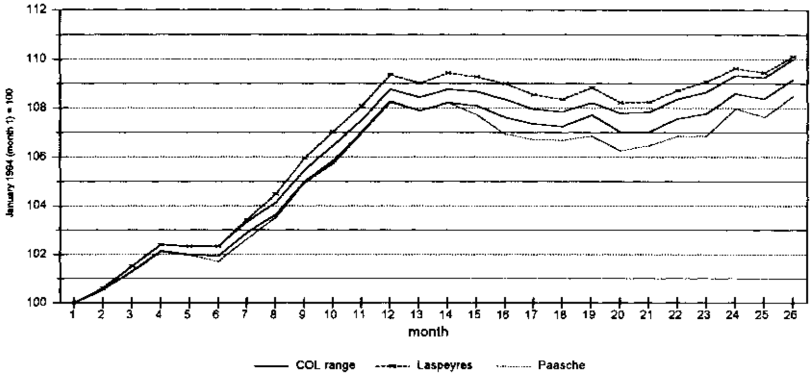


Chart 3b: COL range and L-P bounds
for coffee

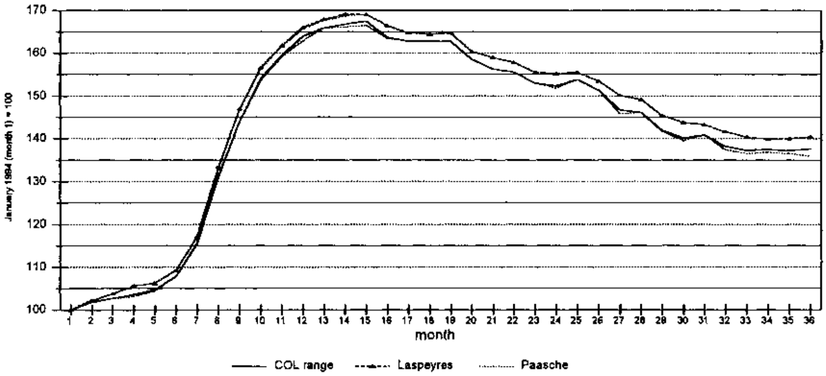
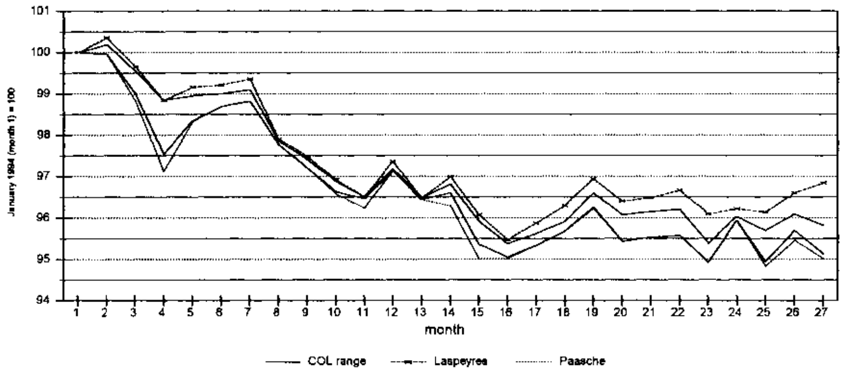


