Catalogue no. 62F0014MIB

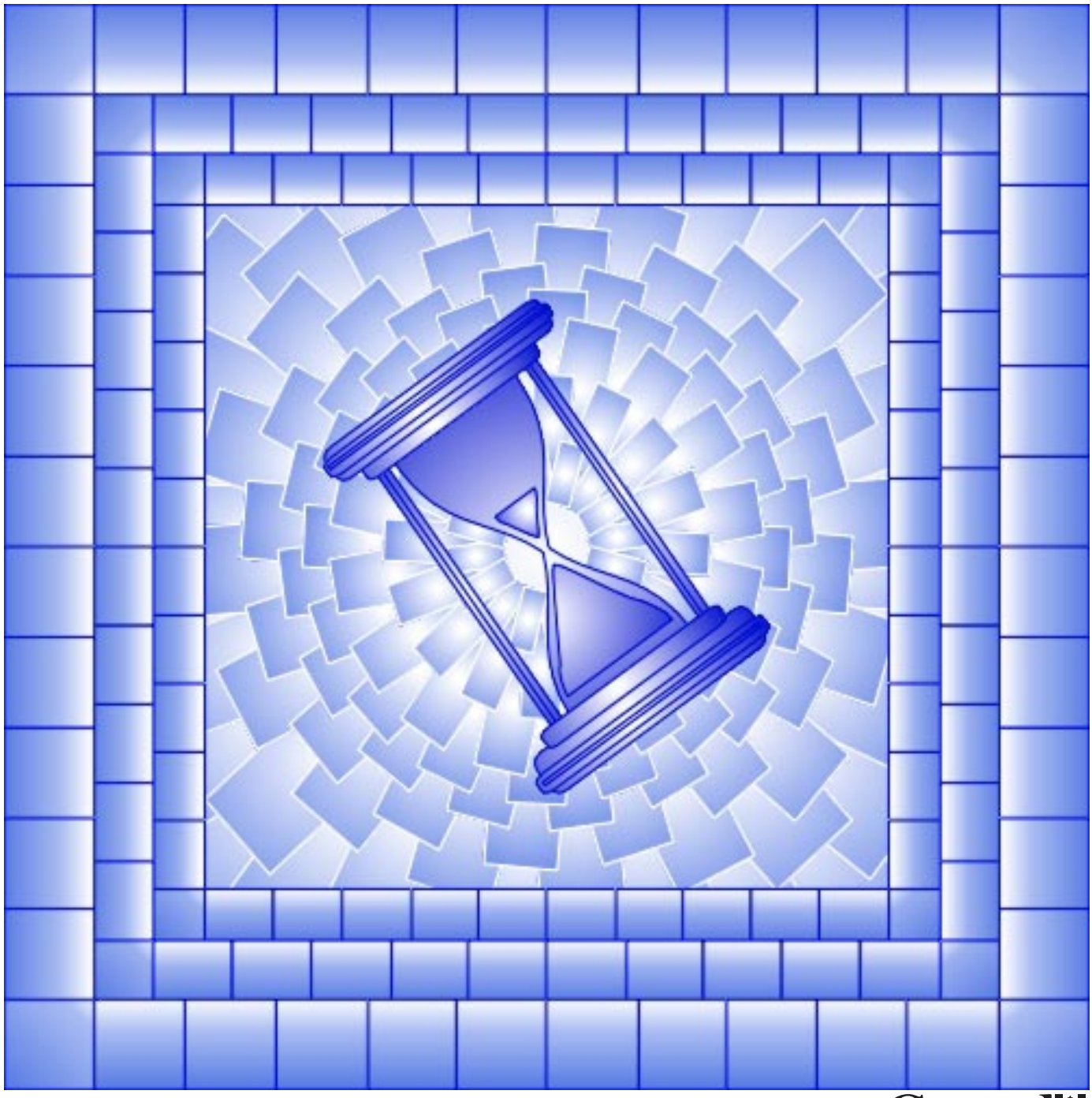
Analytical Series

Prices Division

Bias in the CPI: Experiences From Five OECD Countries

Edited by Louis Marc Ducharme

Nº 10



Statistique Canada
Statistics Canada
Canada

Bias in the CPI: Experiences From Five OECD Countries

Edited by Louis Marc Ducharme
Prices Division, Statistics Canada

Catalogue No. 62F0014MPB, No. 10
ISBN: 0-660-59374-2
ISSN: 1206-2677

13-A8, Jean Talon Building, Ottawa, K1A OT6

Facsimile Number: (613) 951-2848

Telephone: (613) 951-0688

Email: ducharl@statcan.ca

September 1997

Texte français au verso

Bias in the CPI: Experiences from Five OECD Countries

Introduction

Jacob Ryten[1]
Statistics Canada

The debate on measurement bias in the CPI, arising from the Boskin report, is not new and has been around for a number of decades.

But a number of things are new, which is why we decided to include under one cover a review of the experience of five different OECD countries. Several circumstances made the current experience special. Firstly, the rate of inflation has abated substantially in the more important industrialized countries and in some cases is steadily approaching zero. As a result, a bias that might have contributed a fraction of the total annual change in prices threatens today to swamp the entire change and more. Secondly, the experience of the change in prices of computers and their parts and accessories is so unique and so strong that it tends to overshadow everything else. To paraphrase an analogy provided in *The Economist*, if the efficiency of car engines had changed in the same proportions, today's cars could travel round the planet half a dozen times on a fraction of a dollar's worth of fuel. Thirdly, many central banks have chosen to use annual changes in the CPI as their key criterion for intervention in the money market. And lastly, for the first time ever, a blue ribbon commission dared give a number for the estimate of total bias and detailed each of the contributing factors.

These essentially new circumstances created a very special receptivity for the report. As a result, the OECD has listed a thorough discussion of the report and the way it was perceived in member countries as a key item to be discussed by one of its economic policy committees. While it is possible that the report's influence in the United States has waned, in a number of countries it raised the awareness of the need to review formulae, improve underlying household surveys (make them more frequent and more comprehensive) and adjust for quality variation all goods and services, especially complex goods and services typically offered in the market place as bundles.

This publication presents the experience and point of view of five different OECD countries relative to the issues raised in the Boskin report.

While most statistical agencies recognize that their CPIs are not perfect measures of inflation, some agencies of the OECD countries have consistently argued for a research agenda designed to improve their measurements and at the same time make them more comparable

[1] *Assistant Chief Statistician, Business and Trade Statistics*

with counterpart indices elsewhere in the OECD area. It is in this context, that Statistics Canada launched the Ottawa Group on Price Indices, an annual international seminar of like minded statistical agencies anxious to improve the quality of their respective price statistics.

But precisely because these statistical agencies have been so concerned about the CPI reliability, they cannot put their stamp of approval on any point estimate of the "total bias" of their CPI. In this respect Statistics Canada is very much at one with other comparable agencies — NOS in the UK, INSEE in France, CBS in the Netherlands, ABS in Australia to name a few — which have also come to the conclusion that there is too much that is not known for an estimate of bias, no matter how rough, to be advanced.

Along the same lines as others and in particular INSEE, Statistics Canada believes that estimates for particular biases (not all) are possible. In particular, these agencies agree that biases result from the treatment of price-induced commodity substitutions and from the use of improper micro-aggregation formulae. On the former, Statistics Canada, INSEE, NOS and ABS update their basket of consumer goods and services much more frequently than the United States does. On the latter, Canada was the first OECD country, followed by France, to introduce the use of geometric means for the aggregation of prices at the micro level.

The biases that result from the two features mentioned above are the only ones that can be estimated with some confidence and the statistical agencies contributing to this publication see no downside in engaging in inter-country comparisons based on them. But there seems to be a consensus amongst these agencies that biases resulting from the "improper or late treatment of the appearance of new goods," the "inadequate treatment of quality changes in observed products" and the "improper or late treatment of the appearance of new outlets," while important, do not lend themselves to any firm estimate given our present state of knowledge. There is neither sufficient theoretical agreement in the literature on these matters nor are the assumptions required to make estimates free of subjectivity or of controversy.

Accordingly, on the treatment of new goods, the four statistical agencies see no solid evidence that would lead them to modify their current practices. They all have proper procedures that allow them to introduce new goods in their price surveys without necessarily having to wait for the results of the latest consumer expenditure survey. The evidence available is unbalanced because the commodities appear to be selected more for their visibility than for their representativity. The technique of estimating "reservation" prices is a questionable one because the newly introduced goods prices are seldom the reflection of the behaviour of well informed consumers in a free and competitive market.

The five national statistical agencies agree that quality change is the most difficult and important problem in estimating the CPI. They also agree that quality changes are not one-sided. They also believe that the patterns of under- and over-estimation of quality dependent prices are subject to cyclical variations and are also sensitive to whether goods are traditional or in the midst of rapid technological change. Having said that, all feel that the only way to progress is by developing a research agenda and, with the help of other experts, to systematically analyze the effects of quality change on estimated prices.

Measuring Inflation in Australia[1]

Rob Edwards[2]
Australian Bureau of Statistics

Summary

The ABS currently publishes a wide range of separate consumer, producer and international trade price indexes, each relating to a particular segment of economic activity, as well as implicit price deflators and fixed weighted indexes derived from the national accounts.

These individual indexes can be considered as partial indicators as they each relate to a particular economic activity. Each index was developed to meet specific requirements and is released in its own separate, specialized publication, with substantial differences in profile. The Consumer Price Index is frequently used as a measure of inflation but it has a number of conceptual shortcomings for such purposes. In recent years, there has been increasing international attention directed towards developing new approaches to the measurement of inflation.

The purpose of this paper is to briefly outline the framework and current or future developments in the field of price statistics. The paper concludes that although no studies of bias have been undertaken in the Australian CPI, it is believed that any bias is likely to be small.

---

1 *This paper has been presented by the Australian Bureau of Statistics at the eighth meeting of Heads of National Statistical Offices of East Asian Countries, on Recent Issues on Statistical Activities, 20–23 May 1997.*

2 *First Assistant Statistician, Economic Accounts Division.*

Introduction

The ABS currently publishes a wide range of separate consumer, producer and international trade price indexes, each relating to a particular segment of economic activity, as well as implicit price deflators and fixed-weighted indexes derived from the national accounts.

These individual indexes can be considered as partial indicators as they each relate to a particular economic activity. Each index was developed to meet specific requirements and is released in its own separate, specialized publication, with substantial differences in profile.

Given that price measures are used for a wide variety of purposes including analysis of inflation, indexation, contract escalation, derivation of constant price value estimates and international comparisons, the selection of the most appropriate measure(s) for any given application is particularly important. The Consumer Price Index (CPI) is frequently used as a measure of inflation but it has a number of conceptual shortcomings for such purposes. In recent years, there has been increasing international attention directed towards developing new approaches to the measurement of inflation.

Given the differing principal objectives of the various individual price measures published by the ABS, there is scope to enhance the analytical value of the information through drawing the series together and presenting them as a system or family of price indexes. This system would be designed around a cohesive statistical framework developed specifically to support the study of inflation. The framework would also accommodate future developments in the field of price statistics such as extensions in coverage and presentation under alternative classifications. The ABS is presently in the process of developing such a framework.[2] The purpose of the present paper is to briefly outline the framework and current or future developments required to implement it.

Properties of an Analytical Framework for Inflation Measurement

Despite universal usage of the term "inflation," there is no generally agreed definition that is sufficiently precise to support the development of a statistical framework. The measurement of inflation is a complex issue and it is generally accepted that no single price index can measure all aspects of inflation. Nevertheless, there is some convergence of views among analysts as to the ideal conceptual properties that would be possessed by a system of price indexes designed for the analysis of inflation:

i. It would encompass only market transactions. That is, government services that are not marketed and notional transactions such as those where homeowners are deemed to rent dwellings from themselves as landlords, would not be included.

[2] *A discussion paper has been prepared on the framework* (ABS Information Paper: "An Analytical Framework for Price Indexes in Australia," *ABS Catalogue No. 6421.0).*

ii. It would capture the inflationary trend in prices associated with transactions in goods and services; accordingly, it would not include interest rates.

iii. Conceptually, the framework would embrace the entire economy and not be restricted to particular segments.

iv. It would provide for broad, economy-wide price indexes as well as component indexes.

v. It would relate to pure price change and incorporate very recent weighting information.

vi. From an analytic viewpoint, the effects of changes in government charges and taxes could be analysed, and the effects of erratic price fluctuations would be identifiable.

vii. The price indexes would be non-revisable, and provide certainty to users.

Two Possible Approaches

Two possible broad approaches for developing an economy wide system of price indexes are the national accounts approach and the market transactions approach.

