Does the French Consumer Price Index Overstate Inflation?

François LEQUILLER *

Août 1997

The author wishes to thank all those who have taken the time to improve this article. He is especially grateful to the unnamed referee chosen by the editor of Économie et Statistique (where the original French version is being published under # 303, 1997-3), but also to his successor, T. Lacroix, and to L. Viglino, D. Temam, M. Glaude, and G. Laroque. Special thanks to J. Mandelbaum for a magnificent translation.
This paper will soon be available in the Insee Studies series on Insee's Web site (www.insee.fr).

At the time of writing, François Lequiller was deputy head of INSEE’s Department of Consumer Prices, Income, and Living Conditions, which is in charge of the French consumer price index.

Ces documents de travail ne reflètent pas la position de l’INSEE et n’engagent que leurs auteurs.
Working papers do not reflect the position of INSEE but only their author’s views.
I. Introduction

Spurred by several years' debate over a possible overestimation of U.S. inflation, the U.S. Senate appointed a commission of economists chaired by Professor Michael J. Boskin of Stanford University to examine the issue. The commission submitted its final report in December 1996, in which it argued that the U.S. consumer price index (CPI) would be overestimated by 1.1 percentage points per year after 1996 and had been overestimated by 1.3 percentage points a year before 1996. In other words, the Boskin Commission claimed that "true" inflation at the household-consumption stage would be 1.1 percentage points lower than the figure published by the Bureau of Labor Statistics (BLS) in the years to come. For example, if the U.S. CPI rises 3.0% in 1998, the true rise, according to the Commission, should be reckoned at 1.9%.

As the upward biases would add up over the years, the combined total would automatically reach 11.5 percentage points after 10 years. In the United States, the CPI is used to index welfare benefits and income-tax brackets. An upward bias on either item would swell the federal deficit, as indexed benefits would be exaggerated and tax revenues would be diminished by an over-rapid rise in the income-tax brackets. The Commission reckons that this mechanism alone would add about $1 trillion to the public debt by 2008. Amid the present raging debates over the federal deficit, the Commission did not hesitate to describe the CPI bias as the "fourth largest federal program, after social security, health care and defense" (p. ii).

Although the subject has long been discussed in Europe and the U.S., the controversy over the Boskin Commission findings offers an opportunity to review the French situation on this issue.

The French CPI is unquestionably one of the most closely watched of INSEE's economic indicators. It serves as the main gauge of inflationary pressures in monetary and fiscal policy, and, as such, is directly involved in the assessment of one of the convergence criteria defined by the Maastricht Treaty. The detailed price indexes that compose it are used as household-consumption "deflators" in the national accounts. As household consumption represents 60% of GDP, the CPI is crucial to the determination of the economic growth rate. In France, the CPI is used less directly and systematically than in the U.S. for determining social benefits and tax brackets. However, it is the direct indexing instrument for the minimum wage (called SMIC), pensions, and family benefits. It is also the indirect instrument for indexing wages, tax brackets, selected social benefits, and many types of private contracts.

---

1 This extrapolation assumes a constant bias over time—an assumption that has not been fully explored. It also supposes that BLS will not change its methods in the future.
2 Appendix 1 tabulates the results of other studies similar to the Boskin Commission's, for the U.S. and other countries.
3 INSEE, the French National Institute of Statistics and Economic Studies, is the French counterpart to BLS for the calculation of the Consumer Price Index.
4 The official procedures for automatic CPI indexing apply to the minimum wage, retirement and disability pensions, and family benefits. The law no. 93-936 of July 22, 1993, provided that, for a five-year period from January 1, 1994, retirement and disability pensions would first be indexed on the forecast change in consumer prices, then adjusted to the actual change. The law no. 94-629 of July 27, 1994, defined the same mechanism for indexing family benefits, with effect through December 31, 1999. In the sphere of private contracts, alimony settlements are among the most commonly indexed to the CPI. It should also be noted that indexing in France is regulated by the Neiertz Law, which bans linkages to indexes that include tobacco. Note: this article does not discuss the actual principle of indexing, which is the subject of debate both in France and the U.S. (see, especially, Griliches 1995).
For all these reasons, the determination of the CPI—in France and the U.S.—rests on the strict application of standardized procedures, and INSEE, like BLS, devotes substantial resources to the task (box 1). The harmonization of European price indexes, in which France is heavily involved, has confirmed that the basic principles of French CPI determination meet the highest international standards.

---

5 The CPI’s suitability to its various applications is discussed by Glaude 1997.
II. Substitution effects and new products

As may easily be imagined, however, it is no easy task to determine an indicator that is supposed to summarize in a single figure the changes in prices of the billions of transactions involved in household purchases. The process raises many problems, such as the coverage, representativeness, and accuracy of the index (sampling), and the recording of actual transaction prices (followed by that of unadvertised discounts: see box 2). This article addresses two very specific issues: (1) the coverage of households' purchasing behavior (substitution effects between existing products), and (2) the introduction of new products (and/or quality changes).

Both issues have long been at the heart of the CPI debate. Many economists, including the members of the Boskin Commission, stress that the methods for calculating price indexes are based on the hypothesis of fixed consumption structures and on equilibrium assumptions. As a result, it is claimed, the methods are not good at describing an ever-changing economy made of imbalances, new products, and new agents. In particular, as we shall see in §III, the balanced-market hypothesis plays a central role in the treatment of new products in the price index. This hypothesis, it is argued, does not allow for possible temporary imbalances, such as a "price war" triggered by the introduction of new products, or the emergence of new, more efficient producers.

The novelty in the Boskin Commission approach is that, for the first time, an official commission has ventured to produce a quantitative estimate of the alleged upward bias. Like our BLS colleagues, we believe the estimate is questionable. We shall see how statisticians tackle these issues.

Section II examines issues relating to product substitution. We look at the weightings and mathematical formulas used to offset the substitution bias. Our conclusion is that the statistical methods used in France differ from those of the U.S., and that the French methods largely shield the French CPI from the kind of criticism voiced by the Boskin Commission. Section III addresses the issue of new products. We describe the statistical methods used, without hiding their limitations or the need for improvements. We show, however, that errors can occur in both directions and that no convincing studies have yet been produced that make it possible to identify and quantify an upward bias—although that is precisely what the Boskin Commission did. In Section IV, we specifically discuss the issue of outlet substitution. In Section V, we examine the quantification of the potential upward bias in the French CPI compared with the U.S. index. Although we refuse to produce a figure for France to be set alongside the Boskin Commission's 1.1 percentage points for the U.S., our conclusion is that, if an overestimation exists in France, it is probably far weaker.

---

6 This adjective is not intended to be polemical. We must stress that the Boskin Commission managed to maintain a high standard of discussion in its report, notably avoiding all polemic. We should also salute the contribution of BLS (which has been extremely forthcoming despite the difficult context) to improving public knowledge of CPI statistical methods. Indeed, many of the Boskin Commission calculations rely on BLS studies. Consequently, when we refer to "Boskin Commission figures" for the sake of brevity, the reader should understand that those figures are often based on studies by BLS itself, even if the context is, at times, a questionable extrapolation. One of the main controversies over the Commission findings centers on the difficulty of avoiding double counting when assessing the various biases.
The French consumer price index is a complex statistical tool based on the strict application of standardized procedures. More than 160,000 price quotations are collected each month in about a hundred urban areas representative of towns and cities of more than 2,000 inhabitants. All types of outlets are covered, including hypermarkets, supermarkets, hard-discounters, superettes, specialized retailers, “mom and pop stores,” street markets, etc. Thirty thousand outlets are visited each month by the CPI’s 170 full-time or part-time price collectors. One hundred and thirty other INSEE personnel, including several dozen statisticians, are responsible for computing the index. The Institute’s price collectors are specially trained for their assignment. Quality-control procedures are applied at every step of the process. The initial results for the month—known as the “provisional” index—are published toward the 10th of the following month, i.e., sometimes within seven working days of the close of price collection. The index covers the entire range of goods and services included in the definition of household consumption: food, beverages, tobacco; clothing and footwear; housing, heating, and electricity; furnishings, household equipment, and routine maintenance of the house; health; transportation and communications; recreation, entertainment, culture, and education; personal care, restaurants, hotels, and other services to households. The CPI thus covers about 90% of households’ market consumption. Insurance is scheduled for introduction in 1998, along with prices of selected domestic and legal services, fees for government services, and second-hand motor vehicles. As regards non-market services, INSEE is examining ways to include hospitals, clinics, and welfare expenditures.

The weightings of the index groupings and their component products are reviewed each year. The index uses an annually chained Laspeyres formula. Every year, the statisticians in charge of the broad index sectors (food, durable goods, clothing, services, etc.), assisted by field workers, conduct a thorough examination of the characteristics of the 1,000 detailed families of products (variétés or sub-items) that make up the index. These families are modified where necessary to preserve CPI representativeness. One example of such changes concerns the meat sub-items whose consumption has fallen as a result of the mad-cow crisis. New sub-items are introduced each year as additions or substitutes for existing sub-items that have become less representative. A recent example was the introduction of computer floppy disks. The current index is known as the 1990-base index (1990 = 100). This base is France’s sixth-generation CPI since the start of the century. Methods and coverage have improved with each generation. Estimations of sampling errors on the latest generation of the French CPI show that the twelve-month change fluctuates in a very narrow range of +0.1 to -0.1 percentage points (Ardilly and Guglielmetti 1993).
Box 2

Does the CPI take account of all price discounts?

Criticism of the French CPI, especially in recent years, has often focused on the following issue: the CPI, it is argued, does not take into account the "unadvertised" and "private" discounts agreed upon between buyers and sellers.\(^7\)

To clarify the discussion, it should be pointed out that the French CPI is now tracking all advertised discounts, special offers, and promotions on all products. INSEE price collectors are instructed to record prices net of all such discounts.\(^8\)

It is true, however, that INSEE price collectors can only monitor advertised discounts and not actual transactions if the listed prices are changed through direct bargaining between buyers and sellers. INSEE can hardly require its field workers to simulate a purchase every month, as this would violate basic ethical principles. In all likelihood, though, bargaining over advertised prices became more common in France in 1993 owing to the consumption slowdown. The French CPI may therefore have overestimated inflation during a specific period by failing to take into account the greater incidence of unadvertised discounts.\(^9\)

But if we accept that inflation may have been overstated during consumption slumps, we should also admit the converse assumption, namely, that inflation may have been understated during consumption recoveries, when unadvertised discounts tend to become less common. In other words, this « discount bias » is not a structural bias, like those described in this article, but a cyclical bias.

Estimating the discount bias is a particularly delicate task. One major obstacle is the high cost of tracking the phenomenon. This would require access to retailers' bills—which are regarded by retailers as highly sensitive data. However, using the CPI weighting structure and some conventional assumptions, we can show that the impact of the bias on the overall index is probably rather limited. First, the types of product to which discounts are most commonly applied—automobiles, clothing and footwear, household appliances, and furnishings—account for only 14% of the total index. Consequently, if we make the extreme assumption that 10% of all purchasers of these products are offered an additional unadvertised discount of 2% on the previous year's price, this would have an impact of 0.10 x 0.02 x 0.14 = 0.03 points on the annual average of the overall index. This upward bias would be offset by a negative bias if the suppliers' position improved thanks to a consumption rebound.