Chart 3c: COL range and L-P bounds
for detergents



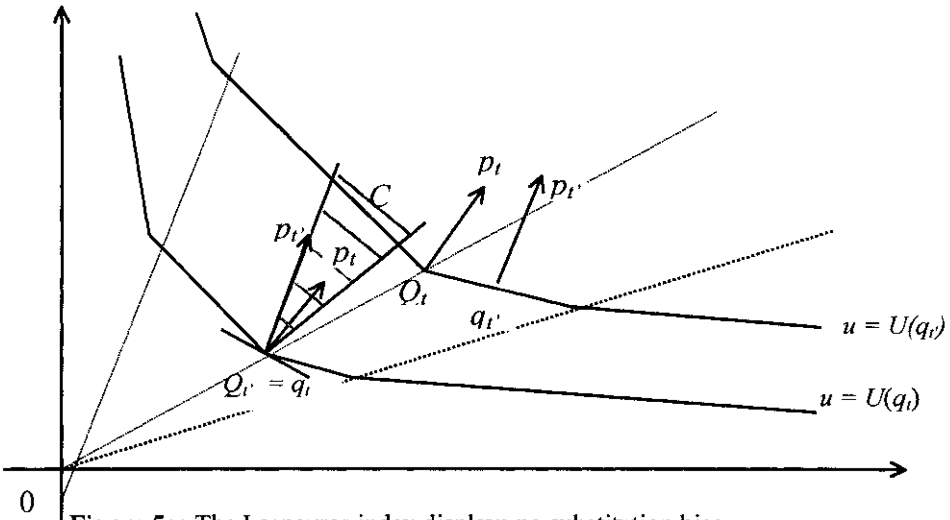


Figure 5a: The Laspeyres index displays no substitution bias

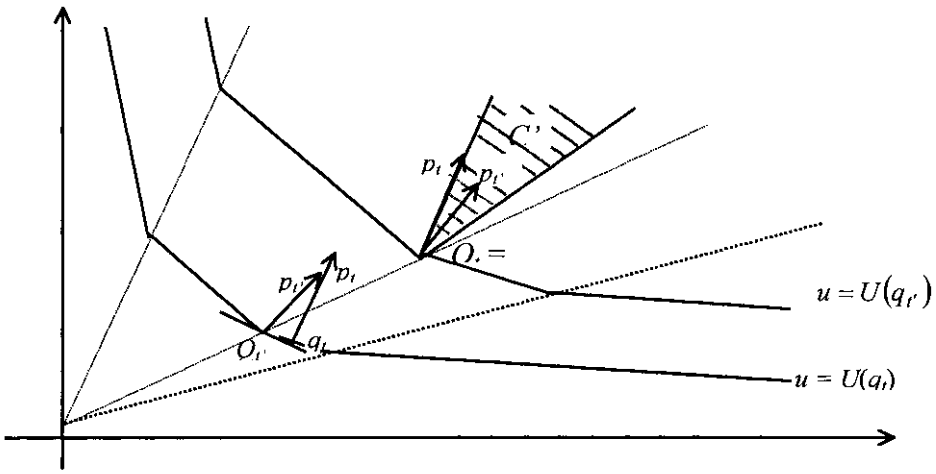


Figure 5b: The Paasche index displays no substitution bias

For the two values to coincide, however, the *relative* variation in prices between t and t' (measured by the "gap" between the directions of p_t and $p_{t'}$) must not be too great. It should lie within the limits allowed by the "degree" of substitutability of the products in q_t, measured by the "bend" of the indifference curve in that point. Specifically, $p_{t'}$ must belong to the *cone C* of prices p such that :

$p\,q_t = C_U(p, U(q_t))$. The cone is shown in figure 5a.

Figure 5b illustrates the symmetrical case of a homothetic utility function for which the COL coincides with the Paasche index while staying below the Laspeyres index.

The conjunction of the two cases described above (existence of a homothetic utility function for which the COL is equal to the Paasche index and another for which the COL is equal to the Laspeyres index) occurs frequently with coffee (chart 3b).

5. Existence of the COL and data disaggregation

5.1. The need for effective data disaggregation

Given the apparently restrictive character of the homotheticity condition, it may seem surprising that the HARP condition is satisfied for such long periods as the 36 months tracked for coffee, 27 months for detergents, and 26 months for edible oils. It is worth noting the narrowness of the COL range : approximately less than one index point for oils and detergents, no more than three points for coffee (coffee prices, however, registered very wide swings).

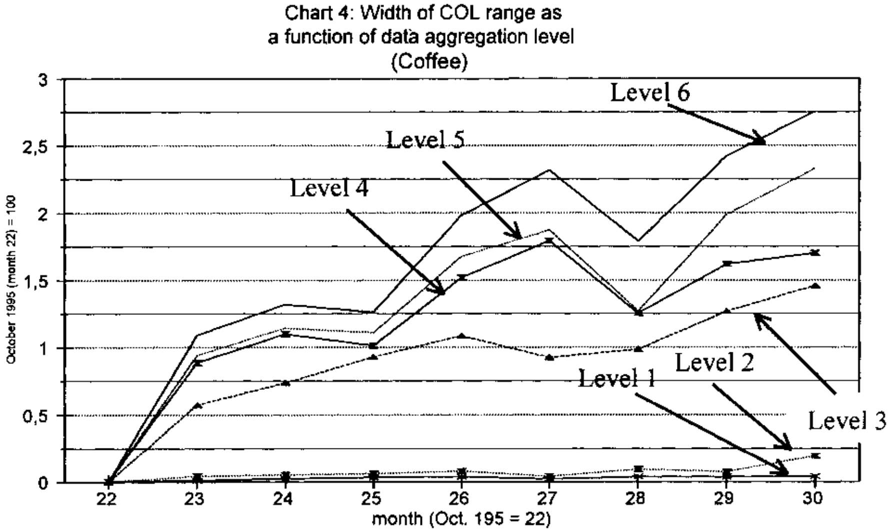
On macro-data, Manser and McDonald (1988) have tried to explain the existence of "rationalizing" homothetic functions by the high disaggregation of the data they used. In other words, the robustness of the HARP condition test would be due to the great detail of the micro-data. To verify this, we tested the condition at various levels of aggregation of coffee and detergent data. The levels are determined by the list of variety characteristics. Besides the type of sales outlet (of which there are four), they consist of the following : producer (or distributor), brand, reference, packaging, and content. Each of these characteristics has different modalities : packaging is defined by the type of container (jar, can, box), the number of packs sold together, and the total weight. Content description for coffee is even more varied : form (beans, ground, espresso), quality (regular or decaffeinated), botanical variety (arabica, robusta, or blend), and provenance. These different contents have been assembled into a smaller number of "items."[11]