The strength of the national accounts approach lies in the fact that the System of National Accounts (SNA) represents the only comprehensive and detailed framework for the systematic and integrated recording of the stocks and flows of an entire economy. As such, most users of economic statistics are familiar with its underlying principles and the major aggregates. The national accounts approach involves two sets of price measures: Implicit Price Deflators (IPDs) and Fixed-Weighted Price Indexes (FWIs), however, they possess a number of conceptual limitations.

IPDs do not compare the price of a constant basket of goods and services between any two periods because the weights change from period to period. As a result, period-to-period movements in the IPDs do not measure pure price change. FWIs provide a measure of pure price change and generally conform to several of the conceptually desirable properties for an economy-wide system of price measure. They do, however, include notional transactions, do not directly provide for the analysis of government taxes and charges, and have weights, which can remain fixed for long periods of time.

Alternatively, the market transaction approach represents a relatively new proposal being considered by a number of international experts as a means of overcoming the perceived deficiencies of the national accounts approach to the measurement of price change. The fundamental rationale for a market transactions approach is the view that inflation in an economy is a phenomenon peculiar to the operation of markets and thus price measures designed for the analysis of inflation should be confined to market activity; that is, non-market transactions (e.g. non-traded goods and services) have no role. This is the approach proposed by the ABS. Specifically, we are proposing the use of a Domestic Final Purchases model — see Diagram 1, which is appended.

The Domestic Final Purchases (DFP) Model

The scope of the price indexes under the DFP model is intended to reflect purchases by Australian residents. As such, it would include prices of imported items but exclude prices of exported items. While the price measures could potentially reflect the prices of *all* domestic purchases, a more meaningful aggregation would be one based only on *final* market purchases, excluding all intermediate purchases.

This approach recognizes that although many transformations take place in an economy (outputs of one firm become inputs of another), eventually a point can be identified where a good or service is exchanged in a marketplace for the last time. Therefore, if the price movements experienced by the final purchasers are recorded, it can be argued that the influences of the various price pressures experienced in earlier stages of the production process have been captured. Purchases of financial assets, existing capital goods (e.g. established housing), labour and intermediate inputs would not be within the scope of the DFP model. On the other hand, purchases of newly produced and imported capital goods, which are generally treated as final transactions within the national accounts, would be within scope.

The range of price indexes under the DFP model that could conceivably be produced include the economy-wide aggregate measure and the Price Index of Domestic Final Purchases, along with its component and sub-component indexes:

- Price Index of Current Consumption Purchases
- Price Index of Household Consumption Purchases
- Price Index of Non-Profit Institutions Serving Households (NPISH)
- Price Index of General Government Consumption Purchases
- Price Index of Capital Purchases
- Price Index of Household Capital Purchases
- Price Index of Private Corporate Capital Purchases
- Price Index of NPISH Capital Purchases
- Price Index of General Government Capital Purchases

(See diagram 2, which is appended.)

These component indexes could be disaggregated further to provide price indexes for key sub-components under alternative classifications, for example, by commodity (Household Purchases of Food, Private Corporate Purchases of Machinery and Equipment, etc.). Alternatively, the aggregate measure could be broken down by type of transaction (e.g. Consumption Purchases/Capital Purchases, Dwelling Purchases/Machinery and Equipment Purchases) and by institutional sector (e.g. Household Purchases, NPISH Purchases).

Conceptually, the scope of the DFP index is most closely aligned with Domestic Final Demand under the national accounting framework (the most notable difference being the exclusion of notional transactions), as opposed to Gross Domestic Product which involves the exclusion of imports and the inclusion of exports.

The characteristics that a DFP index, and its component indexes, would possess conform to the ideal properties identified earlier. The index would:

i. Relate to transactions in goods and services and thus not include interest rates (though in concept, interest rate margins which banks use as a substitute for direct charges for banking services would be included for households).

ii. Include actual market transactions only, that is, exclude changes in stocks, notional transactions, and non-marketed goods and services.

iii. Be based on the purchaser's perspective, with the valuation basis being at purchaser's prices.

iv. Have an economy-wide scope.

v. Include purchases by domestic residents only, and thus exclude exports and include imports.

vi. Include final transactions only.

vii. Reflect current expenditure weights using a chain index formula and adopt a pure price change approach.

viii. Contain component indexes presented under different classification systems.

ix. Form part of a broader statistical framework offering alternative, complementary views of the economy.

Future Development

Under the Domestic Final Purchases price index model, the ABS plans initially to concentrate on the development of a Price Index of Household Consumption Purchases (HCP) designed specifically for the analysis of inflation. There is a particular interest in measuring underlying inflation. Therefore the ABS will be looking at developing a "net price index" that removes the first order effects of tax change, at least. Further, it will be looking at trending techniques to provide a genuine underlying measure.

Since the scope of the DFP model is confined to final transactions, it cannot provide a complete picture of the price experience of the economy. Further, as some users are interested in price indexes that enable identification of price pressures arising from intermediate transactions, thus potentially identifying early inflationary signals, a range of Stage of Production (SOP) producer price indexes will also be incorporated into the model. The aim is to augment the analytical value of the current range of partial producer and international trade price indexes through their presentation in an economy-wide framework.

One complementary price measure already being developed by the ABS is the Labour Cost Index. Wage costs are a significant determinant of general price change, and so play a key role in any framework of price measures. A Labour Cost Index for general government would be required specifically for the DFP price index.

The ABS will continue to produce the CPI. Although widely used for other purposes, such as a measure of inflation, the Consumer Price Index has been designed primarily as a means of assessing changes in the purchasing power of wage and salary earner household incomes and, as such, has played an important role in the income adjustment process. In recent years, there has been recognition of conceptual deficiencies in the CPI for inflation measurement, particularly for policy purposes. This has led to the development of alternative measures such as the index of underlying inflation as defined by the Australian Government Treasury. The forthcoming periodic review of the CPI will provide an opportunity for, amongst other things, a review of the principal purpose of the index.

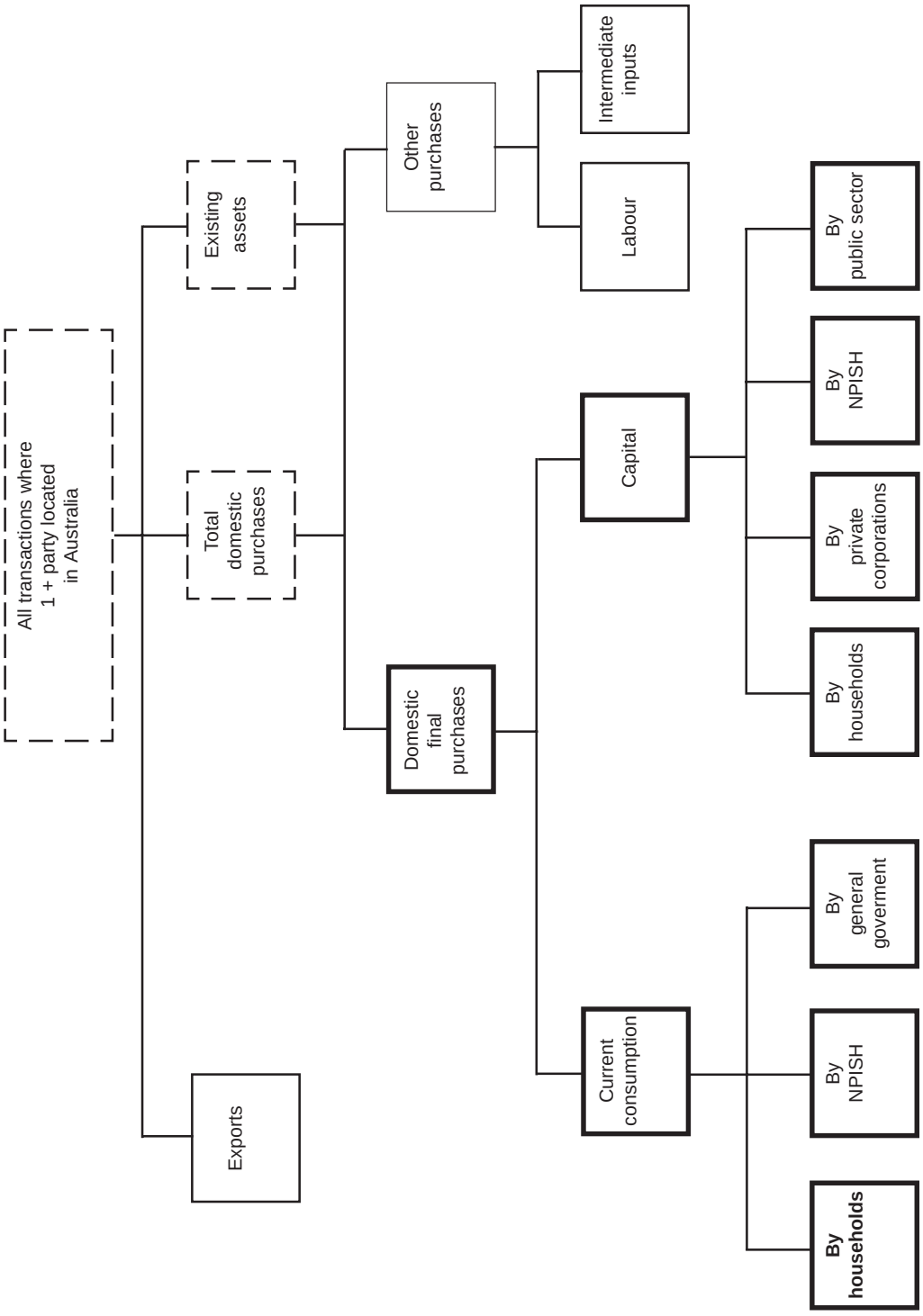
As there is a huge public interest in price measures, it is important that these changes are only implemented after extensive public consultation and explanation of the changes that are being made. The ABS has started this process through the publication of the discussion paper mentioned earlier and a paper outlining the issues for consideration in the forthcoming review of the Consumer Price Index[3] . The ABS is also holding a series of public seminars and bilateral discussions with the key users.

One issue of particular interest is CPI Bias, particularly following the publicity given to the report of the Boskin Commission in the United States of America. Studies of bias in the CPIs of other countries have been undertaken.

No studies of bias have been undertaken on the Australian CPI. We believe, however, that the similarities of our approach to CPI compilation with those of Canada and the United Kingdom would suggest that any bias in the Australian CPI is likely to be small. Nevertheless, there are some changes we can make to reduce the risk of bias, notably the reduction of lower level substitution bias and the extension of quality adjustment to important services included in the CPI.

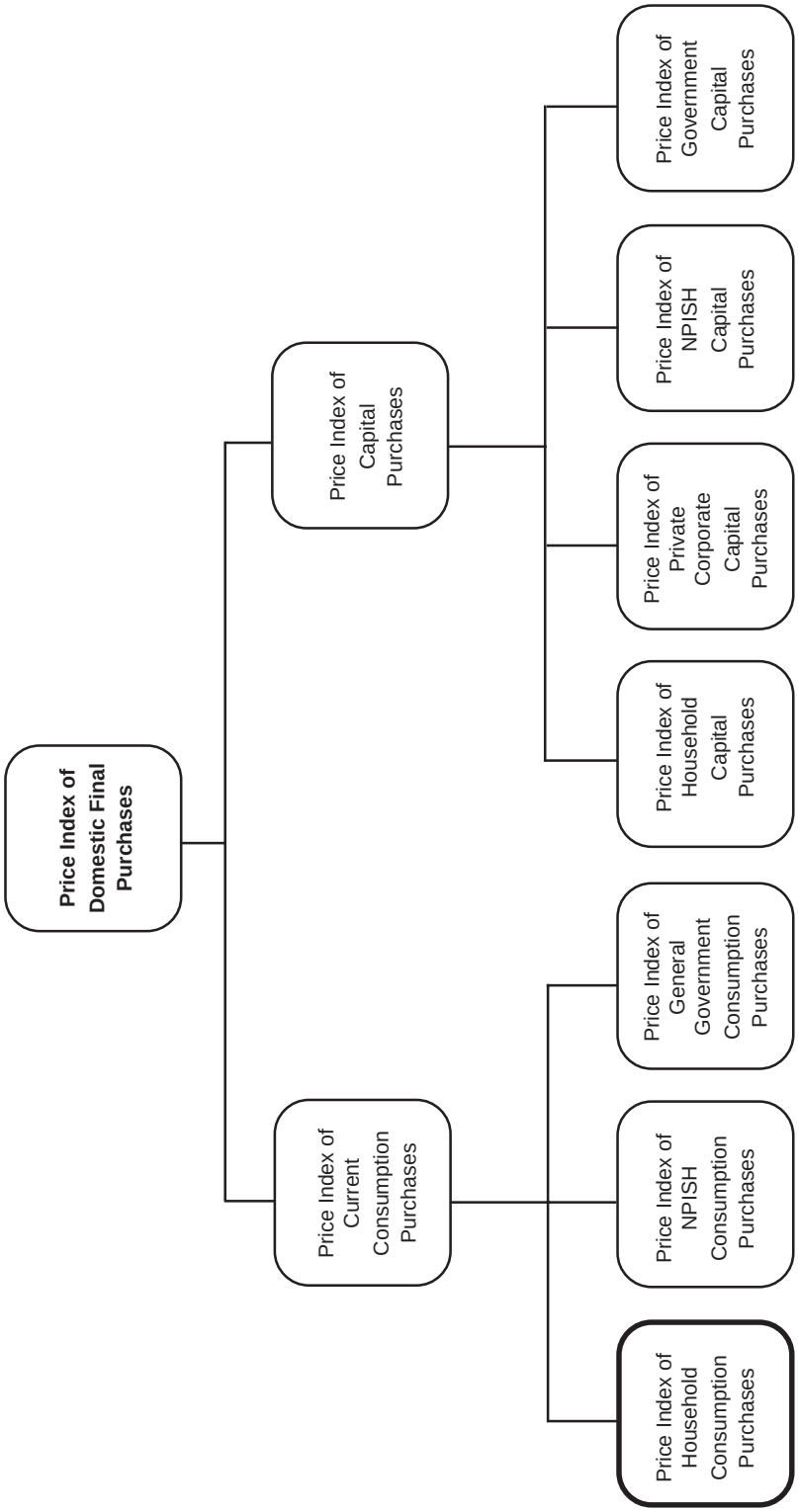
[3] *This discussion paper is due to be released at the end of April.*