---

7 Discount coupons, which have become very popular in recent years, are another variant of this practice.
8 Except for discounts applied for a few hours only ("surprise" special offers announced in stores), clearance sales, and consumer-credit rebates. Our position on clearance-sale prices is that—by definition—the old product cannot be replaced when its price can no longer be observed. As for consumer-credit rebates, CPI coverage does not extend to interest-rate movements.
9 In this connection, it should be remembered that the factor that would influence the price-index change is the variation in unadvertised discounts and not their absolute level. If their level did not vary, the price-index change would be identical regardless of whether or not the unadvertised discounts were included.
II.1. Product substitution: the issue of weightings

The problem raised by product substitution for CPI statisticians is that households change their consumption basket at the same time as prices vary. As we shall see, a CPI whose weightings rested on obsolete information may tend to overstate inflation.

Let us begin by examining the case of a single consumer and try to calculate his or her price index between a base period and the current period. We define the consumer's price index between the two periods as the growth rate of his or her budget outlays that enables the consumer to maintain the same utility level at current prices in the current period as in the base period. This is the concept known as the preservation of "purchasing power."10

If all product prices varied proportionally, such a price index would be very easy to calculate: one would simply pick one product and measure its price change. However, we know that the average rise (or fall) in prices masks different variations from one product to another. In other words, relative prices change at the same time as the general upward or downward movement in prices. To calculate our consumer's price index, we therefore need to track all the products he or she consumes, or at least a representative sample of those products; we then have to determine the average of the changes in prices for the products in the sample. This immediately raises one question: what weighting should be assigned to each of the products included in the average? The only possible weighting, of course, is based on the quantity consumed. The price index is therefore the result of an average of price changes weighted by values of quantities consumed. This definition is inadequate, though, as there are many ways to compute the average. In particular, prices are not the only variables to have changed between the base period and the current period: the quantities of products consumed have changed as well. Since that is so, what quantities should we use to establish the price-index weightings? The quantities consumed in the base period? Those of the current period? Or an average of the quantities consumed in the two periods? Index theory (appendix II, §1) does not offer a single, final answer—not even for a single consumer and, still less so, for multiple consumers. The theory does tell us, however, that—under certain hypotheses—one of the best proxies of an ideal index is a Fisher index: this is an average of an index using base-period weightings (called a Laspeyres index) and an index using current-period weightings (known as a Paasche index). The theory also shows that the Laspeyres index usually overestimates the Fisher index, while the Paasche index usually underestimates it. The mechanism behind this result is simple: the Laspeyres index overweightss the products whose prices register the steepest increases, whereas, by logic, the share of these products in consumer expenditures will decline once consumers accept a degree of substitution between products while maintaining constant utility.

In practice, however, a Fisher index is impossible to calculate, at least during the year and within the short deadlines required for an index such as the CPI. The reason is that the Fisher index cannot be established without the current-period weightings, which take a long time to determine. To calculate the Fisher index for 1997 against a 1990 base, for example, we would need to know, in particular, the quantities consumed annually in 1997. Naturally, those quantities cannot be determined during 1997. That is why all countries calculate the CPI as a Laspeyres index, that is, using fixed weightings derived from the base period. The "age" of the

---

10 For some economists, the term "purchasing power" implies, by construction, that the price index uses the quantities consumed in the base period. Here, however, we use the term in a broader context.
base year used to estimate the weightings varies from one country to another. In France, the weightings used to obtain the overall index from "grouping" indexes are updated every year from recent data. In the U.S., the weighting base is much older.

II.2. A three-level decomposition

The situation is not quite so simple, however, since a price index as complex as the CPI is the result of successive aggregations of indexes; each aggregation level has its independent weightings, with different "ages." In France, for example, statisticians work at a fairly high level of aggregation for the annual weighting updates based on data from year n-2. The other, more detailed levels are not treated identically. In this section, we describe the three stages of the aggregation process used to obtain the overall index from the individual price quotations. The first stage, hereafter referred to as the "lower level" (niveau détaillé), is the calculation stage that yields the highly detailed indexes (often called "micro-indexes") for highly disaggregated product categories (variétés or "sub-items") for a specific geographic region (agglomération or "urban area"). The micro-indexes are obtained from price quotations collected from sales outlets in the urban areas examined. The second stage, hereafter designated as the "intermediary level" (niveau intermédiaire), yields the indexes for product groupings from the micro-indexes. In third and final stage, called the "upper level" (niveau agrégé), the overall price index is calculated from the grouping indexes. In each stage, a Laspeyres formula is used (or was used, as we shall see). Upper-level weightings are taken from the national accounts. Intermediate-level weightings are usually taken from public or private surveys of household expenditures, or from other sources such as private panel surveys of distributors, production and import statistics, etc. At the lower level, the lack of information has led to the convention of assigning an identical, fixed weight to each product. At each of these stages, an overestimation—described hereafter as "substitution bias"—may occur. We shall therefore examine each level separately to assess the possible size of that bias.

II.3. No overestimation at the upper level in the French index

The methods used to calculate the French index protect it from upward bias at the upper level. Since the early 1970s, the weightings of CPI "groupings" are reviewed annually against national-accounting data from year n-2: in other words, the 1997 monthly price index is based on 1995 weightings. The "instantaneous" French index—that is, the monthly changes in the current year—is therefore based on very recent upper-level weightings. The "long-term"
French index, which tracks price changes over several years, is based on the chaining of these instantaneous indexes. The French index is thus referred to as a "chained" Laspeyres index.\textsuperscript{17}

In sum, the age of the French CPI grouping weightings never exceeds two years. We can therefore conclude that the upper-level substitution bias is, in practice, negligible or non-existent in France.\textsuperscript{18} In the U.S., by comparison, the Boskin Commission estimates the upper-level substitution bias at 0.15 percentage points per year based on the present weighting structure, which dates from 1982-84—that is, well over a decade ago. The Commission relied on simulations conducted in the U.S.: these showed that, in present conditions, a Laspeyres index whose weightings are approximately ten years old increases by about 0.1-0.3 percentage points a year more than the same index calculated using a Tornqvist formula, a close equivalent of the Fisher formula (Aizcorbe and Jackman 1993). In France, we compared the official index with an unchained index for the period 1980-90. As expected, the unchained index was found to rise \textit{on average} by 0.11 percentage points more than the official index (Viglino and Montiel 1995).\textsuperscript{19}

\textsuperscript{17} In Europe, two countries use a nearly identical system: Britain and Sweden. Germany updates its weightings every five years. The other European countries update somewhere in between these two frequencies.

\textsuperscript{18} The chaining method is not necessarily a universal cure. In particular, it can lead to positive biases in case of sharp swings in relative prices, as shown in appendix II, §2. However, most of these swings do not occur at a level as aggregated as the product groupings. Indeed, all the numerical simulations for the 1980-95 period show that chaining makes the Laspeyres index more comparable to a Fisher index than before.

\textsuperscript{19} The figure of 0.11 percentage points is an average. The difference is not necessarily identical in all years and many even be negative in some years. Another illustration of this type of gap, but in the other direction, is offered by the consistent difference between the CPI and the Paasche indexes of household-consumption implicit deflators in the French national accounts (box 5).
II.4. A mild residual upward bias at the intermediate level

By contrast with the upper level, the weightings of the French intermediate-level index—i.e., the weightings of the "sub-items" in the "groupings"—are not updated each year. Of course, as far as the CPI budget allows, INSEE gathers and uses all the detailed information published in wholesale/retail trade journals or obtained from trade associations and market-research firms. Admittedly, there is room for improvement at this level. For example, a more routine use of data from market-research firms would be desirable. The French index may thus be subject to an upward bias due to the obsolescence of some intermediate weightings. Unfortunately, there is no way to measure that bias directly. Some indications, however, can be used to provide a very rough range. If a substitution bias exists at the intermediate level, it probably exceeds the upper-level bias, which, as we saw, would be equal to 0.11 percentage points per year in France if we did not calculate a chained Laspeyres index. This assumption is based on the notion that, as the level of index detail increases, so does the substitutability between the index components. On the other hand, many sub-item weightings are regularly updated, even if not annually, and this would tend to reduce the potential upward bias. On balance, a mild upward bias in the range of 0.05-0.10 percentage points a year cannot be ruled out.

Box 3

Harmonizing Europe's CPIs

The CPI has been chosen as the main indicator for the inflation criterion of the Maastricht Treaty. Member States applying for monetary union must show an inflation rate no more than 1.5 points above the average of the three best-performing countries. The criterion’s precision calls for a harmonization of methods to ensure maximum comparability between country indicators. This harmonization drive has made great progress since its launch three years ago. A framework regulation (#2494/95) was adopted by the Council in October 1995. A first major step was the publication of an initial version of the Harmonized Index of Consumer Prices (HICP) in early 1997. Harmonization has both increased the comparability of national indexes and improved each individually. Far from confining themselves to a "common methodological denominator," the Fifteen have sought to incorporate the best method for each field into the implementing regulations, which cover the following areas: (1) geometric mean and other micro-indexes, (2) transition to monthly price collections and improvement of the treatment of missing values, (3) acceleration and coordination of new-product introductions, (4) improvement of weighting updates, (5) improvement of the treatment of quality changes, and (6) harmonization of index coverage. In compliance with (6), France has incorporated insurance prices into its HICP in 1997 (but not yet in its national CPI). Eurostat will inspect the HICPs of each country to ensure compliance with the rules jointly approved by all Member States. Beyond its strictly regulatory aspects, harmonization has fostered common research programs, particularly on quality change.

20 See Buchwald and Saglio (1994) for an example of the non-comparability of the French and German non-harmonized CPIs.
II.5. The geometric mean eliminates the risk of upward bias at the lower level in France

For its lower-level indexes, France previously used a method based on implicitly fixed weightings. As we shall see, the current method is based on a formula that allows for a degree of substitutability and therefore removes the risk of upward bias. For the past three years European statisticians have focused—among other methodological issues (box 3) —on the formulas used by EU member States to calculate "micro-indexes." It emerged that the use of geometric means as micro-indexes had two advantages over the conventional arithmetic-mean formulas (appendix II, §2): (1) the geometric mean allows for possible substitutions between products, whereas the arithmetic mean assumes fixed weightings; (2) the geometric mean, by construction, avoids what has been called "formula bias." Formula bias is a positive bias that occurs when an arithmetic mean is used without precautions in a chaining operation. The 15 EU member States therefore agreed unanimously to extend the use of the geometric mean. This decision—now embodied in a European regulation—will not only improve the calculation methods in each country, but make it much legitimate to compare the price indexes of the 15.

France is therefore gradually substituting geometric means for those of its micro-index formulas that relied on arithmetic means of price ratios. The geometric mean is well-suited to French micro-indexes, which cover highly specific categories of products exhibiting similar technical characteristics and whose prices are collected from the purchasing area known as "urban area" (agglomération). The geometric mean would not necessarily work well with larger product families, for which the substitution elasticities might be lower. Simulations performed on recent years showed that—all other things being equal—the switch would lower the index by an estimated average of about 0.10 percentage points per year for the period examined. The switch will thus gradually eliminate the bulk of lower-level substitution bias and formula bias in France.