[11]Magnien and Pougnard (1998).

The aggregation levels adopted are listed in the table below :

		Coffee		Detergents	
Data aggregation level	**Characteristics adopted**	**Number of varieties**	**Rationalized period**	**Number of varieties**	**Rationalized period**
1	Item	5	22-30	12	13-15
2	Item and sales outlet type (SO)	20	9-32	48	12-16
3	Item, SO, and producer	264	8-36	216	1-20
4	Item, SO, producer, and brand	329	8-36	320	1-23
5	Item, SO, producer, brand, and reference	562	1-36	-	-
6	All	895	1-36	424	1-27

The table also shows the number of varieties for each level and the longest period for which the HARP condition is met, i.e., for which the data are optimal for a continuous, concave, non-satiated homogeneous utility function. The figures effectively show that the period length increases with the level of disaggregation. Moreover, the COL range also widens with the level of disaggregation (chart 4). With detergents, at the most aggregated level (level 1), only three consecutive months were rationalized. In many instances, the HARP condition test actually failed over a two-month period. In these cases, the Paasche indexes (i.e., the ratio of one of the two months to the other) exceeds the Laspeyres indexes.[12] We analyze this phenomenon in §5.2.



[12]These indexes are calculated on aggregated data.

5.2. The P ≤ L inequality

For the Paasche index to be smaller than or equal to the Laspeyres index, one condition *suffices* : the existence of a *homothetic* utility function that rationalizes the data for *two* periods : the base period and the current period (but not necessarily the others). This finding was established in §2.1. Theorem 1 shows that this condition is also *necessary*. The inequality $P_{j/i} \leq L_{j/i}$ i.e. :

$$\frac{p_j q_j}{p_i q_j} \leq \frac{p_j q_i}{p_i q_i}$$

is also written :

$$\frac{p_i q_j}{p_i q_i} \frac{p_j q_i}{p_j q_j} \geq 1$$

which is, quite simply, the HARP condition for the set of periods i and j.

Proposition 3. The Paasche index for t' relative to t is smaller than or equal to the Laspeyres index if and only if there exists a utility function that rationalizes the data (p_t, q_t) and ($p_{t'}$, $q_{t'}$).

Hence the P ≤ L inequality describes the situation of minimal consumer rationality.

For coffee, oils, or detergents, the inequality is satisfied with the entire set of micro-data at their most disaggregated level. If, however, we restrict the analysis to—for example—the "coffee beans" or "espresso coffee" items, and again using disaggregated data, the Laspeyres index is sometimes smaller than or equal to the Paasche index (charts 5a and 5b).