DIAGRAM 1
Market Transactions View of the Economy



From: ABS Information Paper: "An Analytical Framework for Price Indexes in Australia," ABS Catalogue No. 6421.0

DIAGRAM 2
Price Index of Domestic Final Purchases



From: ABS Information Paper: "An Analytical Framework for Price Indexes in Australia," ABS Catalogue No. 6421.0

The Canadian Consumer Price Index and the Bias Issue: Present and Future Outlooks[1]

Louis Marc Ducharme[2]
Statistics Canada

Summary

The debate on the measurement of bias in the CPI has been around for decades. However, given the size of government budgetary deficits, the issue of overestimating inflation and therefore payments in social benefits has triggered the interest in the measurement of the CPI bias. The final report of the U.S. Advisory Commission to Study the Consumer Price Index, chaired by Michael Boskin, states that the U.S. CPI has been overestimated by 1.1% per year since 1996. Following the release of the report, many interested groups have asked the question as to the magnitude of the bias for Canada's CPI. This result raised the question whether the bias in the Canadian CPI was of the same magnitude.

This paper begins by presenting the bias issue in the context of the Canadian CPI and then outlines some of the plans Statistics Canada intends to undertake in the near future to improve the measurement of the CPI. The paper concludes that, although the Canadian CPI may suffer from the same potential problems as the U.S. CPI, the overall effect of these biases is less notable because Statistics Canada started to apply an appropriate methodology earlier. In fact, in recent studies, Crawford (1993 and 1997) tried to estimate an overall bias and concluded that given the generous judgement incorporated in the estimate, it is likely that the bias is, on average, smaller than 0.5%.

[1] This is a shorter version of a paper to be published in *Revista Española*, Fall 1997.

[2] *Director, Prices Division. The author would like to thank Jim Clarke, Robin Lowe, Margaret Parlor, Marc Prud'homme, Jacob Ryten and Bohdan J. Schultz for their comments and suggestions which significantly improved the final version of the paper.*

Introduction

The debate on the measurement of bias in the Consumer Price Index (CPI), arising from the Boskin report, is not new. In fact, it has been around for many decades; for instance, some economists and statisticians may remember that in the late seventies, when inflation was running high, the CPI was under scrutiny for underestimating the true cost of living. Conversely, at the time of the Stigler Commission (1961), there was a comparable concern about the overestimation of the cost of living.

In its final report, the U.S. Advisory Commission to Study the Consumer Price Index, better known as the Boskin Commission, stated that the CPI has been overestimated by 1.1 percentage points per year.[3] Given the widespread use of the CPI in the indexation of social and transfer payments, as well as for the escalation of private contracts and alimonies, this conclusion raised serious concerns in Canada and in other major OECD countries. In Canada, we have long been well aware of the potential biases in the calculation of the CPI and for this reason Statistics Canada has invested over the years in research programs and on-going improvements to minimize their effects. Yet, like several other statistical agencies such as INSEE (France), NSO (UK) and ABS (Australia), Statistics Canada recognized that it is difficult to make conjectures about the overall bias.

The main objective of this paper, as well as the other papers in this publication, is not to judge the estimations or conclusions of the Boskin Commission. Many replies have already been published on the subject.[4] Rather, the purpose is to see how relevant some of the findings are for Canada, how we are currently trying to solve these problems and what we intend to do with them in the future. The paper presents, for five type of bias,[5] a brief description of the problem and the solution actually used in the current calculation of the CPI. It ends with a section on some of the future outlooks facing the Canadian CPI.

---

[3] *The estimate is 1.3% per year before 1996 and 1.1% after 1996, following corrections made by the Bureau of Labor Statistics (BLS) to the calculation of the CPI.*

[4] *Abraham (1997), Ehrlich (1997), Norwood (1997) and Hulten (1997) to name only a few.*

[5] *The bias can be grouped into five categories: i) product substitution, ii) new goods, iii) outlet substitution, iv) quality change and v) micro-formula.*

Product Substitution Bias[6]

The Canadian CPI, like most other CPIs produced by statistical agencies, measures the price change of a fixed basket of goods and services commonly purchased by consumers during a given period. In other words, it measures the change in price of those goods and services purchased by consumers, while trying to keep their quantity and quality constant.

A first potential source of bias arises from the substitution of a product or service for another (this is called the product substitution bias). This type of bias arises because consumers tend to adjust the structure of their purchases to change in relative prices so that price increases have less effect on their situation (level of satisfaction). For example, they may substitute chicken for beef when the price of beef rises relative to chicken. The CPI, however, as it is commonly calculated, cannot fully reflect these savings due to substitution. Indeed, it is calculated as an average price change of various groups of goods and services that form the CPI basket and are referred to as basic commodity classes. To avoid having changes in consumption affect the index, the quantitative proportions between the basic classes are kept fixed from one CPI basket update to another.

In theory, there are two ways of taking into account these fixed proportions. Either we fix the proportions of chicken and beef consumed in a base period (Laspeyres-type index) or in the current period (Paasche-type index). In practice, index makers are forced to derive the fixed proportions from past consumption patterns because current expenditure data are not available. The rigidity imposed by the fixed proportions ignores savings that may have occurred when consumers substituted chicken for beef and using past consumption gives more importance to the items, such as beef, that experienced the largest increase, while their importance in the consumer purchases have decreased. For this reason, the Laspeyres-type index tends to overestimate the consumer price change.[7]

The bias could be avoided by using the so-called Fisher-ideal index ,[8] but this is impractical because current expenditure data are not available to ongoing index production. Instead, baskets are updated only periodically, and the indices are associated with subsequent baskets by linking to each other. The resulting Chain-Laspeyres index is still affected by the product substitution bias, but its size depends on the frequency of basket updates.

The more frequent the update the smaller the bias. Frequent updates minimize the bias from longer-term substitution effects, for example, shifts in food consumption habits, such as butter for margarine. Until recently, the American practice was to update the basket of goods and services every **ten** years, ours is usually updated every **four** years, the French **every** year.[9]

---

6 *This type of bias is also referred to as the macro formula bias.*

7 *Conversely, the Paasche-type index tends to underestimate the consumer price change.*

8 *The Fisher price index is the geometric average of the Laspeyres and Paasche indices, the latter being based on the current-period basket.*

9 *In France, the weights at the level of "poste" (the French equivalent level of the Canadian basic classes) are updated every year using data derived from the National Accounts. For more details, see Lequiller's paper in this publication.*

In addition to more frequent basket updates, Généreux (1983) and Bérubé (1996) showed that the substitution effect could also be reduced by introducing new weights more rapidly. In Canada, weights have been normally introduced three years after the reference period. According to Bérubé's estimation this would lead to an average annual substitution bias of 0.20% over the entire period (1962 to 1994). With the introduction of the 1996 basket change, in January 1998, the delay will be reduced to two years.[10] Again according to Bérubé's calculation, this would reduce the annual substitution bias to less than 0.20%.

The consequence of a given product substitution is often only temporary, hence the overall effect of this substitution bias on the CPI is normally quite limited. This is because of the dynamic nature of the market, which re-adjusts itself to changes in consumption. For example, the demand changes caused by the substitution of chicken for beef will tend to increase prices of chicken in the subsequent periods, which in turn will diminish the effect of initial substitution savings. Several studies, Canadian and other, confirm this conclusion.

New Goods Bias

The second source of potential product substitution bias is due to the fact that new goods appearing on the market are not well taken into account because of the rigidity of the basket of goods and services between two update periods. That is, for each basket the price index makers fixed the proportion of expenditures between the different types of goods (i.e. proportion of consumer expenditures for chicken compare to beef) and also the type of goods that will represent a category (e.g. Kellogg's Corn Flakes to represent "cereals"). This type of substitution bias is better known as the "new goods" bias.

This bias arises because: i) new products or services appear on the market incorporating new technology which allows their prices to decrease rapidly (at least relatively) in the early stage, ii) the prices of other more traditional products decrease as the volume of sales rises and the unit costs of production go down, or iii) new brands of existing products are introduced.

In the past decades, we have witnessed the emergence of many new products and services, especially the new "high tech" products such as personal computers, video games and cellular telephones. For most of these products the price level at the moment of their introduction on the market was relatively high. Then standardization and optimization of the production process allowed for decreases in production cost, and therefore in prices, which in turn led to increases in sales and accelerated the rate of price decline.

[10] *One should note that even for those countries that update their weights annually there is still a time lag of up to two years between the basket reference year and the year of introduction to the CPI.*

If these products are introduced into the CPI with delays and sometimes not until their prices stop declining, it may introduce a bias in the calculation of the CPI. However, since the share of these products in the basket of goods and services purchased by consumers at that time is usually very small, the impact of late introduction[11] on the total CPI is probably also small.

In the Canadian CPI, the delay is not related to the frequency of the basket updating as new products are, in fact, introduced at any time between basket changes, that is, as soon as we are able to collect their prices and as soon as their penetration rate in the household market is substantial. This practice tends to minimize the potential bias that may arise from the introduction of new goods. However, given that the evidence available is unbalanced (commodities chosen in these special studies appear to have been selected for their visibility rather than for their representativity), we are not able to quantify that bias.

A variant of the new goods bias is the bias due to the late introduction of new brands of existing products, hereafter called the "new brand" bias. This bias would occur when the number of new brands increases on the market, giving a larger choice to the consumer and thereby reducing his or her cost of reaching a given level of utility. In order to estimate the effect of the increased brand selection on the cost of living, one can calculate the direct effect of a new brand on the consumer surplus by estimating the corresponding area under the demand curve for the new good between the reservation price and the observed price. Very few empirical studies have addressed this issue and their results are far from conclusive. The technique of estimating a reservation price is questionable because early prices are seldom the reflection of what a consumer would do in a competitive market.[12] For these reasons, statistical agencies do not make direct adjustments for the increased number of brands. Indirectly, some adjustments are made each time the sample is changed to take into account the introduction of new products or varieties.[13]

---

[11] *By late introduction, we do not mean decades later or only at the time of basket change, but when the share in the consumer purchases becomes significant.*

[12] *A more serious objection is that if you compute consumer surplus, you have to do it for all, including losses for items that disappear.*

[13] *In a recent study on coffee using scanner data, Scobie (1996) compared the approach currently used in the CPI to calculate the coffee index with a "scanner coffee index" incorporating more than 526 UPCs (Universal Product Codes) representing a very large spectrum of coffee. His results showed that the "scanner coffee index" was similar to the official coffee index in the Canadian CPI. This indicated that the increased number of varieties did not affect the index much at the aggregate level. In another study on new goods, Scobie (1997) also showed that the price change of new and old products can differ, but they tend to move similarly over time.*

Outlet Substitution Bias

Another source of bias is the substitution of outlets. This source of bias is of a similar nature to the new goods bias, but related to the outlet. It arises when a new outlet or a new type of outlet (e.g. superstore or warehouse type of outlet) arrives in a specific market with lower prices, and no decline in prices is captured in the price index.[14] Here the implicit assumption is that the quality of the new outlet is similar to the existing (traditional) outlets. Is it true? To our knowledge no evaluations of the differences in level of services between the traditional and the new discount-type of outlet have been made.