---

21 See appendix II, §2. The substitution will be applied to about 12,500 out of 25,000 micro-indexes. It began in January 1997 for one-half of the micro-indexes concerned and will be complete by January 1999. To preserve the time-comparability of data series, France preferred a gradual introduction—which nevertheless complies with the European regulation—to a massive, instantaneous switch. It should also be noted that the other 12,500 micro-indexes are based on a price-sums ratio (PSR) formula. Such micro-indexes have been used for years in the CPI for so-called "homogeneous" sub-items and are accepted under the European regulation. The results of a PSR simulation show that, although it can lead to divergences from the geometric mean for individual sub-items, the calculations on several dozen sub-items converge toward the same figures as the geometric mean. PSRs are not vulnerable to formula bias.

22 When the geometric mean was introduced in France, French statisticians analyzed the impact on individual sub-items. One of the main consequences of the analysis was a redefinition of some sub-items to achieve greater precision and make the products more homogeneous.
II.6. A wider difference between geometric mean and arithmetic mean in the United States

The size of the correction procured by the introduction of the geometric mean in France (0.10 percentage points) is well below the Boskin Commission’s figure for the U.S. Yet the latter was obtained with the same type of calculation—i.e., comparing (1) an index prepared from micro-indexes consisting of geometric means with (2) an index based on arithmetic means. For the U.S., the difference is put at 0.50 percentage points, which leads the Boskin Commission to estimate a formula bias of 0.25 percentage points and a lower-level substitution bias of 0.25 percentage points as well. The difference between the U.S. and French results can be explained in two ways.

(1) The U.S. formula bias was, by construction, much greater than the French one. The reason is that 20% of the component products in 70% of the entry-level items in the U.S. CPI are systematically replaced at random each year. This highly sophisticated procedure eliminates product-selection bias and provides an automatic 1/5 renewal of products tracked by the index. However, the procedure was vulnerable to a formula bias similar to the one described in appendix II, §2: whenever the initial price of the new product selected at random was low (because of special offers or seasonal factors), the subsequent price rises in the U.S. index were automatically overstated. Conversely, when the initial price was too high, the subsequent decreases were automatically understated (Moulton 1996). This error proved to be particularly large—an estimated 0.25 percentage points—for the U.S. index of fresh fruits and vegetables. In the French index, the problem affected only one-third of the CPI weighting: one notable exception was, precisely, fresh products, for which another (unchained) formula is used. Nor was the French index affected by special offers, since new products are never introduced when they are on special offer. All the factors that accentuated the bias due to the use of arithmetic means in a chained index in the U.S. were therefore absent in France. In 1996, BLS undertook a specific correction to eliminate the bias with effect from 1996.

(2) The method used by the Boskin Commission to estimate the lower-level substitution bias effectively consisted, as in France, in comparing the official index—all other things being equal—with the same index in which arithmetic-mean formulas were replaced by geometric-mean formulas. Unlike in France, however, the replacement was made not only at the lower level but also at the intermediate level. The use of geometric means at the intermediate level is not always advisable, as BLS itself has pointed out (Moulton 1996). Indeed, the substitution elasticity at the intermediate level may, in some cases, be closer to 0 than to 1. Hence BLS caution about implementing the Boskin Commission's overly systematic recommendation of the geometric mean (Abraham 1997).

23 We use the past tense because—in the U.S. as in France—the formula bias is a thing of the past, since BLS started correcting it in 1996. Rather than introducing a geometric mean, as in France, the Bureau corrected the implicit weightings of the chained index.
24 There might be a product-selection bias if, for example, all price collectors chose the same brand of milk chocolate to represent the "milk chocolate" entry-level item. The U.S. procedure avoids this problem thanks to its probabilistic product-selection method.
25 For the sake of completeness, however, we should note the recent discovery that the problem did exist in the French index at an intermediate level of aggregation (Poinat 1996). The resulting bias proved to be a negligible 0.01 percentage points or so per year.
26 BLS did not, however, revise the index series. The 0.25-point formula bias, therefore, still applies to the years prior to 1996. This explains the difference between the Boskin Commission's 1.3-point bias for the years prior to 1996 and its 1.1-point bias for the years after 1996.
III. New products

One of the main difficulties in constructing price indexes lies in the contradiction between (1) the fixity-of-products principle, which is crucial to the comparison of prices in two different periods, and (2) economic reality, in which new products are constantly launched and obsolete products disappear. The discussion may gain in clarity from a distinction between (1) "product replacement," in which a product tracked by the index is replaced by a similar product, and (2) new products that have few if any equivalents among past products.

III.1. Statistical methods for dealing with product replacement

Product replacement has a heavy impact on the CPI. In the French index, 30% of products whose prices are tracked in conventional outlets are replaced in any given year.\footnote{When stated in terms of price quotations collected during the year, the figure is obviously smaller, since quotations are gathered twelve times a year. This banal observation puts into perspective the issue of product replacement in high-inflation situations. N.B.: The figures and percentages for product replacement given in this section are based exclusively on the price quotations recorded directly in stores by INSEE price collectors. They account for some 90% of total price quotations used in the index. The sectors excluded from these statistics are those for which price quotations are collected from a central source (automobiles, utilities, rail and air transportation, etc.).} In absolute terms, that makes more than 30,000 product replacements a year. A statistic, calculated for the U.S. index but probably applicable to France, gives an idea of the CPI's upward or downward sensitivity to these treatments, at least in periods of mild inflation. In 1995, the U.S. CPI excluding housing (and a few other minor items) was recalculated by canceling the statistical procedures used by BLS in 1995 to eliminate quality-change values. The recalculated index rose 2.5 percentage points more than the original index (Abraham 1997).

III.2. Comparing prices "at constant quality"

Whenever a product is replaced, the price of the new product must be compared with the price of the deleted one. To measure this price change, the price ratio needs to be adjusted for any difference in quality between the old and new products.\footnote{French CPI statisticians call this the "treatment of quality-change value" (traitement de l'effet-qualité).} For example, if a car model without air conditioning is replaced by the same model with air conditioning, a direct comparison between the prices of the two cars is obviously impossible. We will have to estimate the "price" of the air conditioning, for example from the price in the automaker's catalog when air conditioning was optional; we will then have to subtract that price from the price of the new model to obtain the change in price "at constant quality." In this example, the procedure seems fairly easy, since the value of the "air conditioning" option is rather simple to estimate. It is easy to see, however, that the procedure may prove much harder in other cases: the concept of "quality" is often elusive, and its quantitative estimation even more so.

Price-index statisticians readily admit the importance of this problem. The simple theory of indexes outlined earlier is not very enlightening on this point, since it assumes, by definition, that the same products exist in the base period and the current period. Here, we are dealing with precisely the opposite case. However, the theory's reference to the principle of maintaining the consumer's utility level provides a useful guideline for the statistician. When a

27 For brevity's sake, we will refer to "new products" in this section as a shorthand for the more accurate phrase "new goods and services." In fact, most "new products" probably consist of "new services."
28 For brevity's sake, we will refer to "new products" in this section as a shorthand for the more accurate phrase "new goods and services." In fact, most "new products" probably consist of "new services."
product is replaced by another, the price variation should be calculated after the two products have been "restated" in terms of equal utility. Admittedly, this concept is hard to define for a single consumer—and all the more so for millions of consumers.

Of course, it would be wrong to conclude that there are no appropriate statistical methods. Let us begin by clearing away the most obvious misunderstandings. First, the CPI never compares prices of products of objectively different quality. To return to our earlier example, every step will be taken to avoid a direct comparison between the air-conditioned model and the non-air-conditioned model. Second, the recording procedures are designed to enable price collectors to replace the deleted product with the closest possible substitute, in order to minimize the size of the quality-change value and hence the potential error in its estimation. Having defined this framework, let us now examine the four methods used to deal with quality change:

1) **Chaining** (also known as "linking" or "splicing"): This method is probably the most widely used in price indexes, because it is simple and has some basis in economic fact. It is used for 60% of product replacements observed in stores in France. That percentage would be even higher if we extended the concept of replacement to the periodic reconstruction of the sample that takes place every December and to the sectors for which price quotations are collected from a central source. In its pure form, the method assumes the old and new products are observed on the market during the same period $t$ (as we shall see, though, this is not actually the case). The price change for the old product is used to measure the change in the overall index between $t-1$ (and earlier periods) and $t$. Once the old product is deleted, the change in the price of the new product will serve for the overall index between $t$ and $t+1$ (and subsequent periods). In other words, the price changes for the new and old products are "chained"—hence the method’s name. In practice, the method implies that the price difference between the two products in $t$ is exactly equal to the quality difference. This identity somehow validates the price difference recorded in the market, which is assumed to be in equilibrium at the time.

To take an example, let us imagine that the washing machine $Y$ is replaced by a new, more efficient model, $Y'$, sold for FF500 more. The price of $Y$ has been tracked in the index until April 1997. That same month, we also collect the price of $Y'$ in the same store. As the two machines are sold at the same time in the same market, we assume that the FF500 difference measures the value consumers attach to the difference in efficiency. The following month, $Y$ has been deleted, and the price change for $Y'$ is introduced into the index formula. The inflation for $Y'$ between April and May is therefore measured by comparing the May price of $Y'$ minus FF500 with the April price of $Y$.

2) "**Direct comparison**" (or "equivalent replacement"): This method consists in finding a new product that can be regarded as "equivalent" to the old product. The entire price difference between the two products is consequently treated as a "pure" price change (i.e., the "quality" of the two products is considered identical). For the so-called homogeneous sub-items, INSEE price collectors are thus practically obliged, when the product whose price they recorded in the previous month has been deleted, to find another product whose technical

30 It is the only method used for heterogeneous sub-items.

31 An alternative description of the method may, at first sight, seem more neutral in regard to the quality-change valuation but is ultimately equivalent. We can say that the price index is built by chaining monthly links that consist solely of products existing in two consecutive months. But it would be misleading to believe that this description of the method provides a miracle solution to the problem of quality-change valuation. The method merely assumes that the entire price difference measured at the instant $t$ is due to a quality difference.
characteristics can be regarded as identical. This method is applied to 40% of replacements of store-tracked products in the French index.

(3) "Deletion": The discontinued old product is not replaced. Subsequent price movements of its product family are tracked via the "remaining" products. This method is rarely used in the French index.

(4) "Explicit quality-change valuation": This method consists in estimating the quality change value through direct methods such as option costs, estimations of the cost of quality improvement by producers, and econometric methods often referred to as "hedonic regression methods." Properly applied, these methods are the most convincing approach, in statistical terms, for measuring quality change. Unfortunately, they are very expensive: for the moment, statistical agencies—both in the U.S. and in France—cannot afford to implement them except on a very modest scale. In France, they are applied mainly to automobiles, owing to the grouping's importance in the CPI and the high quality of information available in the industry, such as catalogs listing numerous features, option prices, and so on.

None of these four methods is perfect. In any case, it would be a fallacy to believe that there are both theoretical answers to all the issues raised by new products and the practical means to apply those theoretical solutions, even assuming they did exist. It should also be remembered that the price index relies on decentralized collection, and that replacement procedures must therefore be simple and easy to apply. Like the Boskin Commission, however, it is worth examining whether—on balance—these methods consistently produce a bias in one direction or whether the errors can go either way.