Chart 5a: Coffee beans
Paasche and Laspeyres indexes

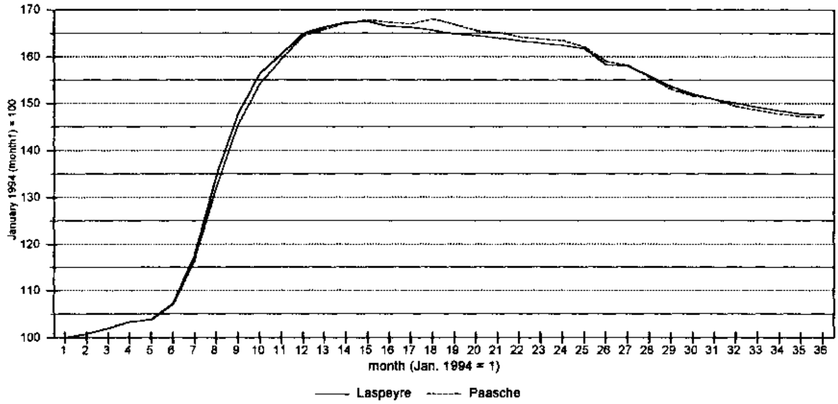
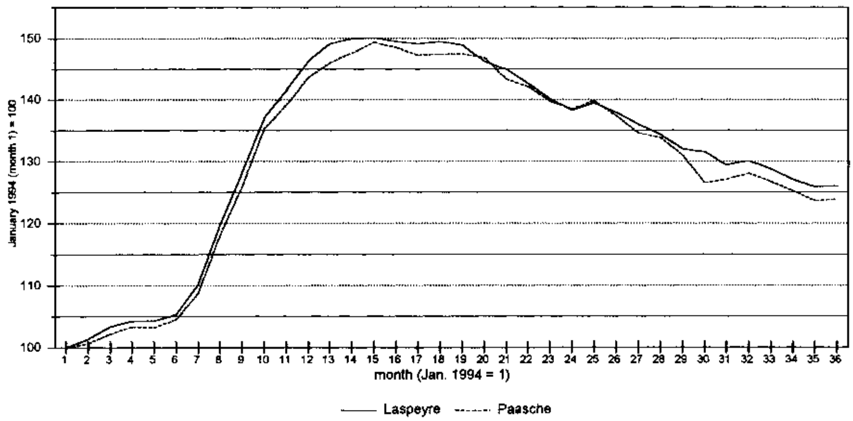


Chart 5b: Espresso coffee
Paasche and Laspeyres indexes



For coffee beans, we combined the quotations for arabica (higher-grade and thus more expensive products) and quotations for blends. We obtained the following price and quantity changes between months t = 1 (the base month) and t' = 18 (June 1995, the month with the sharpest index inversion) :

Coffee beans	average price[1] in month 1 (p_1)	average price[1] in month 18 (p_{18})	change in quantities
arabica	10.3	15.3	- 40%
blend	5.0	8.7	- 21%

1. French francs

In relative terms, however, the larger price increase was for blended coffees :

$$p_{18}^{blend} / p_{18}^{ara.} = 0.57 \text{ and } p_1^{blend} / p_1^{ara.} = 0.49$$

while arabica registered the steeper fall in quantities sold. The movements are plotted in figure 6. This shows a utility function rationalizing the data for dates t and t'. But the function cannot be homothetic : the price and quantity configuration is such that the tangent in A to the initial indifference curve cannot be parallel to the tangent to the indifference curve in $q_{t'}$ (likewise, the tangent in B to the current indifference curve cannot be parallel to the tangent to the indifference curve in q_t). This contradicts the "parallelism" property of the tangents to the indifference curves, which characterizes homothetic utility functions (figure 4).

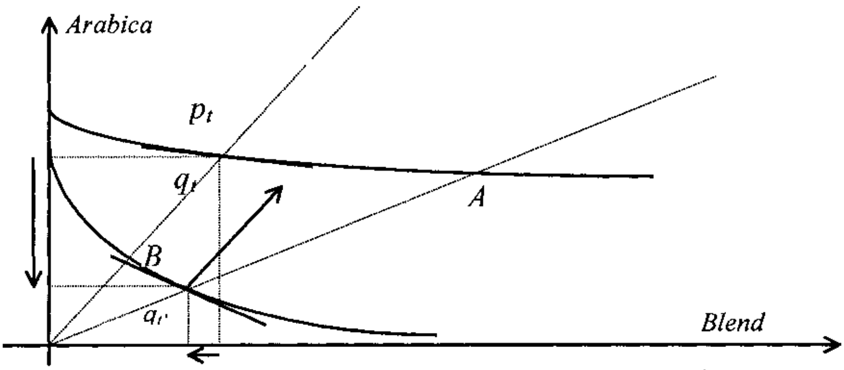


Figure 6 : The paasche index can exceed the Laspeyres index

6. Mathematical demonstrations

The purpose of this section is to demonstrate Theorems 1 and 2 and Propositions 1 and 2. These proofs are based on four lemmas. Lemmas 1 and 2, as well as their proofs supplied here, are taken from Afriat (1981). Lemma 3 combines classic results (from no specific authors) : we have tried to prove it in the simplest possible way. We admit Lemma 4, which allows a demonstration of the necessity of the HARP condition, for this result is not employed in our study. For its proof, we refer the reader to Varian (1983). Proposition 1, and therefore its proof published here, seem to us to be a new finding.