In Canada, the sample of retail outlets for price collection is not random but judgmental, which means that we pick the largest outlets (those that have the highest volume of sales) to represent the universe of stores in which Canadians buy their goods and services. This system is in principle more flexible than a statistical random sample, as we can introduce new outlets at any time and not just when the complete outlet sample is revised.[15] Because we cover the largest stores and we monitor shifts in the market share of these stores, we reduce the potential bias due to the shift in consumers' buying patterns. For example, when a new type of store appears on the market (e.g. Wal-Mart) and gains a significant portion of the market, we start monitoring price movements in that store. Besides, it is most probable that the price movements in the competing outlets have changed, reflecting, therefore, the impact of the introduction of the new outlet.[16]

Quality Bias

Since the CPI measures price changes of goods and services of constant quality, prices must be adjusted to account for quality changes. Each month the price statisticians have to deal with many discontinuities in the products whose prices are collected. When a product disappears from the sample, it should be replaced by a product of similar quantity and quality. If the replacing item is not of the same quantity or quality as the replaced one, adjustments have to be made to the price. Therefore, a measurement bias may arise if the adjustment method overstates or understates the portion of the price change associated with the quality change.

---

[14] *It is important to note that not only must we take into account the price decline generated by this new outlet, but also those of its competitors, so that when the new outlet is linked in the index, the difference in price movement is not lost.*

[15] *In the American CPI the choice of outlets for price collection is derived from a statistical random sample. This means that they select a sample of outlets that represent the total universe of stores in which Americans normally buy their goods and services. The criticism of this procedure has been that, while the sample is representative at the time of selection, it may miss a shift in the consumers' buying patterns (e.g. shift from full-price food retailers to grocery wholesalers) between rotations. This shortcoming has already been taken care of in the United States as a new outlet can be introduced between sample rotations.*

[16] *Two pilot studies will be launched in the fall of 1997 to estimate the impact of the introduction of new outlets as well as the difference in quality of services between the traditional and the new discount-type outlet (e.g. warehouse and superstore).*

In Canada, as in most major OECD countries, three methods are used to make the quality adjustment: i) linking, ii) direct comparison, and iii) explicit adjustments.

The first method, also called "splicing," is the simplest and likely the most often used in practice. This method implies that the differences in price between the new product and the old one represents the effect of change in quality.[17] The second method consists of finding the most suitable replacement for the product that has disappeared from the shelves. In this case, the price collector must find, according to established specifications, another product that has similar characteristics. Since both products are assumed to be of equal quality, the difference in prices is in this case normally considered to be pure price change.

The last method implies explicit adjustments to correct for quality changes. A number of procedures are used in Canada. For most durable goods (for example, electrical appliances) the additional producer cost of the new features is estimated by the price statistician, using information from the industry, and deducted from the price. Another way of estimating the change in quality is to compare prices of optional features and adjust the resulting price change. For example, in the case of automobiles, an evaluation of this type of quality change is done each time a new model is introduced. Cosmetic changes to a product are generally not considered as quality change and no adjustment is made for them.

For clothing, the current practice is to monitor changes for the key quality characteristics, which are important in price setting.[18] Each time there is a quality difference in the product between two periods, the price collectors identify and judge the difference in quality according to these key characteristics. Finally, in the case of microcomputers, econometric methods, better known as hedonic methods, are used. The assumption behind this method is similar to the previous one (i.e. for clothing), but rather than have price collectors judge how the change in quality affects price, change is determined by an econometric equation using shadow prices of key characteristics.

As one can imagine, the explicit methods of adjustments are complex, difficult and costly to apply and that is why they have not been used extensively in Canada, Australia, France or the United States.[19]

There is a prevailing opinion that the quality adjustment techniques tend to underestimate the quality change thus introducing an upward bias in the CPI. It is likely true for the "high tech" products, but evidence of a downward bias also exists for some other products. Schultz

---

[17] *It is assumed that for one period (t) the two products are present on the market. Splicing means that the price movement of the old product (between* t-*1 and* t*) is linked to the price movement of the new product (between* t *and* t+*1).*

[18] *For example, in the case of shirts, the thread-count of the material for a shirt, the type of material (cotton, polyester), the type of collar, long or short sleeves, etc.*

[19] *For instance, in the United States, hedonic techniques are extensively used for clothing and computers.*

(1995) has shown that, contrary to popular belief, the quality adjustments for a selection of clothing articles have led in the long run to an underestimation of price increases. He claims that the patterns of under- and over-estimation of quality-dependent prices are subject to cyclical variations and are also sensitive to whether goods are traditional or in the midst of rapid technological change. Given that this is one of the areas representing the most difficult and important problem in the CPI estimation, it is important to have a continuing review of our quality adjustments to ensure that the resulting biases are not significant. Also, we feel that the only way to make progress in this area is by developing a research agenda and by systematically analysing the effect of quality changes on surveyed prices.

Formula Bias

The CPI bias may also result from the use of specific methods to calculate price changes at the lowest aggregation level. For example, the arithmetic means of price relatives for aggregating prices at the lowest level are upward biased.

Canada has been spared from this type of error since 1978. Indeed, since that time the aggregation at the lowest level has been done using ratios of arithmetic mean prices from matched samples of outlets and items, while the arithmetic means of price relatives partially in use before were dropped.[20]

The ratios of arithmetic mean prices tend to provide unbiased results although they may over-emphasize the influence of more expensive items. Since January 1995, the ratios of geometric mean prices are used instead. They have several advantages. One is that they take into consideration substitution between products, while the arithmetic mean assumes fixed weights (Lequiller 1997). Another advantage is that they give less importance to high priced items in the sample than to the corresponding arithmetic mean. In the Canadian context where samples are small and judgmental, the use of geometric means should reduce the risk of these items creating strong fluctuations of the index.[21]

Total Bias

As stated earlier, statistical agencies, such as Statistics Canada, ABS or INSEE, have been always very concerned about the reliability of the CPI. Although we have recognized the problem of potential biases in its calculation, we have also recognized the difficulties (if not

[20] *Upon Bohdan Schultz's advice, Statistics Canada abandoned, in 1978, the arithmetic mean of price relatives, which was upward biased, and started using the ratio of arithmetic mean prices for which there is no substantial bias.*

[21] *Recent results from an empirical study conducted by Schultz (1994) showed that no large systematic difference should be expected between index numbers obtained using ratios of the geometric mean prices and ratios of arithmetic mean prices. However, there can be a big difference between ratios of arithmetic mean prices and the means of price relatives.*

the impossibility) of giving a point estimate of the total bias. In the previous section of this paper we described and reviewed some of the actual practices used to construct the Canadian CPI in the context of the debate on biases. It appears that because of the foresight of its price statisticians, the Canadian CPI does not suffer seriously from these biases.

Having said that, Statistics Canada believes that it is possible to assess particular types of bias, although not all, and in this respect considers the work of Crawford (1993 and 1997) as a careful and balanced overview of the problem. We agree, in particular, with Crawford's opinion about the biases that result from the treatment of price-induced commodity substitution and from the use of improper micro-aggregation formulae.[22] The biases that result from these two features are the only ones that can be estimated with some confidence and Statistics Canada welcomes inter-country comparisons based on them.

As for biases due to other causes, Statistics Canada does not consider there to be enough evidence to support any precise numerical assessment of their overall effect on the CPI. This does not mean, of course, that they have no effect. The following table shows the estimates made by Crawford (1993 and 1997).

Table 1
Comparison of Bias in the CPI of Three OECD Countries
(in percentage of annual CPI change)

Type of Bias	Canada[1]	France[2]	United States[3]
Product substitution at the aggregate level	0.10–0.20	..	0.15
Product substitution at the intermediate level	0.00–0.10	0.05–0.10	0.25
Product substitution at the detailed level		..	
Outlet substitution	0.00–0.10	0.05–0.15	0.10
Formula	..	..	..
Total (excluding new goods)	0.10–0.40	0.10–0.25	0.50
New goods and quality/change	?[4]	?	0.60
Total (including new goods)	?	?	1.10

.. not applicable.
[1] Based in part on Crawford (1993 and 1997), the upper limit is on average 0.50.
[2] Lequiller (1997).
[3] Estimates for 1997, after changes to the U.S. CPI were made.
[4] Cannot be estimated.

[22] *See the work by Généreux (1983), Bérubé (1996), Szulc (Schultz) (1983,1987,1994) mentioned previously in the paper.*

Future outlook

It has been shown that Statistics Canada, like other statistical agencies, has taken the issues on bias very seriously in the past and intends to do so in the future. A number of new initiatives are about to be carried out by the agencies in order to improve the calculation of the CPI and reduce its potential bias.

The first initiative is to reduce the delay in the introduction of the CPI baskets. For the first time, the new basket weights (1996) will be introduced only two years after the reference period. As was mentioned earlier, this will reduce the product substitution bias. A further step will be to implement annual basket updates. With the introduction in 1997 of an annual Family Expenditures Survey (FAMEX), it will now be possible to make yearly changes in the weights of basic classes. This will further contribute to the reduction of the potential product substitution bias.

In addition to these two initiatives, the standardization of scanner data, based on the extensive use of Universal Product Code (UPC), will provide opportunities for Statistics Canada to correct the product substitution bias at the lowest level of detail. Information on prices and quantity for thousands and thousands of products in hundreds of outlets will also allow statistical agencies to explore the issues of the new goods and outlet substitution biases. Analytical work has already started on the issue of new goods in Canada and in several other countries.[23] There is also great potential to examine the issue of outlet substitution bias.

With the increasing number of new goods introduced on the market, the problem for price statisticians is to get advance notice of their existence in order to be able to introduce them early enough in the product cycle. Here again, cross-fertilization from other fields of research can contribute to further progress. The new "generation of innovation surveys" developed by Statistics Canada could be used to track down the first appearance of a new product or service on the market. This source of data will complement the rich information already available from scanner data and the more traditional sources.[24]

The outlet substitution bias for the Canadian CPI is probably very small. However, the fact that our judgmental sample concentrates on only the largest outlets should not be considered satisfactory. Two empirical studies will be launched in the fall of 1997 to establish the impact of the introduction of new grocery outlets. In order to reduce the gap between their appearance on the market and the moment they are introduced in the sample, a procedure will be implemented.

---

[23] *A number of countries have been exploring the potential of scanner data using selected databases provided free of charge by A.C. Nielsen Marketing Research. See the work of Prime and Saglio (1995) for France, Scobie (1996, 1997) for Canada and Reinsdorf (1995) for the United States.*

[24] *Such as information provided by manufacturers, industrial associations and retailers.*

The next step could be to introduce a proportional probability sample of outlets in some areas to allow for a better representation of smaller shops. Until now consumers have been making most of their food purchases in large supermarkets. However, there is a "renaissance" of speciality shops such as, fresh fruit and vegetable outlets, bakeries, butchers and delicatessens, etc. There is no reason for the price movements of these shops to be different than those of the larger ones. The shift toward proportional probability sampling would address this issue.

A number of new initiatives have been taken by Statistics Canada on the issue of quality change. As outlined earlier, this is one of the most difficult issues. These initiatives are managed on two fronts. In 1996, a program aimed at extending the use of hedonic techniques was launched. The first area to be examined is clothing. Later, other goods which are not characterized by rapid technological change will be first included in the feasibility studies and then into the actual index production. The major constraint here is the cost of this technique. In the case of clothing, the current database allows the development of hedonic models. This is not so for all the other goods, for which larger price samples could be needed.

On the second front, research has started on the occurrence and type of quality adjustments made to the CPI. In a recent study, Lowe (1997) showed the need to make adjustments for commodities purchased occasionally. He reinforced the importance of using regression methods for commodities with clear characteristics and rapid turnover, such as clothing. He also argues that outlet bias and new products (both genuinely new products and new varieties) would be mitigated by larger and more diverse samples. This research fits in with a more general effort to describe and review systematically the procedures used for quality changes, as well as to diversify the variety of prices collected within most of the basic classes.[25]

Finally, progress could not be accomplished quickly enough without the pooling of research findings of other major statistical agencies. It is in that context that Statistics Canada launched, in 1994, the International Working Group on Price Indices (the Ottawa Group), an annual workshop attended by a dozen representatives of statistical agencies, with the objective of improving the methodology and quality of their respective price indices. The agency is also engaging in bilateral co-operation with other statistical agencies on specific topics.[26] Sharing the concerns, the workload and the results between members of the international community can only be beneficial to all.