III.3. Chaining

For the chaining method to work properly, even when applied correctly, we must assume a stable equilibrium for the product market. This is not always the case. In practice, the new product may capture market share from the old one—a possible sign that the instantaneous price difference understates the quality difference. In our washing-machine example, we may assume that if the old model Y is gone from the stores in the following month, it is precisely because the new machine Y' is sold at a price that has eliminated the old one from the market. The theoretical equilibrium price of Y' is therefore more than FF500 above that of Y. To value the quality change at FF500 would understate that change and therefore overstate the price change, since the price would actually have fallen rather than remained stable. This reasoning would hold not only if the price of the new product was higher than the price of the old one (as with our washing machine), but also if the new-product price was equal to or even lower than the old-product price. Let us take another example: Firm F', a more efficient (or aggressive) competitor of firm F, sells a more powerful personal computer than that of firm F at the same price, capturing market share from F in the process. The chaining method would

32 Of the 1,000 sub-items of the French CPI, about half are so-called homogeneous sub-items; the other half are known as heterogeneous sub-items. Homogeneous sub-items are composed of products whose characteristics are defined with great precision and whose price levels are therefore very similar. As a matter of principle, replacements in homogeneous sub-items are made with the direct-comparison method. Heterogeneous sub-items consist of products whose characteristics are defined less precisely. They may therefore contain more disparate items. Although collection instructions advise direct-comparison replacements in every case, the guideline is hard to enforce in practice, and there are many more chaining-based replacements than direct-comparison-based replacements in heterogeneous sub-items.
not recognize the price of the F' computer as lower than that of the F computer, despite the fact that the F' computer is more powerful than the identically-priced F computer. Again, the result would be an overstatement of inflation. That is indeed what statistical studies of the PC market show. The difference between a chaining approach and an approach taking the quality-change value into full account has been estimated at -4.4 percentage points a year for the French PC market in the late 1980s (Moreau 1991). The PC index (Q1 1988 = 100), after full adjustment for the quality-change value, came to 61.6 in Q1 1991; with the chaining method, the index came to a far higher 70.9 for the same period (see chart 1). The handful of other detailed studies on the subject, generally based on hedonic regression methods, reach the same conclusions. But they deal with high-tech goods in very competitive markets and cannot be extrapolated directly to all goods and services. Thus there are probably some cases of a downward bias in the inflation estimate.
Chart 1

The price fall steepens when PC performance is taken into account: 1988-91

*Personal-computer price index (100 = Jan. 1988)*

Continuous line: Price index based on chaining method
Dotted line: Price index incorporating new-product performance
Source: producer prices, INSEE (Moreau 1991)
III.4. The hypothesis of hidden price rises cannot be ruled out

In some markets, one cannot rule out the possibility of hidden price rises applied by the producer when changing products. In such instances, the chaining method would wrongly nullify a price increase. Even in a fairly competitive context, effective marketing—even of a consumer staple—can probably enable a new product to be sold at a price that exceeds the intrinsic quality difference with the old product. There is also the case of regulated prices, where price rises often have to be justified by a quality change. In that case, some quality changes may be voluntarily overstated, which means the chaining method would underestimate the product's price rise in the index. The case of reimbursable pharmaceuticals is a possible example in France because their prices are controlled by the government. Several economists have argued that hidden price rises occur when variants of existing pharmaceutical products come on the market. In the index, these new drugs are treated with the chaining method, which, by construction, cancels the hidden price rises and therefore underestimates inflation in this sector. It is a fact that the pharmaceutical price index grows much more slowly than the overall index, although this does not constitute a proof of the hidden-increase hypothesis, since the difference can be explained by many other factors.

III.5. Some doubts about clothing products

Similar doubts may be raised about clothing products. For a proper perspective on the issue, we need to go back to the chaining method. In theory, it assumes that the deleted product and the replacement product have been observed in the same period. In practice, however, the method is applied in a more summary manner. Typically, the prices of the two products cannot be collected simultaneously, that is, the same month. This means that, in effect, the price of the old product is proxied during the "non-overlap" period (one month, sometimes more) by the change in the prices observed for other products in the same family. The proxied price is then chained. In such cases, therefore, the chaining method is supplemented by a method for estimating missing prices. Unfortunately, another frequent occurrence is that the missing prices cannot be estimated or, rather, that the estimation merely consists in carrying forward the last observed price (before a discount sale). As a result, the chaining method—a constraining hypothesis in itself—is compounded by the implicit and questionable hypothesis of price invariability between the two periods.

For clothing products most exposed to fashion effects (i.e., the products that are most sensitive to the winter/summer collection cycle), up to 80% of the sampled products may be replaced during one year. The chaining method is applied to one-half of these replacements. It totally eliminates the price difference between collections by attributing it implicitly to a fashion effect—which, of course, is unjustified. Moreover, the method is often applied with the implicit assumption that prices will not vary for the entire non-overlap period, which, for clothing products, may be as long as a year. If actual prices are trending up, the index will be biased downward.33 The statistics show that consumer price indexes for dresses, skirts, and other female outerwear are lower than indexes of other clothing products, notably underwear and millinery, which are less sensitive to collection cycles and hence less subject to replacement. A U.S. study (Reinsdorff, Liegey, and Stewart 1996) also suggests that the

33 Reciprocally, if actual prices are trending down, the index will be biased upward.
chaining method may have led to an understatement of inflation for clothing products in the U.S.

In sum, whatever the defects that may occur in the chaining method, they will provide no indication of whether the errors overstate or understate inflation. The same is also true of the second most widespread method, direct comparison, which is used for 40% of item replacements in France.

III.6. Direct comparison

Direct comparisons are made when price collectors manage to find a replacement product whose characteristics sufficiently resemble those of the replaced product to enable the former to be regarded as an "equivalent" of the latter. The entire price difference between the two products is regarded as a "true" price difference (i.e., one assumes there is no quality change). France tries to apply direct comparison as widely as possible in its price-index collection procedures. The method has the immense advantage of apparently not requiring a quality-change valuation. However, it proves defective in the presence of improvements or deteriorations that are not apparent or do not form part of the product's defining characteristics. In such cases, no replacement or treatment is applied despite the obvious need for such an operation in view of the quality change.

One illustration of a possible upward bias due to the defect of the direct comparison method in France is the automobile-repair index. Many price quotations for this grouping consist of hourly fees. In the long run, this approach would therefore overlook a possible decrease in time spent on a well-defined repair procedure. In at least one specific year, French national accountants did make a sizable downward adjustment in the CPI for this grouping. Another example of potential upward bias is the "bank card" service. The French CPI tracks the price of the annual bank-card membership fee. Over the past decade, however, two phenomena have occurred: the number of automated teller machines (ATMs) has risen sharply, and some complementary services (such as travel insurance) have been added to the basic bank-card services. These improvements in the quality of service offered by the bank card have not been deducted from its CPI price. In truth, it would have been very complicated to measure the value that should have been deducted.

Not all the examples, however, point to upward bias. We cannot rule out the possibility that the quality of certain services and the durability of certain goods may have decreased, and that the CPI has failed to record these negative quality changes. For example, one often hears complaints about the falling quality of the baguette (bread loaf), the quintessential symbol of French food consumption. If the accusation was correct (for a single given sales outlet), the decrease in quality would probably have been overlooked by the CPI. The claim of a fall in quality for certain durable goods was made, in particular, during the debate of the early 1970s over the French CPI (Piriou 1983). In the debate over the Boskin Commission Report, several economists spoke of the possible deterioration in public services, such as unsafer subways (Norwood 1995; Kuttner 1995).

III.7. "True" new products
The problem of how to deal with new products that lack a true past equivalent is actually just a variant on the problem of dealing with replacements, but it is even more complex. Examples of "true" new products are a simple service like pizza home delivery, which has expanded considerably in recent years, or high-tech goods such as interactive video games and mobile telephones. These cases raise two questions: (1) When should these new products be included in the index? (2) How should they be included?

Some economists have criticized the lag affecting new-product incorporation into the CPI. Most new-product prices, especially in industries using high technology, follow a pronounced L curve. The new product is launched at a high price, and sold in small quantities. As production is optimized and sales grow, prices come down swiftly. Having then reached "maturity," the price of the "former new" product stabilizes as a new product replaces it, entering a price cycle of its own. Consequently, if CPI producers are routinely late in entering new products into the index, this means they will track only that part of the curve which plots the product's maturity, neglecting the first, downward-sloping part of the curve. Hence—it is alleged—an upward bias in the CPI.

To begin with, we can point out that this accusation overlooks one crucial fact: the weighting assigned to new products in the index would be extremely small if they were included at the very start of their cycle. Also, the French index is less open to the charge than others. CPI statisticians not only update the grouping weightings annually, but conduct a thorough annual review of the 1,000 "sub-items" (variétés) that make up the most detailed product families whose prices are monitored. Each year, nearly a hundred sub-items are altered, introduced, or deleted. The main purpose of this update is to introduce the latest products into the CPI once they have attained significant market share. A European regulation has strengthened this procedure by obliging a member State to introduce a new product into its index (1) as soon as any one of the other 14 member States has done so and (2) if the new product accounts for at least 1 part in 1,000 of the weighting of the country’s overall index.

III.8. A new product increases utility from its introduction

The more crucial question, however, is to define the procedure for introducing these products into the index. The current practice is an implicit use of the chaining method. Let us take the example of mobile telephony. Once we have accepted the fact that this service is a significant household-expenditure grouping, the sub-item will be introduced in December of the current year into the new sample of goods and services defined at each year-end. The change in prices of mobile telephones will therefore be tracked from the following January in the overall index, with its own weighting, like any other product. But the fall (or rise?) in the price of this new product will never have been measured at the time of its introduction. Some economists see this as another source of inflation overstatement, for the introduction of a new product—by definition—improves consumer utility. A simple example will help to explain this (Oulton 1995). Let us assume households take two months' vacation a year abroad, in the form of a package tour. Let us also assume package-tour prices stay constant but the range of destinations expands every year, at that same price. A conventional CPI will not register any

---

34 Household expenditures should not be confused with business expenditures. In the case of mobile telephones, it is clear that the initial purchases were mainly by businesses and were therefore not to be included in the CPI. Only recently has the market been extended to direct purchases by households.
change, even if it incorporates the new destinations as quickly as they are first offered. Yet the real value of the package-tour supply has visibly increased.

The impact on *prices*, however, is very hard to measure. We would need to estimate what some authors have termed the "reservation price," i.e., the price that would eliminate demand in the previous period. There have been few if any practical experiments with estimates of this kind for such recent products. Furthermore, if we could estimate the reservation price, it would not necessarily follow that the new product should be introduced as soon as possible into the CPI. This is because the new product's reservation price in the period prior to its introduction would be extremely delicate to estimate at the very time when the market for the product was very narrow and confined to the upper strata of consumers willing to pay very high prices. In particular, introducing the product at that time and with that type of treatment might well entail a price rise offsetting a fall that—owing to the product's delayed introduction—would not have been observed. Here again, it is not so easy to conclude that all index biases are overstatements.