The theory rests on the minimization of "chained Laspeyres" indexes for quantities. The following values

$$A_{ij} = Inf_{i,k,l,\dots,m,j} \left\{ \frac{p_i q_k}{p_i q_i} \frac{p_k q_l}{p_k q_k} \dots \frac{p_m q_j}{p_m q_m} \right\}$$

effectively play a key role. The sequences of periods i, k, l, ... , m, j are not necessarily arranged in chronological order and their length is random, so that the *inf* is not necessarily a *min*.

Lemma 1 (Afriat, 1981) : *If the HARP condition is met, then :*

(*i*) $A_{ij} = Min_{i,\, k,l,\dots,m,\, j} \left\{ \frac{p_i q_k}{p_i q_i} \frac{p_k q_l}{p_k q_k} \dots \frac{p_m q_j}{p_m q_m} \right\}$ (A_{ij} *is therefore a* minimum)

(*ii*) $A_{ij} > 0$ for all $i, j \in E$;

(*iii*) $A_{ij} A_{jk} \geq A_{ik}$ for all $i, j, k \in E$;

(*iv*) $A_{ii} = 1$ for all $i \in E$;

(*v*) $A_{ji} \geq 1/A_{ij}$ for all $i, j \in E$.

Proof. (*i*) According to the HARP condition, in the expression A_{ij} we can assume the periods k, l, ... , m are, in pairs, distinct and different from i and j. The lower bound is therefore a minimum. Condition (*ii*) results, trivially, from (*i*). According to (*i*), there exists a sequence of periods i, l, m, ... , n, j and a sequence of periods j, r, s, ... , t, k such that

$$A_{ij} = \frac{p_i q_k}{p_i q_i} \frac{p_k q_l}{p_k q_k} \cdots \frac{p_m q_j}{p_m q_m}$$

and

$$A_{jk} = \frac{p_j q_r}{p_j q_j} \frac{p_r q_s}{p_r q_r} \cdots \frac{p_t q_k}{p_t q_t}$$

From the definition of A_{ik} we get :

$$A_{ik} \leq \frac{p_i q_l}{p_i q_i} \frac{p_k q_l}{p_k q_k} \cdots \frac{p_m q_j}{p_m q_m} \cdot \frac{p_j q_r}{p_j q_j} \frac{p_r q_s}{p_r q_r} \cdots \frac{p_t q_k}{p_t q_t} = A_{ij} A_{jk}$$

for i, l, m, ... , n, j, r, s, ... , t, k is a sequence of periods from i to k. The property (*iii*) is therefore satisfied. According to HARP, $A_{ii} \geq 1$. Moreover, by definition, $A_{ii} \leq \frac{p_i q_i}{p_i q_i} = 1$, hence (*iv*). The property (*v*) results from (*iii*) and (*iv*). ♦

Lemma 2 (Afriat, 1981). *Let* $i, j \in E$. *If the HARP condition is met, there exists a set of numbers* $U_1, \ldots, U_T > 0$ *such that* :

(*i*) $U_l / U_k \leq A_{kl}$ *for all* $l, k \in E$;

(*ii*) $U_i / U_j = A_{ji}$.

Proof. Let $U_k = A_{jk}$ for all k in E. Then equality (*i*) of Lemma 2 follows immediately from inequality (*iii*) of Lemma 1. ♦

Lemma 3. *Let $U_1, \dots, U_T > 0$ such that :*

$$U_l / U_k \le \frac{p_k q_l}{p_k q_k} \quad \textit{for all } l, k \in E.$$

(*i*) *Condition* (*i*) *of Lemma 2 is satisfied;*

(*ii*) *The utility function defined by*

$$U(q) = \ Min\left\{U_i \frac{p_i q}{p_i q_i} \ , \ i \in E\right\}$$

is homogeneous, concave, monotonic, continuous, non-satiated, and data-rationalizing;

(*iii*) $U(q_i) = U_i$ *and* $C(p_i) = \dfrac{p_i q_i}{U_i}$ *for all* $i \in E$.

($C_U(p)$, or more simply $C(p)$, denotes the *unit cost* function associated with U : $C(p) = Min\{pq, q \in R^n_+ \text{ and } U(q) \ge 1\}$).