[25] *As part of the ongoing review of the price sample, an effort needs to be made to diversify the variety of commodities in the sample in order to allow for a wider representation of price movement in the CPI.*

[26] *A two-year co-operative program with INSEE on price statistics for services will begin in the fall of 1997.*

Bibliography

Aizcorbe, A. and P. Jackman, (1993) "The Commodity Substitution Effect in CPI Data, 1982–1992," Monthly *Labour Review,* December.

Armknecht, P. and D. Weyback, (1989) "Adjustment for Quality Change in the U.S. Consumer Price Index," *Journal of Official Statistics,* vol.5, no.2, pp.107-123

Bérubé, C., (1996) *Selecting a Formula for the Canadian CPI, 1962–1994*, Prices Division Analytical Series, No. 7, Catalogue No. 62F0014MPB, 7 p.

Boskin, M., E. Dulberger, Z. Griliches, R. Gordon and D. Jorgenson, (1996) *Toward a More Accurate Measure of the Cost of Living*, Final report of the Advisory Commission to Study the Consumer Price Index to Senate Finance Committee, December.

Crawford, A., (1993) *Measurement Biases in the Canadian CPI*, Bank of Canada, Technical Report No. 64, September, 54 p.

Crawford, A., (1997) *Measurement Biases in the Canadian CPI*, Paper presented at the Bank of Canada Conference, Ottawa, May.

Diewert, E., (1996) *Comment on CPI Biases*, UBC Discussion Paper No. 96-07, February, 14 p.

Ehrlich, E., (1997) "The Downside of Bad Data," *Challenge*, March–April, pp. 13-37.

Généreux, P., (1983) "Impact of the Choice of Formulae on Canadian Consumer Price Index," in Diewert, W.E. and C. Montmarquette, (eds.) *Price Level Measurement*, Proceeding from a Conference Sponsored by Statistics Canada, Ottawa, December, pp. 489-512.

Gordon, R., (1990) *The Measurement of Durable Goods Prices*, University of Chicago Press, 714 p.

Gordon, R., ed., (1996) *The Economics of New Goods*, University of Chicago Press.

Hulten, C., (1997) "Quality Change in the CPI," *Challenge*, March-April, pp. 48-74

Lequiller, F. (1997) "L'indice des prix à la consommation surrstime-t-il l'inflation?" *Économie et Statistique*, no. 30, pp. 3-32

Lowe, R., (1997) "Item Selection and Quality Change," Paper presented at the Third Meeting of the Ottawa Group, Voorburg, April, 12 p.

Moulton, B., (1996) "Bias in the Consumer Price Index: What Is the Evidence?" *Journal of Economic Perspectives*, vol. 10, no. 4, Fall, pp. 159-177

Norwood, J., (1997) "How Right is the Boskin Commission?" *Challenge*, March–April, pp. 38-47.

Szulc, B.J., (1983) "Linking Price Index Numbers" in Diewert, W.E. and C. Montmarquette, C. (eds.), *The Measurement of Price Level*, Proceedings from a Conference sponsored by Statistics Canada, pp. 337-357.

Szulc, B.J., (1987) "Price Indices Below the Basic Aggregation Level," *Bulletin of Labour Statistics*, vol.2, Geneva, pp.ix-xvi.

Schultz, B. J. (Szulc), (1994) *Choice of Price Index Formula at the Micro-aggregation Level: The Canadian Empirical Evidence*, Paper presented at the First Meeting of the International Working Group on Price Indices, Ottawa, November, pp. 93-127

Schultz, B.J. (Szulc), (1995) *Treatment of Changes in Product Quality in Consumer Price Indices*, Paper presented at the Second Meeting of the Ottawa Group, Stockholm, November.

Scobie, H., (1996) *A Study of Scanner Data: Coffee*, Prices Division Analytical Series, No. 11, Catalogue No. 62F0014MPB

Scobie, H., (1997) *Use of Scanner Data—The Impact of New Goods*, Prices Division Analytical Series, No. 12, Catalogue No. 62F0014MPB

Shapiro, M. and D. Wilcox, (1996) *Mismeasurement in the Consumer Price Index: an Evaluation*, NBER, Working Paper no. 5590, May.

Statistics Canada, (1995) *The Consumer Price Index Reference Paper: Update Based on 1992 Expenditures*, Catalogue No. 62-553, Ottawa, 126p.

Statistics Canada, (1996) *Your Guide to the Consumer Price Index*, Catalogue No. 62-557-XPB, 15 p.

Wynne, M. and F. Sygalla, (1994) "The Consumer Price Index," *Economic Review*, *Federal Reserve Bank of Dallas*, 2nd quarter.

Does the French Consumer Price Index Overstate Inflation?[1]

François Lequiller[2]
INSEE

Abstract

The debate over problems in measuring inflation is not new. It has recently been revived by the publication of a report by an Advisory Commission to the U.S. Senate. The Commission, chaired by Michael J. Boskin, found that the U.S. Consumer Price Index (CPI) overestimates inflation by 1.1 percentage points a year. This article shows that the potential bias in the French CPI is on a far lower order of magnitude.

It is hard to summarize the changes in a multitude of prices with a single figure. Even in the best-case theoretical scenario — a single consumer faced with a spending decision — the treatment of substitutions between existing products raises important problems. Nevertheless, it is possible to provide a fairly accurate description of the various possible alternatives and the statistical procedures used in France largely shield the country's index from criticism on this point.

The introduction of new products creates serious difficulties that have not been entirely resolved in the United States, in France, or elsewhere: "new products" is used here in the broad sense to denote (1) genuinely new products on the market and (2) products already sold elsewhere but introduced in a new sales outlet, replacing existing products or not. The Boskin Commission estimates the upward bias in the U.S. CPI due to new products at 0.6 percentage points per year. The Commission's claim rests on fragile and probably exaggerated estimates. Our conclusion converges with the opinion of several U.S. statisticians.

---

1 *This is a shorter version of a paper published in* Économie et Statistique*, no. 300, 1997.*

2 *At the time of writing, Francois 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.*
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), but also to his successor, T. Lacroix, L. Viglino, D. Temam, M. Glaude, and G. Laroque. Special thanks to J. Mandelbaum for a magnificent translation.*

Introduction

Spurred by a longstanding debate over the possibility of a consistent 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 it expected the U.S. Consumer Price Index (CPI) to 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, would be no more than 1.9%.

The French CPI is unquestionably one of the most closely watched of INSEE's economic indicators. It is the main gauge of inflationary pressures in monetary and fiscal policy, and supports 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 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.

For all these reasons, the determination of the CPI—in France as in the United States—rests on the strict application of standardized procedures, and INSEE, like BLS, devotes substantial resources to the task. It is no easy task to determine an indicator that summarizes in a single figure the changes in prices of the billions of household purchases. The process raises problems, such as the coverage, representativeness, and accuracy of the index (sampling) and the recording of actual transaction prices (followed by that of unlisted discounts).

In the first part of this article we will study the problems of substitution between products. We will review the weighting and the calculation formulae utilized as solutions. We conclude that the statistical methods used in France differ from those of the United States, and that the French methods largely shield the French CPI from the criticisms voiced by the Boskin Commission.

The second part will describe the statistical methods used to address the appearance of new products. We will describe the statistical methods used, without concealing that these methods have their limits and leave plenty of room for improvement. We show, however, that errors can occur in both directions but no convincing studies have been produced to identify and quantify an upward bias—as opposed to what the Boskin Commission did.

In the third part, we will specifically discuss the issue of outlet substitution, and in the last part, we will examine the quantification of the potential upward bias in the French CPI compared with the U.S. index. While there is no basis to estimate a counterpart figure to the Boskin Commission's estimate of 1.1 percentage point bias, our conclusion is that, if an overestimation exists in France, *it is probably far smaller*.

Product Substitution: The Issue of Weighting Patterns

Households change their consumption basket *at the same time* as prices vary. A CPI with weights based on obsolete information may overstate inflation.

We define a consumer's price index between the two periods as the growth rate of his or her budget outlays that makes it possible 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."[1]

Unfortunately, relative prices change at the same time as the general upward or downward movement in prices. To calculate a consumer's price index, we need to track all the products consumed, or at least a representative sample of those products, in order to determine the *average* of the changes in prices for the products in the sample. This raises the question: what weighting should be assigned to each of the products included in the average? The only possible weighting 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 imprecise, 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?

Theory tells us that one of the best proxies of an ideal index is a Fisher index which is an average of an index using base-period weights (called a Laspeyres index) and an index using current-period weights (known as a Paasche index). Theory also shows that a Laspeyres index usually overestimates the Fisher index, while a Paasche index usually underestimates it. A Laspeyres index gives more weight to the products whose prices register the steepest increases, whereas, by logic, the share of these products in consumer expenditures declines once consumers rearrange their expenditures to maintain constant utility.

In practice, 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 a Fisher index cannot be established without current-period weights which take a long time to determine. To calculate a 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. The "age" of the base year used to estimate the weights varies from one country to another. The

[1] *For some economists, the term "purchasing power" implies by construction that the price index used the quantities consumed in the base period. Here, however, we use the form in a broader context.*

older the year, the bigger the overestimation. In France, the weights used to obtain the overall index from "grouping" indexes are updated every year from recent data.[2] In the United States, the base weights are much older.

A Three-level Decomposition

The CPI is estimated through successive aggregations of indexes; each aggregation level has its independent weights, with a different "age." In France, statisticians work at a fairly high level of aggregation for the annual weight updates based on data from year *y*-2. More detailed levels are not treated identically.

We will also 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" aggregation *(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 areas"). The micro-indexes are obtained from price quotations collected from sales outlets in the urban areas examined. The second stage, or "intermediary level" (*niveau intermédiaire*), yields indexes for product groupings from the micro-indexes. In the final stage, or "upper level" (*niveau agrégé*), the overall price index is calculated from the grouping indexes. At each stage, a Laspeyres formula is used (or was used, as we shall see). Upper-level weightings are taken from the national accounts.[3] Intermediate-level weights 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.[4] At each of these stages, a "substitution bias" may occur.

The Upper Level

The methods used to calculate the French index protect it from upward bias at the upper level. Since the early 1970s, the weights of CPI "groupings" are reviewed annually against national accounting data for year *y*-2: in other words, the 1997 monthly price index is based on 1995

---

[2] *The French term* poste, *translated here as "grouping," denotes the broad product categories that constitute the first level of French CPI publication (the U.S. equivalents of "groupings" are the item strata). Examples of groupings include "fresh fruits," "automobiles," and "men's hairdressers." There are 265 groupings in the current index, known as the "base-1990" index. The "base-1980" index had 295 groupings.*

[3] *In the French national accounts, the public surveys on household expenditures and other statistical sources on consumption are collated in sufficient detail to serve as a basis for CPI weightings. In the United States and most other European countries, the weightings are derived simply from public surveys on household expenditures.*

[4] *Thanks to the computerization of the wholesale/retail trade and the standardization of bar codes, explicit weightings at this level of detail—which seemed like statistical science fiction only a few years ago—may become reality in the foreseeable future.*

weights. Accordingly, the current French index — that is, the monthly changes in the current year — is based on recent aggregate weights. The "long-term" French index, which tracks price changes over several years, is based on the chaining of these current indexes. The French index is thus referred to as a "chained" Laspeyres index.[5]

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.[6]

The Intermediary Level

In contrast to the upper level, the weightings of the French intermediate-level index — that is, the weightings of the "sub-items" in the "groupings"— are not systematically 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. If a substitution bias exists at the intermediate level, it probably exceeds the upper-level bias. On balance, a mild upward bias in the range of 0.05–0.10 percentage points a year cannot be ruled out.

The Lower Level

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. It emerged that the use of geometric-means as micro-indexes had two advantages over the conventional arithmetic-mean formulas: (1) the geometric mean allows for possible substitutions between products, whereas the arithmetic mean assumes fixed weightings; and (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.

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

---

[5] *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.*

[6] *For confirmation, we compared the official index with an unchained index for the period 1980–90. As expected, the unchained index was found to rise* on average *by 0.11 percentage points more than the official index (Viglino and Montiel 1995).*

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.[7] The size of the correction procured by the introduction of the geometric mean in France (0.10%) is well below the Boskin Commission's figure for the United States. 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. The switch will thus gradually eliminate the bulk of lower-level substitution bias and formula bias in France.