Despite these uncertainties, many economists—including members of the Boskin Commission—believe that, as our competitive economies generate an overall improvement in the quality and range of products, most price-index biases lead in one direction: an understatement of quality improvements, and a consequent overstatement of inflation. As we have seen, however, there is no evidence today to go beyond a series of presumptions. In the following pages, we will show that all the figures circulating on this source of upward bias—including the Boskin Commission number—are fragile and probably exaggerated. In the current state of our knowledge,natural text content for III.9. New-product bias cannot be effectively estimated... III.10. ...by extrapolating partial studies...
To put the matter into perspective, durable goods account for 10% of the total French CPI. The direct impact of the overestimation on the French index would therefore come to 0.1 points. However, as the Boskin Commission\textsuperscript{36} itself points out, the estimated annual bias of 1 point actually includes several types of biases at the same time, and cannot be regarded as an indication of the new-product bias alone. The use of the durable-goods bias to estimate an overall new-product bias is therefore misconceived on two counts: (1) It extrapolates a figure to product categories not initially examined by Gordon and for which the most conservative hypothesis expressed is that there is no upward bias. As we have seen, though, we cannot rule out the possibility of downward bias. (2) There is a double-counting error. Yet that is the main source of the figure of 0.6 points extensively cited in several articles.

### III.11. ...or by the Boskin Commission approach

Rather than extrapolating partial studies, as other similar studies had done, the Boskin Commission therefore launched into an estimation, product category by product category, based on broad economic assumptions and on direct estimates. The Commission found an overall figure for new-product bias of 0.6 points per year. The Commission's arguments are generally very interesting and well-supported. However, some of its American critics have described its approach as a series of "guesstimates." To illustrate this attack, let us take the Commission's approach to fruits and vegetables. It estimates the bias due to the neglect of quality changes for fruits and vegetables at 0.6 points a year for the past 30 years.\textsuperscript{37} This figure derives directly from the following statement: "How much would a consumer pay to have the privilege of choosing from the variety of items available in today's supermarket instead of being constrained to the much more limited variety available 30 years ago? A conservative estimate of the value of extra variety and convenience might be [...] 20 percent for produce where the increased variety in winter (as well as summer farmers' markets) has been so notable [...]." As this effect would not have been taken into account by the CPI, we need only take the 30th root of 1.2 to obtain 1.006—hence the 0.6-point annual bias for fruits and vegetables.

This argument is questionable for three reasons: (1) Why 20%? Why not 10% or even 30%? The figure is totally subjective and rests on no other data than the Commission's own opinion. (2) No allowance—even subjective—is made for a potential decline in the intrinsic quality of fruits and vegetables. (3) The Commission implicitly enlarges the notion of quality to include concepts that clearly exceed the range of currently measurable factors. Indeed, if we look closely at the assumption on fruits and vegetables, we realize that it concerns not so much the quality of the products themselves but the "increased variety in winter." This implies that a broader product range, in itself, improves consumer utility. While this proposition is conceptually acceptable, is remains impossible to quantify and thus cannot be included as a quality factor in an operating procedure. It is crucial to bear in mind that the CPI can only track the quantifiable components of quality change. Economists toy with many explanatory models, but very few of these have attained statistical operational status.

\textsuperscript{35} Gordon finds an overstatement of 1.54 points a year for the period 1947-83, which he breaks down as follows: 2.21 points for 1947-60, 1.24 points for 1960-73, and 1.05 points for 1973-83.

\textsuperscript{36} Of which R. Gordon was a member.

\textsuperscript{37} The fact that the upward bias for fruits and vegetables is identical to the Commission's figure for the overall index is a coincidence.
The extension of the Boskin Commission's procedure for fruits and vegetables to many other product categories cannot conceal the shakiness of the initial assumptions—hence of the 0.6-point estimation of the overall upward bias. Moreover, the Commission neglected the cases where the methods used for new products (clothing in France or the U.S., for example) could lead, instead, to an understatement of inflation. This neglect provides another argument for a far lower bias figure. But we do not want to enter into this kind of reasoning; accordingly, we refuse to advance a figure of our own. In such cases, statisticians must simply admit their ignorance, and work to reduce it rather than to venture uncertain estimates.

We now turn to the issue of how the CPI should treat the gains in market share by mass merchandisers selling at lower prices than conventional outlets. This issue is clearly linked to the two issues discussed above. If a product transaction is defined not only by the physical characteristics of the product itself but also by the related services surrounding it, the phenomenon of new outlets can be compared to (1) product substitution (gains in market share by existing mass merchandisers) and (2) new-product introduction (opening of new outlets). The scale of the new-outlet phenomenon, however, calls for separate treatment.

38 In France, the CPI's exclusion of hospital services would, paradoxically, help reduce a bias if there was one. All U.S. authors agree that hospital-service price inflation is grossly overstated, as the existing indexes rarely take into account the improvements in health-care results due to new medical technologies.
IV. New sales outlets

France, like other countries, has witnessed the growth of new outlets offering products at lower prices. Year after year, these mass-merchandising outlets have captured market share from conventional stores. After supermarkets, followed by hypermarkets, the trend has been fueled more recently by the emergence of "hard-discounters" and, in the service sector, by franchise chains—for example in automobile repair or photography. Deregulation has led to the same phenomenon in air transportation.

In this section, we show that the CPI calculation method—in France and elsewhere—may not fully track the fall in consumer prices in a given region or market due to the establishment of a new outlet or service producer. The method may also understate substitution bias if the weightings of each outlet type are not updated in a timely manner.

The method used for a new store is the chaining method—the same one we described previously for new products. This consists in introducing the new price quotations at an index level equal to the index of prices of earlier quotations in the same urban area. Let us take an example where the price of a liter of soda in urban area A was FF12 in December 1996, yielding an index of 112.3 on a 1990 = 100 base. A new supermarket opens in that urban area, and the first price quotation recorded in the outlet in the same period is FF8 a liter. The starting point for the soda-price ratio in the new supermarket will also be set at 112.3. As a result, its merger with the other price ratios in the urban area will not cause a fall in the soda-price index in A.\(^{39}\) The index will not record a change unless the small shopkeepers (or the other mid-sized or mass outlets) whose prices were tracked beforehand lowered their own prices in response to competition from the new supermarket.

IV.1. Price differences are regarded as differences in quality of customer service

In fact, it is as if statisticians regarded the total price difference between the two outlet types for an identical product as somehow due to a difference in the quality of customer service. Admittedly, shopping in a conventional store and shopping in a supermarket are not equivalent, even if the product sold is exactly the same. Shopping close to home, personalized service, and friendliness are often listed as arguments in favor of convenience stores. Actually, low prices are not the only reason for the growth of mass merchandising. The phenomenon is also largely linked to the rise of the "automobile civilization," the expansion of suburbs, and the advent of home freezers: all these developments have enabled consumers to make large, grouped purchases in outlets offering a new kind of customer service. However, the steady gains in market share by mass-merchandisers also originate in the keen competition and fierce price-wars between outlet types. This situation constitutes equally solid evidence for arguing that the statisticians' implicit assumption—i.e., the entire price difference is due to the

---

\(^{39}\) Because of the lower price charged by the supermarket, the soda-price index does not fall in urban area A. This will lead to a sizable difference between (1) the estimated "volume" (in the national accounting sense) of soda sales in the urban area at the time of the supermarket opening and (2) the change in the number of liters sold. This is because the procedure for introducing the new supermarket into the price index implies that a liter of soda sold in a supermarket is less "good" (\(^{3}\)) for the consumer and therefore "weights" less in the total volume sold than a liter sold in a conventional store at a higher price. The underlying notion is that the supermarket provides less "customer service" to soda purchasers than a conventional store.
difference in service—is exaggerated. The omission of price decreases due to the growth of mass merchandising is plainly a source of upward bias in the CPI. The appropriate statistical treatment would be to assess the value attached by consumers to their shift from one type of outlet to another. Some studies on the subject have been conducted in the U.S., but have not yet been translated into operating procedures. One proposal would be to regard one half of the price difference between outlet types as a "pure" price difference and the other half as a difference in service. This solution, however, is arguably just as arbitrary as the alternative outlined earlier.

An upward bias might also occur if the weightings of each outlet type were not updated regularly: the problem here is similar to product-substitution bias. The current procedure for annual revision of the French CPI sample does include an update of the outlet-category weightings. Under the procedure, the CPI regional units, once a year, transfer price quotations from outlet types that are losing market share to outlet types that are gaining market share. This implicitly increases the weighting of the latter type. For this purpose, the regional units use market-share data compiled by INSEE's product specialists. The quality of the procedure notably depends on the quality and timeliness of the market-share data available. As we saw in our discussion of intermediate-level substitution bias, there is probably some room for progress in this area.

**IV.2. Less than 0.2-point outlet-substitution bias in France in the late 1980s**

The most thorough assessment of the impact on the CPI of gains in market share by mass merchandisers is a French study published in 1995 in *Économie et Statistique* (Saglio et al. 1995). Extrapolating a detailed monograph on chocolate bars, the authors compare the difference between an index calculated using the conventional method and an index in which the entire price difference between outlet types is treated as a "pure" price difference. The difference between the two indexes is estimated at 0.2 percentage points a year in the 1980s. This figure, called "outlet-substitution bias," thus probably overstates the bias due to new outlets, if one accepts that the price difference is at least partly due to a difference in customer service. If we assume that only half the difference between the indexes is a price difference, the bias is reduced to 0.1 points per year.

Can we regard this 0.1-point figure, whose starting point has been estimated for the 1980s, as a representative value for the years ahead? There are two factors to consider here. (1) The mass-merchandising boom in France has probably lost momentum. The slowdown will be furthered by a recent law that places new restrictions on superstore openings. This is an argument for adopting a lower bias figure. (2) The initial 0.2-point estimate may itself be on the low side, since it does not take into full account the emergence of new outlet types and of deregulation in services, transportation, and telecommunications. Some striking examples of this are the rise of franchise chain stores in auto repair and photography, and the spectacular impact of deregulation on prices and market share in domestic air transportation. Indeed, in 1993, the price index for air transportation services estimated in the French national accounts had to be estimated well below the CPI. Much scope for price wars remains as new players enter the banking and insurance industries—for example, through telephone banking and

---

40 The alternative index is based on average prices. The difference between standard price indexes and average prices is explained in box 4.
insurance—not to mention teleshopping on the Internet. Overall, the defect of the method of introduction of new outlets may have created a small upward bias of about 0.05-0.15 points a year in the French CPI. This range is consistent with the Boskin Commission's figure for the U.S.
Box 4

Price indexes and average prices

Information on certain price movements is provided by other, non-INSEE sources. Among the more significant are the studies based on data originating from data bases of market-research firms. These firms monitor in extreme detail the changes in prices and quantities purchased. This is done by surveying consumer or merchandiser panels. From these sources, analysts calculate average prices by product family, whose movements sometimes differ substantially from the corresponding INSEE price index. The average price is generally equal to the product family's total sales divided by the quantities sold (by weight or in units). This method is not the same as the one used in the price index.