Proof. (*i*) Let $i, j \in E$. According to Lemma 1-(*i*), there exist $m, r, \dots, k$ such that

$$A_{ji} = \frac{p_j q_m}{p_j q_j} \frac{p_m q_r}{p_m q_m} \ . \ \dots \ . \frac{p_l q_k}{p_l q_l} \frac{p_k q_i}{p_k q_k}$$

We get :

$$\frac{U_i}{U_j} = \frac{U_m}{U_j} \frac{U_r}{U_m} \ . \ \dots \ . \ \frac{U_k}{U_l} \frac{U_i}{U_k} \le$$

$$\frac{p_j q_m}{p_j q_j} \frac{p_m q_r}{p_m q_m} \ . \ \dots \ . \frac{p_l q_k}{p_l q_l} \frac{p_k q_i}{p_k q_k} = A_{ji}$$

for all $i, j \in E$.

(*ii*) We need only show that U rationalizes the data : the rest is self-evident. If U does not rationalize the data, then there exist $j \in E$ and $q \in R_+^n$ such that $U(q) > U(q_j)$ and $p_j q \leq p_j q_j$. According to the first inequality, and by construction of U, there exists i such that :

$$U_i \frac{p_i q_j}{p_i q_i} < U_k \frac{p_k q}{p_k q_k} \quad \text{for all } k \in E$$

whereas $U_j \leq U_i \dfrac{p_i q_j}{p_i q_i}$ by hypothesis for $U_1, \ldots, U_T$. Hence $U_j < U_j \dfrac{p_j q}{p_j q_j}$ and therefore

$\dfrac{p_j q}{p_j q_j} > 1$, which contradicts the second inequality : $p_j q \leq p_j q_j$.

(*iii*) By hypothesis, we have

$$U_i \leq U_j \frac{p_j q_i}{p_j q_j} \quad \textit{for all } i, j \in E.$$

with an equality if $j = i$. Therefore :

$$U_i = \quad Min\left\{U_j \frac{p_j q_i}{p_j q_j} \quad , \; j \in E\right\} = U(q_i) \quad \text{for all } i \in E.$$

In addition :

$$\begin{aligned} p_i q_i &= C(U(q_i), p_i) && \text{for } U \text{ rationalizes the data} \\ &= C(U_i, p_i) && \\ &= U_i C(p_i) && \text{for } U \text{ is homogeneous} \end{aligned}$$

so that $C(p_i) = \dfrac{p_i q_i}{U_i}$. ♦

Lemma 4. (Varian, 1983) *If U is a homothetic, non-satiated utility function that rationalizes the data, then the HARP condition is met.*

We can now supply the proofs of Theorems 1 and 2, and Propositions 1 and 2.

Proof of Theorem 1. The HARP condition is sufficient according to Lemmas 2 and 3, with the observation that $A_{kl} \leq \dfrac{p_k q_l}{p_k q_k}$. The condition is necessary according to Lemma 4. ♦

Proof of Theorem 2. (*a*) Let U be a homothetic, data-rationalizing utility function and C the associated unit-cost function. We have :

$$P_{i/j} \leq \frac{C(p_i)}{C(p_j)} \leq L_{i/j} \qquad \text{for all } i, j \in E$$

i.e.,

$$\frac{p_i q_i}{p_j q_i} \leq \frac{C(p_i)}{C(p_j)} \leq \frac{p_i q_j}{p_j q_j} \qquad \text{for all } i, j \in E.$$

Let

$$U_i = \frac{p_i q_i}{C(p_i)} \quad \text{for all } i, j \in E. \qquad (*)$$

Thus

$$\frac{p_j q_j}{p_j q_i} \leq \frac{Uj}{U_i} \leq \frac{p_i q_j}{p_i q_i} \quad \text{for all } i, j \in E.$$

Therefore, according to Lemma 3-(*i*)

$$1/A_{ji} \leq \frac{Uj}{U_i} \leq A_{ij} \qquad \text{for all } i, j \in E.$$

Hence, with (*) :

$$1/\Delta_{ji} \leq \frac{C(p_i)}{C(p_j)} \leq \Delta_{ij} \qquad \text{for all } i, j \in E.$$

Thus, $\text{COL}_{i/j}(U)$ lies in the interval $[1/\Delta_{ji}\,;\Delta_{ij}]$ for all $i, j \in E$.