New Products[8]

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.

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.[9] In absolute terms, that makes more than 30,000 product replacements a year. In 1984, for example, the U.S. CPI excluding housing (and a few other minor items) rose 3.4%. Recalculated *without* the series in which product replacements had occurred during the year, the same overall index showed an increase of merely 0.14% (Armknecht and Weyback 1989).

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.[10] For example, if a car model

---

7 *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.*

8 *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."*

9 *When stated in terms of price quotations collected during the year, the figure is obviously smaller, since quotations are gathered 12 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.).*

10 *French CPI statisticians call this the "treatment of quality-change value"* (traitement de l'effet-qualité).

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 catalogue 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 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.

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.

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.[11] 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.[12]

---

[11] *It is the only method used for heterogeneous sub-items.*

[12] *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 FF5OO *with the April price of* Y.

Direct Comparison (or "equivalent replacement"): This method involves 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 characteristics can be regarded as identical.[13]

This method is applied to 40% of replacements of store-tracked products in the French index.

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.

Explicit Quality-change Valuation: This method involves 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 United States 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 catalogues 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.

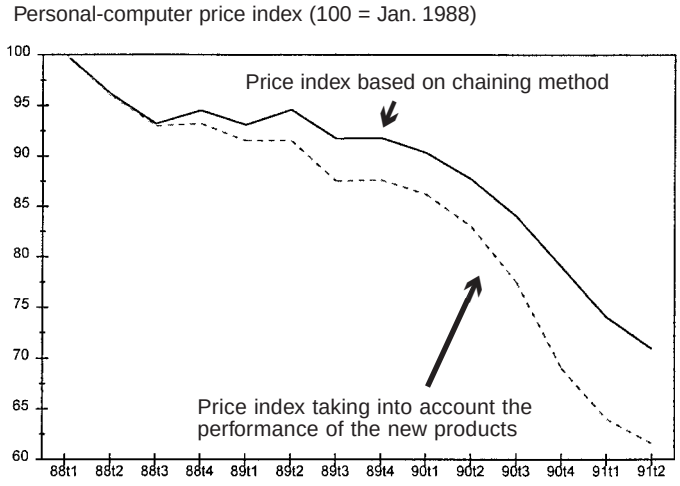
[13] *Of the 1,000 sub-items of the French CPI, about half are so-called homogeneous sub-items; the other half are known as heterogeneous sub-items. Homogeneous sub-items are composed of products whose characteristics are defined with great precision and whose price levels are therefore very similar. 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.*

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. An 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 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 1991b). 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 to 1991



Source: Prix de vente industriels, INSEE (Moreau 1991).

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 (Jacobzone et al. 1997). 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.

Similar doubts may be raised about clothing products. 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.[14] 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 chaining method may have led to an understatement of inflation for clothing products in the United States.

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.

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

[14] *Reciprocally, if actual prices are trending down, the index will be biased upward.*

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 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 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 less safe subways (Norwood 1996; Kuttner 1995).

"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.[15]

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 telephones. Once we have accepted the fact that this service is a significant household-expenditure grouping,[16] 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 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 difficult 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

[15] *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 one part in 1,000 of the weighting of the country's overall index.*

[16] *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.*

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, one cannot conclude so easily that the errors are always going to be overestimations.

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, putting any figure at all on this problem makes little sense.

The fact is that there are no general studies allowing even an approximate quantification of the bias issue. The only objective approach would be to recalculate all the indexes with the "right" methods. This would require a knowledge of such methods, together with data suitable for estimating quality-change values with those methods. In practice, there are many cases where economists themselves do not know what the right method would be. Even if such a method did exist, the data for the estimation would be very expensive to gather. All the authors who have written on new-product bias are therefore reduced to extrapolating partial studies or, like the Boskin Commission, relying on intuitive estimations. Both approaches are shaky and questionable.

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. This figure derives directly from the following statement: "The average consumer 30 years ago would have been prepared at that time to pay 20% more in order to have the present day selection of fruits and vegetables in winter rather than the limited range of those days. 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, it 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 *quantifable* components of quality change. Economists toy with many explanatory models, but very few of these have attained statistical operational status.

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 United States, for example) could lead, instead, to an understatement of inflation. This neglect provides another argument for a far lower bias figure.[17] 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.

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 of 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 litre 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 litre. 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

[17] *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.*

in A.[18] The index will not record a change unless the small shopkeepers (or the other mid-sized or mass outlets), whose prices were tracked previously, lowered their own prices in response to competition from the new supermarket.

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 favour of convenience stores. The blossoming of supermarkets cannot be explained just by low prices. It is to a large part linked to the automobile civilization, to the development of suburbia, and to home deep-freezers; all these phenomena allowed important bulk buying, representing a different commercial service. However, the steady gains in market share by mass-merchandisers also originate from 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 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 United States, 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.

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 1995; Prime and Saglio 1995; Dubeaux and Saglio 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

[18] *Because of the lower price charged by the supermarket, the soda index does not fall in urban area A. This will lead to a sizeable difference between (1) the estimated "volume"* (in the national account sense) *of soda sales in urban area at the time of the supermarket opening and (2) the change in the number of litres sold. This is because the procedure for introducing the new supermarket into the price index implies that a litre of soda sold in a supermarket is less "good" (?) for the consumer and therefore "weighs" less in the total volume sold than a litre 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.*

treated as a "pure" price difference. The difference between the two indexes is estimated at 0.2% 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% per year.

The Overall Estimation

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 United States (Baker 1996). In particular, one would find a massive share of the population below the poverty line in the 1960s.

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.[19]

Table 1
United States and France: Upward Bias in CPI

Upward-bias type	United States[1]	France
	Estimated value, % per year, for years after 1996	
Upper-level substitution	0.15	..
Intermediate-level substitution	0.25	0.05-0.10
Lower-level substitution		..
New outlets	0.10	0.05-0.15
Total Excluding New Products	**0.50**	**0.10-0.25**
New products	0.60	?[2]
Total Including New Products	**1.10**	?

.. not applicable.
[1] Source: Boskin Report, December 1996.
[2] 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% is greatly exaggerated.

[19] 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. 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 than the reforms just mentioned.

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. In fact, the CPI is an essential tool for all users of economics, ordinary citizens as much as decision-makers. The task will therefore require time and money.[20] 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 dishwashers. 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 labour 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 bias and outlet-substitution bias.[21] This approach is consistent with the Boskin Commission recommendations and with initiatives in many other countries. Some detailed studies on the "outlet-substitution bias" from such data have already been published.

---

[20] *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 additional funding, which has recently enabled its head to launch a program to improve the index (Abraham 1997).*

[21] *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.*

Bibliography

Abraham, K. (1997), *Testimony of the Commissioner of Labor Statistics Before the Senate Budget Committee*, January.

Aizcorbe A., and P. Jackman (1993), "The Commodity Substitution Effect in CPI Data, 1982-1991," *Monthly Labor Review*, December.

Ardilly, P. and F. Guglielmetti (1993), "La précision de l'indice des prix," *Économie et Statistique*, no. 267, pp. 13-27.

Armknecht, P. and D. Weyback (1989), "Adjustment for Quality Change in the U.S. Consumer Price Index," *Journal of Official Statistics*, vol. 5, no. 2, pp. 107-23.

Baker, D. (1996), "The Overstated CPI: Can It Really Be True," *Challenge*, September/October, pp. 26-33.

Berndt, E., Z. Griliches and N. Rappaport (1995), "Econometric Estimates of Price Indexes for Personal Computers in the 1990s," *Journal of Econometrics*, July, vol. 68, no. 1, pp. 243-268.

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

Buchwald W., and A. Saglio (1994), "Des indices de prix à la consommation plus comparables entre la France et la RFA," *Économie et Statistique*, no. 275-276, pp. 9-24.

Bureau of Labor Statistics (1993), "The Anatomy of Price Change," reprints from the *Monthly Labor Review*, December.

Bureau of Labor Statistics (1995), *Report from the BLS for the House Budget Committee*, United States Congress, April.

Crawford, A. (1993), *Measurement Biases in the Canadian CPI, Technical report no. 64*, Bank of Canada, September, p. 54.

Deaton and Mullbauer (1980), *Economics and Consumer Behaviour*, Cambridge (U.K.): Cambridge University Press.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers," *Journal of Econometrics*, vol. 4, no. 2, pp. 115-145.

Dubeaux, D., and A. Saglio (1995), "Modification des circuits de distribution et évolution de prix alimentaires," *Économie et Statistique*, no. 285-286, pp. 49-58.

Glaude, M. (1997), "Du bon usage des indices de prix à la consommation," *La Revue Francaise du Marketing*, no. 161.

Gordon, R. (1990), *The Measurement of Durable Goods Prices*, University of Chicago Press for the NBER.

Griliches, Z. (1995), *Senate Finance Committee Testimony*, U.S.

Griliches, Z. and I. Cockburn (1994), "Generics and New Goods in Pharmaceutical Price Indexes," *American Economic Review*, Dec., vol. 84, no. 5, pp. 1213-1232.

Jacobzone, S., E. Martin, V. Perrin and J. Werle (1997), "A Hedonic Insight at the Pricing of Prescription French Drugs: Are Prices Really Rising that Slowly?" memorandum, Crest-INSEE, March.

Kuttner, R. (1995), "Why It's Too Soon to Monkey with the CPI," *Business Week*, November.

Montiel, S. (1996), "Impact sur l'IPC de la moyenne géométrique," *Notes internee INSEE*, no.361/F32O, July 96; no.90/F320, March 21, 1995; no. 144/F320, May 24, 1995.

Moreau, A. (1991a), "Un indice de prix pour les micro-ordinateurs," *Courrier des statistiques*, no.58-59, pp.52-54.

Moreau, A. (1991b), "Micro-informatique, concurrence par les prix," *INSEE Premiere*, no. 175, December.

Moulton, B. (1996), "Bias in the Consumer Price Index: What Is the Evidence?" *Journal of Economic Perspectives*, vol. 10, no. 4, pp. 159-177, Fall.

Norwood, J. (1996), "Statement Before the Senate Finance Committee," U.S. Congress, June.

Oulton, N. (1995), "Do UK Indexes Overstate Inflation?" *National Institute Review*, May.

Pirieu, J.Y. (1983), *L'indice de prix*, Paris: Editions La Découverte/Maspero.

Poinat, F. (1996), "Estimation du biais potentiel sur l'IPC issu de la méthode de chaîmage des agrégats intermédiaires," *Note interne INSEE*, no.315/F320.

Prime, M. and A. Saglio (1995), "Indices de prix et prix moyens : une étude de cas, "*Économie et Statistique*, no. 285-286, pp. 34-48.

Quaranta, V. (1995), "Sensitivity to the Choice of the Formula for Compiling Elementary Aggregates," Paper for the Task Force III on the European Program for CPI Harmonization.

Saglio, A. (1995), "Changement de tissu commercial et mesure de révolution des prix," *Économie et Statistique*, no.285-286, pp. 9-33.

Schultz, B. (1994), "Enchaînement des indices de prix," in W.E. Diewert and C. Montmarquette, eds., *La mesure du niveau des prix*, Proceedings of a conference held under the auspices of Statistics Canada, Ottawa.

Shapiro, M. and D. Wilcox (1996), "Mismeasurement in the Consumer Price Index: an Evaluation," NBER Working Paper, no.5590, May.

Viglino, L. (1995), "À propos des indices de variété par agglomération, réflexions sur les travaux d'harmonisation européenne," *Note interne INSEE*, no.91/F320, March 21, 1995; "De l'utilité de la moyenne géométrique, Communication CEE/OIT Genève 1995," *Note interne INSEE*, no.261/F320, Sept. 13, 1996.

Viglino, L. and S. Montiel (1995), "Effet du chaînage annuel," *Note interne INSEE* no. 176/F320, June 19,1995.

Wynne, M. and F. Sygalla (1994), "The Consumer Price Index," *Economic Review*, pp 1-22, Federal Reserve Bank of Dallas, Second Quarter.

The Boskin Report from a United Kingdom Perspective

David Fenwick[1]
Office for National Statistics

Summary

From a U.K. perspective, the Boskin Report raises no new issues; it simply gives some issues greater prominence. At the U.K.'s Office for National Statistics, as in other national statistical agencies around the world, a substantial amount of research has been conducted over a number of years into methodology associated with consumer prices indices, and this work is continuing.