The average price often combines products of different quality (from entry-level to top-of-the-range), in different packagings (sales by unit or in packs of several units) distributed through different types of outlets. The larger the product family, the more the average-price calculation will mix different characteristics. If households buy more entry-level products, the average price will fall. If, instead, they prefer top-quality products, the average price will rise. Not all these movements can be equated with price changes as measured by the price index. Their possible causes range from income effects to fashion effects and other phenomena. To take a borderline example, the average purchase price of an automobile may rise from one year to the next simply because households have bought more large-engined cars and fewer small-engined cars in the second year: this shift is due to a change not in the price of either category, but in income and tastes. Clearly, this is not a price rise as defined by the price index, since the households' utility level will not have remained constant.

The approach most commonly used for the price index is to assume that the difference in price between two products—even similar ones—is indicative of a difference in their quality. The average-price computation rests on a diametrically opposite hypothesis. The fullest study on the difference between a price index and an average price concerns the chocolate-bar market in France (Saglio 1995). The study showed that the average price of a chocolate bar (irrespective of brand, sales outlet, and packaging) fell 1.6% a year between 1988 and 1990 whereas the bar's Laspeyres price index fell only 0.2% a year. The 1.4-point annual difference between the two indexes breaks down as follows:

- 0.5 points is due to a product-range effect (substitution of cheaper, i.e. low-end, products for more expensive, i.e. high-end, products);
- 0.1 points is due to a packaging effect (substitution of purchases of bars in packs of two or three units, at a lower unit price, for purchases of bars by the unit);
- 0.8 points—the largest component—is due to outlet substitution (consumers switching from conventional retailers to hypermarkets).

The substitution of low-end products, particularly "best-price" items, for high-end products is not regarded as a fall in price, since the product qualities are not comparable. Outlet-substitution bias is discussed at length in §III of the main text.

Other phenomena may also explain the differences between average prices and the price index. Annual shifts in purchase volume toward sales seasons have no effect on the price index, although INSEE now takes full account of advertised sales, discounts, and promotions. The explanation is that the CPI measures a price level at a given instant (the month) and not an average price for the year weighted by monthly sales. The latter figure would, of course, be influenced by the shift in sales volume toward sales seasons.
V. Conclusion: a mild upward substitution bias, but a difficult and expensive adjustment

Table 1 reports the estimates we have arrived at. As we announced in the introduction, we are not proposing an overall upward-bias figure for France, since—for the reasons set out earlier—we have not quantified new-product bias. However, we make no secret of our belief that the Boskin Commission's 0.6-point estimate for new-product bias is, in any event, highly exaggerated. Several U.S. statisticians and economists (including Moulton 1996) have also voiced their doubts about the Commission's figure. One line of criticism, while not totally convincing, does stress the fact that a long-run extrapolation of the Commission's figure would entail a drastic revision of the economic history of the U.S. (Baker 1996). In particular, one would find a massive share of the population below the poverty line in the 1960s.

Table 1: United States and France: upward bias in CPI

<table>
<thead>
<tr>
<th>Upward-bias type</th>
<th>United States</th>
<th>France</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated value, % per year, for years after 1996</td>
<td>Estimated value, % per year, for years after 1996</td>
</tr>
<tr>
<td>Upper-level substitution</td>
<td>0.15</td>
<td>-</td>
</tr>
<tr>
<td>Intermediate-level substitution</td>
<td>0.25</td>
<td>0.05-0.10</td>
</tr>
<tr>
<td>Lower-level substitution</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>New outlets</td>
<td>0.10</td>
<td>0.05-0.15</td>
</tr>
<tr>
<td>TOTAL EXCLUDING NEW PRODUCTS</td>
<td>0.5</td>
<td>0.10-0.25</td>
</tr>
<tr>
<td>New products</td>
<td>0.60</td>
<td>?*</td>
</tr>
<tr>
<td>TOTAL INCLUDING NEW PRODUCTS</td>
<td>1.1</td>
<td>?</td>
</tr>
</tbody>
</table>

*As indicated in the text, we find it impossible to volunteer any estimate for this line and consequently for the overall total on the bottom line. However, we believe the Boskin Commission's estimate of 0.6 percentage points is greatly exaggerated.

Disregarding new-product bias, the Boskin Commission findings point to an upward bias of 0.5 points. Within the same scope of coverage, our estimates for France yield a range of 0.1-0.25 points. As we have seen, the gap is due to the different statistical methods used in the French and U.S. indexes. Some of the methods used in France are included in the Commission's recommendations to BLS.41

---

41 The recommendations include geometric means, annual chaining, and the annual revision of entry-level items (our "sub-items").
True, a range of 0.1-0.25 points is relatively modest and, most importantly, has little impact on monetary policy or on the growth rate (box 5)\textsuperscript{42}. Naturally, however, one cannot admit an upward bias without doing something to correct it. INSEE is working to improve its methods. The recent introduction of the geometric mean was part of this program. The forthcoming introduction of insurance into the CPI, the gradual shift to a more extensive monthly system of price-quotation gathering, and an improved treatment of missing values in the CPI will achieve substantial progress. However, the adjustment for the mild upward bias in the two lines of table 1 where an overestimation may exist would be far more complex and expensive that the reforms just mentioned.

The CPI is a vital tool for all economic agents—including ordinary citizens and decision-makers alike. The methods used to construct it must be transparent, relevant, and reliable. The present article itself admits that the estimate of this weak upward bias (0.1-0.25 points) is speculative. INSEE cannot, therefore, use the figure as such. Rather, the Institute needs to improve its operating methods with the aid of information based on hard data. The task will therefore require time and money.\textsuperscript{43} INSEE’s experience in assessing the quality-change value for PCs showed the high cost of a reliable measurement of quality-change value even in one particular category. The Institute is continuing to explore this approach, and an econometric method for estimating quality-change values will soon be operational for other selected durables, notably dish-washers. On the other hand, INSEE does not have the means to generalize these methods single-handedly.

In conclusion, however, we can point to two paths for the future:

(1) European statistical institutes are beginning to pool their research resources on quality-change values. This raises the hope of large economies of scale in the years ahead. Several multilateral task forces set up on Eurostat's initiative have begun work this year, notably on clothing, personal computers, durable goods, and automobiles. INSEE is counting heavily on the build-up of this international research program, which is expected to extend, in particular, to North American countries. The automotive and durable-goods markets have become global, and it makes no sense to squander resources on attempting identical estimations in all countries. A division of labor in the international network of statistical institutes will benefit all users.

(2) INSEE is considering a more massive use of detailed data from private market-research companies. Thanks to their high information content and their prompt availability, these sources will probably be among tomorrow's solutions for correcting lower-level substitution

\textsuperscript{42} A greater bias might have an impact if monetary policy targeted specific inflation levels. Indeed, it is significant that the Federal Reserve chairman, Alan Greenspan, is one of the harshest critics of the U.S. CPI. An economist pointed out that one could construe the CPI as a vindication of Greenspan's monetary policy. It will be noted that the Maastricht Treaty inflation criterion is defined in relative terms, with price growth not to exceed that of the three best-performing countries by more than 1.5 points. As EU member States use very similar methods to calculate their CPIs, virtually all the countries would exhibit the same bias. As a result, the impact on the Treaty criterion would be nearly neutralized. To assess the impact of a large upward bias on monetary policy, we would also need to know more about the behavior of the bias under sharp inflation swings. At present, there is not enough data on this point.

\textsuperscript{43} The French CPI statistical team is smaller than its U.S. counterpart. As a result of the controversy over the Boskin Commission findings, BLS obtained significant additional funding, which has recently enabled its head to launch a program to improve the index (Abraham 1997).
bias and outlet-substitution bias. This approach is consistent with the Boskin Commission recommendations and with initiatives in many other countries.

44 A.C. Nielsen France, for example, has offered INSEE the free use of highly detailed data on selected markets, under a joint research program beginning in 1997. The same company supplied the data that enabled INSEE to carry out the study on chocolate bars mentioned earlier in this article.
Box 5

Impact on the national accounts

One of the basic purposes of a consumer price index is to enable statisticians to determine the change in consumption "volume" (also known as "real consumption" or "consumption in constant price") from the change in consumption in nominal value. These movements are summarized in the National Accounts. If we assume that the accounts in nominal value are fixed, any correction in the price index will therefore affect the "volume" (in what is termed the "volume/price" breakdown). As household consumption accounts for 60% of GDP, any correction on the "volume" of that consumption may have a substantial impact on the GDP growth figure.

If the CPI did indeed overstate inflation by a significant margin—let us say x%—how would this affect the GDP growth number? The question is all the more deserving of attention as the sizable overstatement of U.S. computer price changes—revealed in the late 1980s—caused U.S. national accountants to make a very substantial upward adjustment in their growth estimates.

With regard to France, we will show that the impact of a CPI overstatement on growth would be very weak.45

Contrary to what one might readily assume, the following simple argument is wrong:
(a) Since national accountants use the CPI in their computations, the x% overstatement in the index also applies to the national-accounting price index.
(b) Since the nominal value series in the national accounts are fixed, lowering the price index means increasing the volume.
(c) And since household consumption accounts for 60% of GDP, the impact on GDP is equal to 0.6 times x%.

Reality is nowhere near as simple, as the following pages will show (with apologies for the frequent but necessary technicalities).

Correction of the upper-level substitution bias: no impact

We must begin by distinguishing the "upper-level substitution bias" from the other biases allegedly responsible for the total CPI overstatement of inflation. The reason for this is that the French national accounting system does not directly use the overall CPI. Instead, it aggregates the intermediate-level consumption price indexes (roughly equivalent to U.S. item strata) using a distinctive procedure (which differs from the CPI method), in order to produce its own overall index. One of the relatively little-known consequences is that the price index for household consumption in the national accounts is not the same as the CPI despite the fact that national accountants use the detailed CPI-based indexes to prepare their accounts, while CPI producers use the national accounts to set their weightings!

The first step in measuring this difference is to correct the coverage of the national accounts, as they define household market consumption more broadly than the CPI. The following items have to be excluded from the national accounts: consumption for own account (including imputed rents), insurance, clinics, private hospitals, institutions for the

---

45 We shall here disregard the fact that the reasons for the CPI overstatement may also apply to other price indexes used in measuring GDP components other than household consumption.
elderly, and air transportation (before 1992). Second, it is important to realize that the French national accounts actually use two implicit price deflators. One is obtained by chaining the quotients of the division of the accounts in nominal value by the accounts "at year n-1 prices." The other deflator is calculated by dividing the accounts in nominal value by the accounts "at 1980 prices." Thus there are two implicit price deflators to be compared with the CPI. We shall refer to the first index as the "base n-1, chained" deflator and the second as the "base-1980" deflator. The two series, set by convention to 100 in 1980, are shown in table A.