(*b*) Reciprocally, let $i, j \in E$. Let us consider the series $U_1, \dots, U_T$ of Lemma 2. Let U be the utility function defined in Lemma 3-(*ii*) and C the associated unit-cost function. We have :

$$\text{COL}_{i/j}(U) = \frac{C(p_i)}{C(p_j)} = \frac{p_i q_i}{p_j q_j}\frac{Uj}{U_i} = \frac{p_i q_i}{p_j q_j}A_{ij} = \Delta_{ij}$$

Likewise, we construct another utility function such that :

$$\frac{C(p_j)}{C(p_i)} = \Delta_{ji} \quad \text{i.e., } \text{COL}_{i/j}(U) = 1/\Delta_{ji}.$$

Thus, the bounds of the interval $[1/\Delta_{ji}\,;\Delta_{ij}]$ are $\text{COL}_{i/j}$ indexes.

We must now prove that *all* the points of the interval $[1/\Delta_{ji}\,;\Delta_{ij}]$ are $\text{COL}_{i/j}$ indexes. This result is the object of Proposition 1. ♦

Proof of Propositions 1 and 2.

Let

$$K = \{C_1, \dots, C_T > 0\,, \quad 1/\Delta_{ji} \le \frac{C_i}{C_j} \le \Delta_{ij} \text{ for all } i, j \in E\}$$

$$= \{C_1, \dots, C_T > 0\,, \quad C_i \le \Delta_{ij}\, C_j \text{ for all } i, j \in E\}.$$

This set is convex. Let us take, for all $i, j \in E$, the numeric function $\phi_{ij} : R^n_{++} \to R$ defined by $\phi_{ij}(C_1, \dots, C_T) = \dfrac{C_i}{C_j}$. Since this function is continuous and K connected, according to the theorem of intermediate values, the set $\phi_{ij}(K)$ is an interval. Under the definition of K, this interval is included in the interval $[1/\Delta_{ji}\,;\Delta_{ij}]$. According to point (*b*) of the demonstration of Theorem 2, the interval $\phi_{ij}(K)$ reaches the bounds of the interval $[1/\Delta_{ji}\,;\Delta_{ij}]$ (the C_l values take the form $C(p_l)$ for a utility function chosen as in (*b*)). Thus, $\phi_{ij}(K) = [1/\Delta_{ji}\,;\Delta_{ij}]$ for all $i, j \in E$.

Let us now fix $i, j \in E$ and take $x \in [1/\Delta_{ji}\,;\, \Delta_{ij}]$. There exist $(C_1, \ldots, C_T) \in K$ such that $x = \dfrac{C_i}{C_j}$. Positing $U_k = \dfrac{p_k q_k}{C_k}$, we have

$$\frac{U_l}{U_k} = \frac{p_l q_l}{p_k q_k}\frac{C_k}{C_l} \leq \frac{p_l q_l}{p_k q_k}\Delta_{kl} = A_{kl} \leq \frac{p_k q_l}{p_k q_k} \qquad \text{for all } k,\, l \in E.$$

Therefore $U_1, \ldots, U_T$ satisfy the conditions of Lemma 3. Let U be the associated utility function as in Lemma 3-(*ii*) and C the associated unit-cost function. The U function has the form mentioned in proposition 2 with $\alpha_t = \dfrac{U_t}{p_t q_t}$; moreover, it rationalizes the data. According to Lemma 3-(*iii*), we have $C(p_k) = C_k$ for all $k \in E$, so that $x = \dfrac{C(p_i)}{C(p_j)}$. In other words, $x = \mathrm{COL}_{i/j}(U)$. Thus, all the elements of the interval $[1/\Delta_{ji}\,;\, \Delta_{ij}]$ are $\mathrm{COL}_{i/j}$ indexes. ♦

7. Conclusion

The most important finding—which is far from immediately obvious—is no doubt the observation that the prices recorded and the quantities traded in the market can be rationalized by a *homothetic* utility function, all the more easily as the data are strongly disaggregated. The cost-of-living index is thus not only a theoretical *concept*. Its existence is demonstrated for consumer staples such as coffee, edible oils, and detergents. Admittedly, the calculation of COLs requires the use of scanned data, but it is reasonable to assume that such data will become more affordable in the relatively near future.

Our study provides a gauge of the quality of standard price indexes, measured in terms of their closeness to the COLs. It confirms the reputation of the Fisher index as a very good proxy for the COL, since it lies almost always inside the COL bounds. If we consider a Fisher index value inside the bounds, there exists a data-rationalizing homothetic utility function for which the Fisher index serves as a COL (Proposition 1); there also exists a quadratic utility function that defines the Fisher index (§2.2), but there is no evidence that the utility functions coincide, or even that a quadratic utility function rationalizes the data.