Our view is that the Retail Prices Index (RPI) presently remains the most accurate single measure of consumer inflation in the U.K. It is produced using the best available methodology following advice from an independent RPI advisory committee. We believe that many of the issues raised in the Boskin Report for the U.S.A. have less force in the U.K. This view is also shared by a number of independent commentators. Nevertheless, along with many other countries, we are undertaking further research and analysis to investigate the issues raised.

[1] *Director, Consumer Prices and General Inflation Division.*

Introduction

From a U.K. perspective, the Boskin Report raises no new issues; it simply gives some issues greater prominence. At the U.K.'s Office for National Statistics, as in other national statistical agencies around the world, a substantial amount of research has been conducted over a number of years into methodology associated with consumer prices indices, and this work is continuing.

Our view is that the Retail Prices Index (RPI) presently remains the most accurate single measure of consumer inflation in the U.K. It is produced using the best available methodology following advice from an independent RPI advisory committee. We believe that many of the issues raised in the Boskin Report for the U.S.A. have less force in the U.K. This view is also shared by a number of independent commentators. Nevertheless, along with many other countries, we are undertaking further research and analysis to investigate the issues raised.

Issues Raised by Boskin

The Boskin Report concludes that the U.S. Consumer Price Index (CPI) has been overstating inflation by somewhere between 0.8 and 1.6 percentage points a year. This conclusion is based on five main arguments.

Commodity Substitution

Firstly, it is claimed that the U.S. CPI takes inadequate account of commodity substitution — changes in consumer spending habits that result from consumers substituting one item for another in response to relative price changes. The argument is that consumers will tend to spend less on goods or services where prices have increased in relative terms and more on those where there has been a relative drop in price. Due to the time lag before these shifts in expenditure are reflected in the weights used to calculate the price index, too much weight will be given to prices that are rising rapidly and not enough weight to prices that are rising more slowly or falling, thereby causing an overstatement of average price rises. Clearly, the amount of any such bias depends on the frequency and speed of introduction of revised weights used in the "representative" shopping basket. Any bias associated with substitution between products in different weighting groups will be greatly reduced if expenditure weights are updated frequently. However, frequent updating of expenditure weights will not affect any substitution bias between products in the same weighting group. The magnitude of this bias is easily found by comparing the published index with a superlative index (Diewert 1976), which is to a very large extent free from this bias. However, a superlative index can only be calculated retrospectively when the following year's weights become available, so cannot be used for the routine production of a price index.

An important methodological difference between the U.S. CPI and the U.K. RPI is that the RPI basket and weights are reviewed annually, while in the U.S. CPI they are reviewed every 10 years, so any commodity substitution bias in the U.K. RPI will be less than in the U.S. CPI. This annual update takes into account changes in the relative amounts of money consumers spend on different categories of goods and services. These changes are measured annually by the Family Expenditure Survey and, for food, by the National Food Survey and are then taken into account each year when the relative weights given to the items in the representative "shopping" basket are revised. In addition, the individual items selected for pricing are updated from these surveys and the latest market research material so that less popular items can be removed from the basket to make room for more popular ones. Thus any bias in the RPI from commodity substitution is likely to be smaller in the U.K. RPI due to the annual updating of the RPI basket and weights.

Outlet Substitution

Boskin's second argument is that the U.S. CPI does not take adequate account of outlet substitution — consumers switching to cheaper outlets. The economic reasoning is similar to commodity substitution. Consumers will tend to spend less in outlets where prices have become relatively expensive and more in those outlets which become relatively cheaper. If these shifts in consumer spending habits are not reflected in the sample of outlets where prices are collected or in any subsequent re-weighting, there will be over-representation of outlets whose prices have risen in relative terms and under-representation of outlets whose prices have fallen in relative terms, thereby causing an overstatement of average price rises. Clearly, the more often the sample of outlets used for price collection is reviewed and the more up-to-date the sampling frame used for shop selection, the more limited the scope for bias to arise from outlet substitution. In addition, sample re-weighting further reduces this type of bias.

In the U.K., price collection is not restricted to large urban areas as in the U.S. and unlike the latter there is a regular cycle of updating the fieldwork areas and of enumerating and updating the sample of the outlets in those areas. Also, for most items, shop type weights are applied to data from independent and multiple shops and there are individual weights for major chains from which prices are collected centrally. These weights are reviewed each year and should compensate for much of the change in shopping patterns. The evidence is that discount warehouses, which are considered to be a major source of outlet bias in the U.S. CPI, are less prevalent in the U.K. In summary, it is maintained that any outlet substitution bias is likely to be smaller in the U.K. than in the U.S.

Formula Bias

Thirdly, the Boskin Report argues that the formula used in the U.S. CPI to average the raw price data ignores the shift in consumer spending patterns between similar products in the same weighting group which results from relative price changes. This formula bias depends both on the choice of formula used to average the raw price data and on consumer reaction to price change, in particular the elasticity of demand for different items within a commodity grouping. Various formulae can be used to combine individual price quotations within the basic commodity groups (see Appendix). The U.S. CPI uses only the average of relatives. The use of the geometric mean as recommended in the Boskin Report is regarded by some (see Dalen 1992) as a way to compensate for substitution bias due to shifts in buying patterns between different brands or varieties of the same commodity. As explained in the Appendix, the average of relatives treats each price quote as equally important. However, if one price rises faster than the others, there is a tendency for consumers to switch away from the brand, variety or outlet represented by that price, so it should be down-weighted to avoid upward bias. It can be shown (Dalen 1992) that the use of the geometric mean is equivalent to assuming that expenditure shares remain constant so that if one price doubles while the others stay the same, the quantity purchased of the brand, variety or outlet represented by that price will halve. In reality, the sensitivity of consumers to price change will vary from item to item. Thus, the blanket application of the geometric mean, and its implicit assumption that items priced have an elasticity of demand of one, represents a simplification of what happens in reality. A more detailed technical background and evaluation of different formulae used for combining price data are in the Appendix.

In the U.K., the ratio of averages is used except for items where the prices within an elementary index are likely to vary a lot, such as for furniture. In these circumstances the average of relatives is used (see Appendix). The sensitivity of consumers to price changes will vary from item to item. Thus for each item one would ideally estimate how much variety and outlet substitution is likely to occur and, in the light of this, decide which formula would be most appropriate. Making the wrong choice would introduce bias in the estimate of price change for that item. The choice of formulae is not clear-cut. Further investigation is needed.

Quality Adjustment

Boskin's fourth point is that the U.S. CPI takes inadequate account of quality improvements, particularly for durable goods. The reasoning behind this objection is that a distinction should be made between price rises due to inflation and rises due to higher quality. If the latter are not fully allowed for and excluded, there will be an over-statement of inflation. Thus, the Boskin Report asserts that in the U.S. context this quality adjustment bias adds to the inflation rate. Note that there may be some items or commodity groups where quality will have deteriorated. Also, there will be cases where the quality change is more apparent than real.

The estimates of quality adjustment bias for the U.S. are not applicable to the U.K. because of differences between the two countries. For example, it is agreed that medical treatment is a problem area for the U.S. CPI. However, medical treatment has a very small weight in the U.K. RPI as most treatment is free at the point of delivery. It is arguable that the introduction of an adjustment for changes in quality is not appropriate for some purposes, for instance for a compensation index. Also, quality bias may be negative in some areas. For instance, it may be argued that the quality of some services, such as public transport, has declined in recent years. The issues of quality change and how to deal with it have been considered for many years (Department of Employment 1986). Quality adjusted price indices are being developed for those sections thought to be most affected (clothing, new cars, computers, audio-visual goods). The difference between these indices and those produced by existing methods will provide estimates of the numerical effect of excluding quality adjustment. Experimental quality adjusted price indices for new cars and computers have already been included in the Harmonised Index of Consumer Prices (HICP) produced for the European Union. It must be considered whether it is sensible to introduce an adjustment for quality change in the RPI. Then there is the question of even-handedness: any quality adjustments must take account of quality deterioration as well as improvement. Quality adjustment is a complex issue both from a conceptual and measurement viewpoint. Further investigation is warranted.

New Goods and Services

The fifth and final issue identified by Boskin relates to the treatment of new goods and services. It is argued that the U.S. CPI overstates inflation because it is slow to incorporate new items into the index or misses them entirely. It is asserted that the prices of new items often fall rapidly in the period immediately after first entering the market; by excluding them, the CPI fails to pick up these price reductions. Also, the longer-term price trends of some new items may be atypical of the general trend in prices within the relevant commodity groups. It may be difficult to distinguish between bias arising from the exclusion of new items and bias resulting from inadequate quality adjustment. A further issue arises in connection with monitoring and incorporating changes in the value of sales during periods of rapid price change which occur immediately after the new goods have been introduced into the market. The sales value of rival products will change as well as that of the new product. This particular issue was not directly addressed by Boskin and is concerned with both new goods and commodity substitution.

In the U.K., the basket of representative items used to construct the RPI is updated annually and in this way is superior to the U.S. CPI. However, research will be conducted to help inform further consideration of this issue.

Conclusion

It is well recognised that no single inflation measure can meet all users' needs. Because of this, the ONS produces other inflation measures as well as the RPI, for example the Producer Prices Indices and the GDP deflator. Even in the context of consumers' expenditures, particular types of households and each individual person may experience different rates of inflation.

It must be noted that the concept of bias very much depends on fitness for purpose. Thus Diewert (1996) argues that bias from new goods and services should be ignored for some purposes, for instance for a compensation index for those on relatively low incomes. These people do not purchase such items, since all their income is spent on basic essentials. It has been argued that quality adjustments should also be excluded from a compensation index if, in effect, consumers have no choice but to accept such changes in quality because products without the quality change no longer exist and so cannot be purchased.

The Boskin report argues that the U.S. CPI should be a cost of living index. Already, the U.S. CPI aspires to be such an index, and many methodological decisions are taken in a cost of living framework. By contrast, RPI advisory committees have recommended that the RPI should not be a cost of living index.

The problems identified in the Boskin Report are inherent in all price indices but the methodology of the U.K. RPI is in a number of significant respects superior to that of the U.S. CPI. In particular, the representative "basket" of goods and services through which the average change in prices for the RPI is calculated is updated annually, unlike the USA where the basket has been, until recently, updated every 10 years. Thus we are much more able to take account of new items and commodity substitution. Also, the outlet sampling procedures used for the U.K. RPI make outlet substitution less problematic.

The arguments on the other two major issues raised in the Boskin Report — quality adjustment and the use of the geometric mean — are less clear-cut. There are two key issues in considering quality adjustment. The first is whether adjusting for quality is sensible in the context in which the RPI is used, the second is that any such adjustment should take equal account of quality deterioration as well as quality improvement. The argument over whether the geometric mean is better than the two formulae currently used in the RPI depends on consumer reaction to price changes and how this varies between different items. In some circumstances, the geometric mean can result in a downward bias. In other circumstances, the geometric mean can produce a higher inflation figure than current methodology.

It is vital to consider the issues that arise from the Boskin Report on a proper factual basis. The ONS will be undertaking further research to assist in this.

Appendix: Formula Effects

An elementary index is an index for a particular good or service identified specifically for the purpose of price collection and is computed from data within a stratum of outlets (depending on the item, prices may be stratified by region, by multiple/independent or both.) The two methods used in the U.K. RPI to produce elementary indices are **average of relatives (*AR*)** and **ratio of averages (*RA*)**. The Boskin Report recommends use of the **geometric mean (*G*)**. If prices $p_{1,0}$ to $p_{n,0}$ are obtained in the base period (which, for the RPI, is January each year) and matching prices $p_{1,t}$ to $p_{n,t}$ are obtained for a particular commodity in the elementary index in a subsequent month, then

$$AR: \quad I_t = \frac{1}{n}\sum_{i=1}^{n}\frac{p_{it}}{p_{i0}} \tag{1}$$

$$RA: \quad I_t = \frac{\sum_{j=1}^{n} p_{jt}/n}{\sum_{i=1}^{n} p_{i0}/n} \tag{2}$$

$$G: \quad I_t = \left(\prod_i \frac{p_{it}}{p_{i0}}\right)^{1/n} = \frac{\left(\prod_i p_{it}\right)^{1/n}}{\left(\prod_i p_{i0}\right)^{1/n}}$$

where I_t = value of elementary index at time t (3)

RA is less affected than *AR* if there is a $p_{i,0}$ that is abnormally low, for example due to January sales, while the corresponding $p_{i,t}$ is not low. However, *RA* has the disadvantage that if one matched pair of prices (i.e. prices for exactly the same good available in both the base and the current month) relates to an object of much higher price than the other pairs, this pair dominates the calculation. *AR* is thus used when the prices within an elementary index are likely to vary a lot, such as for furniture.