Table A  
Discrepancies between CPI and national accounts

<table>
<thead>
<tr>
<th>Year</th>
<th>CPI Annual Average</th>
<th>National-accounting deflators, using CPI coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Annual average</td>
</tr>
<tr>
<td>1980</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>1985</td>
<td>158.0</td>
<td>157.3</td>
</tr>
<tr>
<td>1990</td>
<td>184.0</td>
<td>181.4</td>
</tr>
<tr>
<td>1995</td>
<td>205.3</td>
<td>201.8</td>
</tr>
</tbody>
</table>

Annual average change, 1980-95

<table>
<thead>
<tr>
<th></th>
<th>CPI</th>
<th>National-accounting deflators, using CPI coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.27%</td>
<td>5.14%</td>
</tr>
<tr>
<td></td>
<td>4.99%</td>
<td></td>
</tr>
</tbody>
</table>

As we can see, the 15-year changes in the CPI and in the two series diverge. The CPI ends the period at 205.3, or 3.5 points above the base n-1 chained deflator at 201.8, which is 3.9 points higher than the base-1980 deflator. Over the 15 years, the CPI gains an average 5.27% a year, compared with 5.14% for the base n-1 deflator and 4.99% for the base-1980 deflator. The difference per year between the first and second indexes is 0.13 points; between the first and third; 0.28 points; between the second and third, 0.15 points.46

These differences are due to the fact that the national accounts use different formulas for the upper-level aggregation. The national-accounting deflators are Paasche indexes whereas the CPI is a Laspeyres index.47 More accurately, the base-1980 deflator is a pure Paasche index whereas the base n-1 deflator is a chained Paasche index. The CPI, as mentioned earlier, is a chained Laspeyres index. The differences between the indexes are therefore hardly surprising. The consequences are familiar:48 as (quasi-general) rules, (1) a Paasche index is lower than a Laspeyres index; (2) a chained Paasche index is higher than a pure Paasche index; (3) a chained Laspeyres index is lower than a pure Laspeyres index; (4) a chained Laspeyres index is higher than a chained Paasche index.

46 The differences calculated for the period 1985-95 are identical in ranking and magnitude.
47 For the definition of Paasche and Laspeyres indexes, see appendix II, §1.
48 For a partial description of them, see appendix II, §1.
What would happen, then, if the CPI were adjusted downward to eliminate the upper-level substitution bias—i.e., by a change in the upper-level aggregation formula? The answer may come as a surprise to some: the correction would have no impact whatsoever on the official GDP growth statistic, since national accountants use an entirely different upper-level aggregation formula. In any case, this is not even an issue in France, since—as we have seen—its CPI is free of upper-level substitution bias. But even in the U.S., where the Boskin Commission estimated the inflation overstatement due to the bias at 0.15 points a year, the adjustment for the bias would have no effect on the growth figure.\footnote{The U.S. has been using a chained Fisher index as a national-accounting deflator since 1995. As in the French situation reported in table A, the U.S. national account deflator is therefore in all likelihood well below the U.S. CPI—which, as pointed out earlier, is a Laspeyres index based on 1983-85 weightings. The Boskin Commission recommends the introduction of a Fisher index (or a Tornqvist index, which is similar) into the CPI. If this were done, the probable result would simply be a convergence of the U.S. CPI inflation estimate toward the national-accounting estimate, with no change in the latter.}

Estimating the impact on growth of adjustments for other biases: a complex task

On the other hand, if the CPI required a downward adjustment owing to inappropriate procedures at the intermediate and lower levels and/or the procedures for handling new products, there is reason to believe that the changes would have an extensive impact on the growth figure. This is because the adjustment would focus on the intermediate and lower-level indexes—the very ones used directly in the national accounts. Assuming fixed nominal values, we may therefore suppose that the downward adjustment in the price indexes would automatically entail an upward adjustment in the volume figures. However, even for past years, the initial assumption of this argument—"data in nominal value are fixed"—does not hold, at least for some of the GDP components. For example, the national accounts calculate a portion (admittedly, a small one) of nominal household consumption not from a direct reading of a nominal series but by applying the price-index series to a volume series. Thus a downward adjustment in the price index would not affect the volume figures, but it would have an impact on the nominal values.

This conclusion may seem surprising in view of the fact that GDP—at least in the past—has been largely calculated in nominal terms by pooling the nominal accounts of enterprises, government units, and other economic agents. Two factors need to be considered, however: (1) there is a margin of uncertainty on the actual figure for value added in nominal terms, if only because business surveys are not exhaustive; (2) some other nominal GDP components are not quantified with great precision. One notable example is the change in inventories, which is very hard to estimate on a nominal basis in the national accounts, despite the sources provided by corporate accounts. The problem is due to differences in the inventory-valuation procedures used in the national accounts and in business accounting.

On balance, the impact of a downward adjustment in the CPI on volume series would certainly be smaller than initially assumed. Calculating the exact impact, however, would demand a costly simulation of the mechanisms used to construct the national accounts. In any case, as we have seen, the overstatement in the French index is probably small. As the bias would be further lessened by the procedures used in compiling the national accounts, the adjustments on past growth series would be anything but drastic.
REFERENCES


BLS (1995), Report from the BLS for the House Budget Committee, United States Congress, April 28, U.S.


Appendix I

Some other estimates of potential CPI upward bias:
United States, Canada, United Kingdom

<table>
<thead>
<tr>
<th>Bias type</th>
<th>1 U.S. Boskin</th>
<th>2 U.S. Wynne-Sygalla</th>
<th>3 U.S. Shapiro-Wilcox</th>
<th>4 U.S. CBO</th>
<th>5 Canada Crawford</th>
<th>6 U.K. Oulton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substitution</td>
<td>0.4</td>
<td>&lt;0.2</td>
<td>0.2</td>
<td>0.1-0.3</td>
<td>0.1-0.2</td>
<td>small</td>
</tr>
<tr>
<td>New products</td>
<td>0.3</td>
<td>?</td>
<td>0.2</td>
<td>0.06</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Quality change</td>
<td>0.3</td>
<td>?</td>
<td>0.25</td>
<td>0.1-0.2</td>
<td>0.0-0.2</td>
<td>?</td>
</tr>
<tr>
<td>New outlets</td>
<td>0.1</td>
<td>?</td>
<td>0.1</td>
<td>0.2-0.3</td>
<td>0.0-0.1</td>
<td>small</td>
</tr>
<tr>
<td>Formula</td>
<td>0.2</td>
<td>?</td>
<td>0.25</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>1.3*</td>
<td>&lt;1</td>
<td>1.0</td>
<td>0.6</td>
<td>&lt;0.5</td>
<td>small but &gt;0</td>
</tr>
<tr>
<td>Plausible estimation range</td>
<td>1.0-1.8</td>
<td>0.6-1.5</td>
<td>0.4-0.8</td>
<td>0.16-0.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This figure of 1.3 percentage points differs from the 1.1 points listed in table 1 of the main text. The reason is that the values shown here apply to the years prior to 1996, which, according to the Boskin Commission, exhibit a formula bias of 0.2 points eliminated after 1996.

4: Congressional Budget Office study, 1995.
Appendix II

1. An outline of price-index theory

There are two conventional approaches to price-index theory: the "economic" approach and the "axiomatic" approach.

The economic approach is based on the simple theory of the consumer used in microeconomics. The price index is defined as the ratio of two expenditures. The denominator is the consumer's actual expenditure at the starting date. The numerator is a theoretical expenditure equal to the minimum expenditure that would preserve the consumer's initial utility level, given the price vector of the current period. In the chart below, which illustrates a simple two-product situation, the index is the ratio of the budget in $A'$ to the budget in $A$.

The COLI is the ratio of two budgets at constant utility

How to read this chart: A represents the situation in the base period. This point corresponds to the optimal situation in terms of the budget constraint represented by the straight line $D_0$ and the utility curve $U_0$. B represents the situation at the current period, with the new budget constraint represented by the straight line $D_1$ and the new utility curve $U_1$. $A'$ is defined as the point located on the first utility curve that is tangent to the parallel to the budget constraint $D_1$. $A'$ therefore corresponds to a (theoretical) situation in which the consumer remains on the same utility curve $U_0$ but consumes the basket of goods corresponding to current relative prices. The price index is equal to the ratio of the (theoretical) budget in $A'$ to the budget in $A$. One can see in this example that the basket changes from $A$ to $A'$, while the consumer remains on the same utility curve. This "ideal" price index, usually referred to as the cost-of-living Index (COLI) in the English-language literature, depends on the reference situation. There are actually two COLIs, one on the initial utility curve, the other on the final utility curve. The COLI could be calculated directly if the consumer utility curves were known, but this is never the case. Some series allow a rough estimate, however, provided we make certain assumptions about the shapes of the utility curves. If the curves display an "elasticity of substitution" as in the chart below,

---

50 The economic approach is attributed to the Russian economist Könnus. For a good description, see Deaton and Mullbauer 1980.

51 French statisticians prefer to describe this as a "constant-utility index," as the term "cost-of-living index" has been used in France in a different context.

52 This means that the consumer can substitute a quantity of product 1 for a different quantity of product 2 and remain on the same utility curve.
then we can easily prove that the Laspeyres and Paasche indexes respectively overstate and understate the COLI. It will be recalled that the Laspeyres index weightings are the expenditures based on base-period quantities, while the Paasche-index weightings are the expenditures based on current-period quantities. There is no need to be a great mathematician to reproduce these results using plain economic common sense. All other things being equal, the Laspeyres index overestimates the COLI because, as it based on initial-period quantities, it will overweight the product with the steepest price increase while the weight of that product in the COLI decreases. The Laspeyres index will therefore be positively "biased" relative to the COLI. The opposite can be shown for the Paasche index. If, instead, the utility curves display no substitutability (in which case they are said to exhibit a "complementary factors" profile), the COLI will be identical either to the Laspeyres index or to the Paasche index. If the utility curves display what is referred to as a Cobb-Douglas profile \((Aq^a q^b)\), then the COLI will be equal to the geometric price mean weighted by consumer expenditures. The most general provable finding is that the Fisher index is a good proxy for the COLI for a fairly large family of utility curves, but subject to restrictive assumptions such as the requirement that the curves should be homothetic (Diewert 1976).

The second approach is known as the "axiomatic" approach. It was developed by Irving Fisher, the U.S. economist who gave his name to the Fisher index. The approach lays down a set of more than thirty rules that the price index must obey. One example is transitivity: the index of 2 relative to 0 must be equal to the product of (index of 1 relative to 0) times (index of 2 relative to 1). Another example is that the index must be independent of the quantity chosen, i.e., the index must remain constant regardless of whether one of the prices is collected in grams, kilos, etc. The next step in the axiomatic approach is to determine which of the many possible averaging formulas "pass" the maximum number of tests. Unfortunately, no formula passes all the tests, but the Fisher index is one of those that passes the most. This finding, which corroborates the conclusion of the economic approach, suggests that the Fisher index—or other so-called superlative indexes—is probably one of the most suitable index formulas. In fact, it is the preferred index in the new international manual of national accounting, the 1993 System of National Accounts (SNA 93). While index theory recommends the Fisher index, it should be remembered that this superiority is demonstrated in a single-consumer framework. The aggregation to n consumers poses tough problems that await a solution. CPIs, however, are macroeconomic, not microeconomic indexes. Some economists therefore refuse to talk about a "bias" between the Laspeyres and Fisher indexes since it has not been proven than the Fisher index constitutes a benchmark macroeconomic indicator.