More generally, the determination of the COL bounds comprises the following limitation : we cannot say whether two successive COL lying within these bounds are related or not to the same utility function. Consequently, we cannot measure a price movement with precision, but merely the bounds of the movement. The precision of these bounds will obviously depend on the width of the COL interval. However, that width is of the same order of magnitude as the 95% confidence interval for the precision of the CPI groupings (for example, for the "edible oils" grouping of the French CPI, the width of the confidence interval is 2 % for the December 1998 change on one year earlier).

Unlike the Fisher index, the geometric mean frequently oversteps the COL interval (except for coffee), and is thus an inferior proxy of the COL. Like the Laspeyres index, but less so, it overweights the products whose prices rise most steeply. Apparently, therefore, the Cobb-Douglas function does not always take into adequate account the substitutions between products or sales outlets. In other words, the geometric mean—adopted by most countries for all or part of their item indexes—offers one major advantage : it avoids the drift of the arithmetic-mean formula when chaining is practiced.

Regrettably, there is no unique COL at a given date—here, a given *month* of the period studied. This is due to the *discreteness* of price and quantity observations. The scanner data, however, are collected weekly, which ought to provide a good proxy for continuous price and quantity monitoring. A classic theoretical result is worth recalling here : with continuously observed data, and under certain assumptions, the COL is unique and coincides with the Divisia index. However, by increasing the frequency of quotations while remaining at a detailed product-definition level, it is hard to demonstrate the convergence of the chained indexes toward that unique index, probably because such an approach strongly increases the non-observability of the varieties (Magnien and Pougnard 1998).

References

S.N. Afriat (1967), "The Construction of a Utility Function from Expenditure Data," *International Economic Review*, vol. 8, pp. 125-33.

S.N. Afriat (1977), *The Price Index*, Cambridge (U.K.) : Cambridge University Press.

S.N. Afriat (1981), "On the Constructibility of Consistent Price Indices Between Several Periods Simultaneously," in A. Deaton (ed), *Essays in Applied Demand Analysis* (Cambridge : Cambridge University Press).

M. Boskin, E. Dulberguer, Z. Griliches, R. Gordon, and D. Jorgensen (1996), *Toward a More Accurate Measure of the Cost of Living*, Final Report to the U.S. Senate Finance Committee, December.

W.E. Diewert (1973), "Afriat and the Revealed Preference Theory," *Review of Economic Studies*, vol. 40, pp. 419-26.

W.E. Diewert (1981), "The Economic Theory of Index Numbers : a Survey," in A. Deaton (ed), *Essays in the Theory and Measurement of Consumer Behaviour* (Cambridge : Cambridge University Press), pp. 163-208.

W.E. Diewert (1990), "The Theory of Cost-of-Living Index and the Measurement of Welfare Change," in W.E. Diewert (ed), *Price Level Measurement*.

W.E. Diewert and C. Parkan (1978), *Test for Consistency of Data and Nonparametric Index Numbers*, Working Paper 78-27, University of British Columbia (Canada).

W.E. Diewert and C. Parkan (1985), "Test for Consistency of Data," *Journal of Econometrics*, vol. 30, pp. 127-47.

F. Lequiller (1997), "L'indice des prix à la consommation surestime-t-il l'inflation?" *Economie et Statistiques*, no. 303 (available in English on the INSEE Web site as "Does the consumer price index overestimate inflation?" *INSEE Studies*, www.insee.fr)

F. Lequiller (1998), "Biais des IPC : où en est-on?" *INSEE Méthodes*, no. 84-85-86, pp. 291-312.

F. Magnien and J. Pougnard (1998), "Etude du chaînage d'indices de prix à l'aide de micro-données," *INSEE Méthodes*, no. 84-85-86, pp. 247-82.

M.E. Manser and R.J. McDonald (1988), "An analysis of substitution biais in measuring inflation, 1959-85," *Econometrica*, vol. 56, no. 4, pp. 909-30.

B. Moulton (1996), *Constant Elasticity Cost-of-Living Index in Share-Relative Form*, Bureau of Labor Statistics (BLS), mimeo.

H.L. Varian (1982), "The nonparametric approach to demand analysis," *Econometrica*, vol. 50, no. 4.

H.L. Varian (1983), "The nonparametric tests of consumer behaviour," *Review of Economic Studies*, vol. 50, pp. 99-110.

S. Warshall (1962), "A Theorem on Boolean Matrices," *Journal of the American Association for Computing Machinery*, vol. 9, pp. 11-12.