The French practice is to construct a Laspeyres index based on very recent weightings and then to chain the index. The use of a recent base year justifies two arguments: (1) the instantaneous index (the one that measures the monthly change in prices in the current year) is just about the best possible, since it is based on the latest weightings; (2) the long-term index formed by chaining these links is, in most cases, a reliable gauge.53 Indeed, at this level of aggregation, the simulations confirm that the chained index closely resembles the Fisher index and that, like the latter, it is bounded by the Laspeyres index and the Paasche index.

Summary of main index formulas

Laspeyres

\[
IL_{t/t_0} = \frac{\sum p_{t,i}q_{0,i}}{\sum p_{0,i}q_{0,i}} = \frac{\sum p_{t}q_{0}}{\sum p_{0}q_{0}}
\]

also written:

53 The chaining approach is not necessarily a panacea. In particular, it can lead to positive biases in the event of wide swings in relative prices, as illustrated below in §2 of this appendix. Such swings, however, rarely occur at the upper level. All the numerical simulations for the period 1980-95 effectively show that chaining makes the index converge toward—rather than diverge from—a Fisher index.
\[ IL_{t',0} = \sum w_0 R_{t',0}, \text{ with } w_0 = \frac{p_0' q_0}{\sum p_0 q_0} \text{ and } R_{t',0} \text{ the ratio } \frac{p_{t'}}{p_0}. \]

**Paasche**

\[ IP_{t',0} = \frac{\sum_{i} p_{t',0} q_{t',i}}{\sum_{i} p_{0} q_{t',i}} = \frac{\sum_{i} p_{t} q_{t}}{\sum_{i} p_{0} q_{t}} \], which is also written as follows, if we set \( w_i = \frac{p_{0,i} q_{0}}{\sum p_{i} q_{t}} \):

\[ \frac{1}{IP_{t',0}} = \sum w_i \frac{p_0}{p_t}, \text{ which is known as a "harmonic mean" formula.} \]

**Fisher**

\[ IF_{t',0} = \sqrt{IL_{t',0} \cdot IP_{t',0}} \]

It should also be noted that, in a demand-dominated market (i.e., when an increase in the relative price causes demand to fall), we find the following well-known inequality:

\[ IP_{t',0} \leq IF_{t',0} \leq IL_{t',0} \]

In the area of consumer prices, this inequality prevails in most situations.
2. Micro-indexes, geometric mean, and formula bias

A micro-index is an index for the lowest level of aggregation in the price index. At that level, weightings are rarely applied. In other words, the formula is based solely on prices and not on other variables such as quantities, which are unavailable to statistics institutes at such a detailed level. Two formulas were long used in the French CPI: (1) the "price-sums ratio" (hereafter PSR), a ratio of average unweighted prices; (2) the arithmetic mean of price ratios (hereafter AMPR). The PSR formula was used and will continue to be used for about half of the 25,000 micro-indexes of the French CPI, namely, those that track the so-called homogeneous sub-items. The AMPR formula was used for the other half, i.e., heterogeneous sub-items.

<table>
<thead>
<tr>
<th>Micro-index formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let ( p(i,t) ) stand for the price level of product ( i ) in a given lower-level aggregation in period ( t ), and ( S ) the sample of ( n ) products whose prices remain constant between dates ( b ) and ( t ). We have:</td>
</tr>
<tr>
<td><strong>Price-sums ratio (PSR):</strong></td>
</tr>
<tr>
<td>[ I_p = \sum_{i \in S} p(i,t) / \sum_{i \in S} p(i,b), ]</td>
</tr>
<tr>
<td><strong>Geometric mean of price ratios (GMPR):</strong></td>
</tr>
<tr>
<td>[ I_{i \text{GM}} = \left( \prod_{i \in S} p(i,t) / p(i,b) \right)^{1/n}, ]</td>
</tr>
<tr>
<td><strong>Arithmetic mean of price ratios (AMPR):</strong></td>
</tr>
<tr>
<td>[ I_{i \text{AM}} = (1/n) \sum_{i \in S} (p(i,t) / p(i,b)) ]</td>
</tr>
</tbody>
</table>

The choice of formula at a lower level was long thought to have no major significance. However, recent experiments in Canada, the U.S., France and other countries have shown that this is far from true, and that this choice sometimes had a greater impact than the choice of weightings at a higher aggregation level. The lower-level formula, therefore, automatically affects international index comparability. This explains the unifying regulation adopted under the European Union CPI harmonization program.

The EU regulation recommends the use of two formulas: the price-sums ratio (PSR) and/or the geometric mean of price ratios (GMPR). It rejects the use of the arithmetic mean of price ratios (AMPR). The latter, used in France for half of all micro-indexes, is therefore already being replaced by the GMPR.

As we will see below, the geometric-mean formula has three major advantages: (1) it is immune to "formula bias," i.e. the bias in the arithmetic mean due to chaining; (2) it is suitable for heterogeneous sub-items—which exhibit fairly diverse price levels—because it does not give greater weight to high prices than to low prices, as the PSR formula does; (3) it is based on a fixed weighting by value and not by volume: thus, when relative prices change, the implicit quantities used to weight the index change in an inversely proportional manner. The geometric mean is therefore compatible with a degree of price/volume elasticity that works for most of the micro-indexes: these are compiled from price quotations for products that are fairly substitutable in a consumer's shopping area. The geometric mean has long been recommended by the Statistical Department of the International Labor Office and is also in standard use at Statistics Canada since January 1995. Statistics Sweden uses a formula that resembles the geometric mean. Most EU countries will adopt it as well, in compliance with the harmonization regulation. Other, non-EU countries including the U.S. are considering its introduction. The geometric mean is one of the solutions recommended by the Boskin Commission.
To explain why the geometric mean is a "good" micro-index formula, let us begin by showing its superiority to AMPR in chained micro-indexes. A very simple example will suffice (Schultz 1994).

Let A and B be two products whose prices move as reported in table A during the five periods shown. Between period 1 and period 2, the price of product A doubles while that of product B is halved; then the price of product B doubles while that of product A is halved, and so on. It will be noted that, in the odd-numbered periods, prices return to their period-1 levels. Each month, we calculate an AMPR index for the change from the previous month ("AMPR t/t-1 index"); we then chain this index. As shown in table A, the chained index will rise to 244 in period 5 even though prices in that period are identical to those of period 1. This mechanism is known as "formula bias."

Table A : A chained AMPR index produces a positive bias

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of A</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Price of B</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>AMPR index t/t-1</td>
<td>125*</td>
<td>125*</td>
<td>125*</td>
<td>125*</td>
<td></td>
</tr>
<tr>
<td>Chained AMPR index</td>
<td>100</td>
<td>125</td>
<td>156**</td>
<td>195</td>
<td>244</td>
</tr>
</tbody>
</table>

* $125 = 100 \cdot (1/2) \cdot (20/10 + 10/20)$
** $156 = 100 \cdot (1.25 \cdot 1.25)$

This example is admittedly extreme. Although they are far less pronounced in real life than in this example, the « bounces » in relative prices that explain this drift can nevertheless be quite jumpy, causing situations of this type. It should be noted that the phenomenon is not confined to periods when the long-term trend is stable, as in our example. The bias will also occur if prices are moving on an uptrend or downtrend but in irregular jumps. The formula bias was detected in real price indexes using formulas derived from the chained AMPR. In the U.S., for example, formula bias in the overall CPI has been estimated at about 0.25% a year. This was also the case—but on a very small scale—with the French index, which is chained every year in December.

By contrast, it is easy to calculate that the geometric mean (GMPR) would yield a constant index (value: 100), free from the chained-AMPR formula bias. This is shown in table B.
**Table B : The geometric mean is bias-free**

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of A</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Price of B</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>GMPR index</td>
<td>100*</td>
<td>100*</td>
<td>100*</td>
<td>100*</td>
<td>100*</td>
</tr>
</tbody>
</table>

*100 = 100.[(20/10).(10/20)]^{0.5}*

But the geometric mean is not the only way to avoid the problem. We can also use an unchained AMPR, i.e., an AMPR consistently based on period 1. Taking the same example, we get:

**Table C : The unweighted arithmetic mean behaves differently**

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of A</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Price of B</td>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>AMPR index</td>
<td>100</td>
<td>125*</td>
<td>100**</td>
<td>125*</td>
<td>100**</td>
</tr>
</tbody>
</table>

*125 = 100.(1/2)(20/10 + 10/20)*

**100 = 100.(1/2)(10/10 + 20/20)**

The unweighted AMPR thus avoids long-term bias in the initial formula. However, there is a clear difference with the geometric mean in even-numbered periods. Between periods 1 and 2, for example, the GMPR index stays at 100 while the unweighted AMPR index moves from 100 to 125. This difference is due to underlying assumptions on the quantities of each product. The AMPR formula is actually equivalent to a Laspeyres index whose quantities are equal to the inverse of base-period prices (for product A, this would give $q_0 = 1/p_0 = 1/10$; for product B, $q_0 = 1/p_0 = 1/20$). As this is a Laspeyres index, the quantities are kept constant in periods 1 and 2.

The constant-quantities assumption is not very satisfactory, however. If products A and B are relatively substitutable—as is generally the case in micro-indexes—consumers will presumably react to the sharp rise in A’s relative price by substituting purchases of product B, whose price is falling, for purchases of product A. The 25% increase is therefore overstated. The geometric mean automatically performs some substitution, which explains why its value stays at 100. Under certain hypotheses about the utility-curve profile, the geometric mean indicates a unit elasticity of substitution. In other words, if the price of an item rises x%, the implicit quantity consumed will fall x% as well. This property is not extreme. Some products may have a higher elasticity of substitution. The geometric mean seems to be well-suited to simulating consumer behavior in regard to similar products, in a shopping area the size of an urban area. That is the case with French CPI micro-indexes. It is the combination of these properties that tilted the decision in its favor for use in micro-indexes. This choice in no way implies that the geometric mean is appropriate for the higher levels of CPI aggregation, where elasticities of substitution would probably be weaker. Nor is it suited to dealing with temporarily null prices (give-aways).

Other theoretical and empirical evidence supports this conclusion. For example, we could show that the harmonic mean of price ratios (which is the same thing as a Paasche index whose quantities are equal to the inverse of the current period) would produce an index of 80 in period 2 and thus overstate the general fall in prices. The GMPR formula is therefore a halfway measure between the AMPR and HMPR solutions.\(^{54}\) Furthermore, in our example, the geometric mean is equal to the product of the arithmetic mean and the harmonic mean ($1 = 1.25 \times 0.8$). One can also prove that, when the consumer utility curves are Cobb-Douglas, the geometric mean weighted by expenditure values is shaped exactly like the COLI. Applying this observation to micro-indexes, which

\(^{54}\) This finding is just another way of describing the well-known inequalities between harmonic, geometric, and arithmetic means.
assume equal expenditures for all products, we can deduce that the simple geometric mean is equal to the COLI if the utility functions are Cobb-Douglas. We can also show that the weighted geometric mean is equal to the Divisia index when the expenditure values are constant (Viglino 1995). This conclusion is obviously valid when those values are—in addition—all identical, as in the GMPR formula. Lastly, these convergent findings are bolstered by empirical calculations showing the great similarities between the GMPR formula and the Fisher micro-index formula, which is equal to the geometric mean of the HMPR and AMPR formulas (Quaranta 1995).