AR shows a greater price rise (or smaller fall) than *RA* if the price relatives $p_{i,t}/p_{i,0}$ are negatively correlated with the base prices, which is often the case in practice. However, if the price relatives are positively correlated with the base prices, *AR* shows a smaller price rise or larger fall than *RA*.

Another way to describe the difference between *AR* and *RA* is to consider the expression

$$I_t = \sum_{i=1}^{n} w_i \frac{p_{it}}{p_{i0}} \qquad (4)$$

Ideally, the w_i should reflect actual expenditure, but there are no data currently available to estimate the weights on this basis at this very low level of aggregation, so some assumptions must be made. If it is assumed that all the w_i are equal, this formula becomes *AR*. Thus *AR* is appropriate if each price quote within the aggregate is considered to be as important as any other. However, if the weights are assumed to be proportional to the base price $p_{i,0}$ this formula becomes *RA*. Thus *RA* is appropriate if expenditure is proportional to price.

G is always lower than *AR* (unless the price relatives are all equal, when *G* = *AR*). If most price relatives are roughly equal but there are a few outliers, *G* is not raised so much as *AR* by large price relatives, but is lowered more by small ones. For example, suppose n=10 and nine of the price relatives equal 1. If the other price relative is 2, *AR* is 1.1 and *G* is 1.072. However, if the other price relative is 0.1, *AR* is 0.91 and *G* is 0.794.

G may be higher or lower than *RA*. If the coefficient of variation of the prices (standard deviation divided by mean) is higher in month *t* than in the base month *0*, then *RA* > *G* and vice versa.

Bibliography

Boskin *et al.* (1996) "Toward a More Accurate Measure of the Cost of Living," *Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index*.

Dalen, J. (1992) "Computing Elementary Aggregates in the Swedish Consumer Price Index," *Journal of Official Statistics*, pp. 129-147.

Department of Employment (1986), *Methodological Issues affecting the Retail Prices Index* (Cmnd 9848, 1986).

Diewert, W. E. (1976) "Exact and Superlative Index Numbers," *Journal of Econometrics*, vol. 4, pp. 115-145.

Diewert, W. E. (1996) "Sources of Bias in Consumer Price Indexes," The University of New South Wales School of Economics, Discussion Paper 96/4.

Current Status of the Debate on the Consumer Price Index in the U.S.

Jack E. Triplett[1]
Bureau of Economic Analysis and
National Bureau of Economic Research

Summary

The final report of the Boskin Commission (after its chair Michael Boskin) arrived like a huge boulder dropped into a quiet pool of water. It made an enormous splash in the U.S. and the tidal waves and ripples have spread out all over the world. But eight months after the report's December 1996 publication, the centre of the pool where the boulder landed has become almost completely still.

This paper presents some of the reasons why no opportunities were grasped following the publication of the report. The paper concludes that one missed opportunity is the creation of a more unified economic statistical system. Although Statistics USA is not going anywhere, there is still hope for the future.

[1] *Chief Economist, Bureau of Economic Analysis.*

Introduction

In his introduction, Jacob Ryten has well summarized the responses and reactions to the Boskin report by statistical agencies in other countries. It seems appropriate to go back to the country where the current round of worldwide public discussions began and review the current status of the debate on the CPI in the U.S.

To put it succinctly, in the U.S. the CPI issue is dead. I don't think I'm exaggerating. The report of the CPI Commission (or Boskin Commission, after its chair, Mike Boskin) arrived like a huge boulder dropped into a quiet pool of water. It made an enormous splash in the U.S., it probably had the greatest public impact — certainly the greatest immediate impact — of any report in the history of economic measurement. The tidal waves and ripples from the report have spread out all over the world, as Jacob Ryten has explained in this publication, and have had their own impact. Now, a scant eight months after the report's December 1996 publication, the centre of the pool where the boulder landed has become almost completely still.

One reason I say the centre is now quiet is because of the budget accord that has just been reached in the U.S. between the President and the Congress. The agreement contains no provision for adjustments or cuts in the CPI escalator for Social Security payments or the escalator for tax brackets. The CPI has disappeared as a budget issue.

A second measure of stillness in the centre of the pond is action on improving the CPI. Some CPI changes are being made, and they certainly go in the right direction. The Bureau of Labor Statistics (BLS) has proposed budget increments for useful improvements — increasing the size of the Consumer Expenditure Survey (the basis for CPI weights) from 5,000 to 7,500 consumer units, improving processing systems in order to speed up CPI weight changes, computing a superlative index number in five years, and increasing data collection on quality change and new products. Though those are clearly improvements, they are relatively small, and it remains to be seen whether they will fare better in the remainder of this year's budget process than have other proposals for data improvements in past budget cycles.

For this international audience, I thought I would review the reasons why the drive to improve or change the CPI, or to change the government escalation formulas that use the CPI, seems to have stalled. Some things can be learned from the experience. I emphasize that this is my personal report from inside the beltway, as it's said in Washington. My intention is solely to indicate what seems to have transpired and why it transpired. Among the reasons for what happened are some characteristics of the CPI Commission's report, which I will point out, but these remarks are not intended as a criticism of the Commission's reports, or as a review of the substance of the report, even though some comments on the substance are necessary.

First Round

In the first round of the Washington debate on the CPI, Alan Greenspan proposed some two and a half years ago that Congress could save a lot of money on Social Security by reducing the escalator that tied payments to the CPI because, he contended, the CPI overstated inflation. This was very attractive to Congress because it seemed a way, a politically acceptable way, to cut a very large portion of the budget. It provided a cover, in other words, for an action that many members of Congress wanted to take and that they felt was necessary to take, but which was politically unpopular.

Second Round

In Round Two, the Boskin Commission was appointed. The Commission had two charges. First: What is the number by which the CPI overstates inflation? Second: What should be done to improve the CPI?

Now, in a very real sense those two charges worked against each other. I suspect that even some members of the Commission would now agree.

The first way they worked against each other is from the political perspective. If economists know what the right inflation number is, why fund improvements in the CPI? You don't need to spend the money. Indeed, even though members of the Commission have gone out of their way to compliment the professionalism of BLS staff and their contributions to what is known about measurement problems in the CPI, the message that many readers carried away from the Commission's report was lack of progressiveness, or worse, of the BLS in disposing of the resources that were already at its command.

Second, in Washington, the CPI debate was almost entirely a debate over the number — the Commission's 1.1 percentage point bias estimate got the headlines. The CPI measurement issues were much too complicated for most members of Congress to understand, too complicated for the press and certainly for most of the public. That wasn't true of everybody, I do not want to be misunderstood. I heard a presentation by a very influential Congressman who said, as nearly as I recall: "We'll just give the BLS money to calculate geometric means for the lower level estimator and a superlative index number for the higher level aggregation." There is sophistication in Congress on these issues, and in the press. But by and large it was not the CPI measurement issues that got the public attention, it was the number, the 1.1. That was the news story: Is that 1.1 number "correct"? If you understood Washington, you wouldn't have expected anything else, which is a second reason why the two charges to the Commission worked against each other.

Third Round

So, Round Three was a protracted and acrimonious public debate over the number. The CPI Commission's report has a lot of good things in it. But the public discussion of the report concerned the number almost exclusively, and it was certainly true that the 1.1 number was vulnerable, vulnerable in the sense that it was too easy for people to criticize it.

In Washington, and probably in all politics (and for good reasons), when people find a small error in a number, or in the logic or calculation method for the number, it diminishes the credibility of other parts of a report. And I think that's essentially what happened. In an academic discourse if you have some small number error it is excused because it is the big picture that counts. But in a public debate it's often the small points about a number that matter. Statistical agencies understand that very well: Making an error in your number is much more publicly damaging than making an error in your procedure.

Part of the 1.1 bias estimate was non-controversial — the roughly 0.5 percentage point of substitution and other technical biases. It was the 0.6 point estimate of quality change bias that attracted public criticism. We read in the press, for example, of real estate agents who understood (to put the matter into economists' language) that no hedonic function for housing ever goes through the origin, so when the Commission in its quality change chapter, calculated the *average* price per square foot of rental housing, it over-adjusted for the value of quality change. This did not come from economists in the form of technical criticisms, but from real estate agents picking up on the small points. Reporters also found things to criticize and, as one said to me, "If this was a sound report I shouldn't be able to find things wrong with it." People did their own back-of-the-envelope calculations that suggested that some of the back-of-the-envelope calculations the report contained were done a little bit too quickly. I did a couple, Brent Moulton did some for a paper published in the *Brookings Papers* on Economic Activity series, and I've seen others. I don't think those calculations had a big impact but they had some.

Though it did not enter the public discussion, I think others saw more fundamental problems with the report's chapter on quality change bias. It seems to me that the chapter treated research that showed downward CPI bias very differently from research that showed upward bias. So, perceptions of problems with the report's quality change chapter subtracted as well from professional backing for the 1.1 number.

I do not want to be misunderstood. I respect the effort that went into the report, and the professional concern for improving economic statistics that take motivated members of the Commission. But I am giving my assessment of what happened. This is a blow-by-blow account of what happens to a political issue inside Washington. The report had a fundamental flaw in having numbers in it that people could too readily pick apart, because the discussion of the details of the numbers diverted attention away from the recommendations at the end of the report, some of which were very profound recommendations that have long-term potential.

Then there was Round something like 3 1/2, which were the polls. The polls were clearly a strategic blunder. Mike Boskin initially took a poll (I take it there was some external push on this) that was rather selective and that got into the newspapers. So this was a bit of a problem. But more to the point, it conferred validity on other polls.

The *Wall Street Journal* took a confidential poll of 320 academic economists. Nearly three-quarters of them (71%) said the CPI bias was less than 1.1 percentage points. The "Blue Chip Economic Indicators" people polled their forecasting panel, 86% of whom said 1.1 was too high. The polls clearly increased doubts in the minds of many people about the l.l number. Had Boskin not taken the first poll, he could at least have said, "Polls don't matter, it's the expertise that matters, and the Commission has the expertise." But once he took his poll, that made it legitimate for others to poll. And also, to reiterate the point I made before, the Commission's recommendations were more than ever lost in the press reports about the 1.1, whether the various polls were or were not biased, and so forth.

Fourth Round

At this point, Congress was in the position of taking action on the Commission's report. I have been told of a private meeting in Congress where some members were admonished: "An expert panel has stated that the CPI bias is 1.1, Congress should pass legislation to put this adjustment into Social Security payments." One back-bench member finally raised his hand and said, "Wait a minute, I don't understand this. We were going to get the BLS to take care of it, so we didn't have to vote."

And that, of course, is exactly the point. If Congress had to vote whether to cut the 1.1 from Social Security payments, or add it to the tax rates, then they would have to take a stand on reducing benefits or raising taxes, and if they had been in a position to do that in the first place, there would have been no need for a CPI Commission to generate an automatic cut. The attention given to the 1.1 number had all along interfered with the objective of improving the CPI, but at this point it began to interfere as well with the objective of cutting the deficit through a CPI adjustment. At a meeting at Brookings, a Congressman said, "The spotlight is very bright on Congress right now, the brighter is the spotlight, the harder it is for Congress to act." The great public attention on the 1.1 number inhibited direct Congressional action on the escalation formulas.

In fact, to conclude Round Four, Congress and the President did not act. After this enormous water splash surged up and fell out all over the world, the budget escalation formulas were not changed. It is true that the budget incorporates reduced projections for the CPI rate of increase to the year 2002, but after all, the projection may change for a whole host of reasons other than improvements in the CPI (more confidence that the Federal Reserve has broken the back of inflation and will continue to hold the line, for example). At the very end of the budget process, I understand that there was an agreement to take a small amount off the escalation

formula (perhaps 0.15 points, this being the estimate of CPI commodity substitution bias), but in the end there was no change in the formula, largely because a new estimate of the budget deficit showed that deficit reduction had largely been attained, without a change in the escalator. So there is no action on the issue that originally made CPI measurement issues so newsworthy — the possibility of reducing the Federal budget deficit by cutting 1.1 points off the CPI inflation escalator.

Fifth Round

Round Five should have been the long-range moves to improve the Consumer Price Index. Round Five is the round that unfortunately didn't happen. As I have already noted, there have been a number of small moves forward. The changes that the BLS have now proposed are clearly in the right direction. But no one can argue that the changes that have been proposed to the CPI are anywhere nearly consistent with this big splash that went up in December. There were three reasons why Round Five did not come off.

One reason was that Congress and the Administration wanted initiatives that would reduce the CPI. At one point, there was political discussion about giving BLS resources to improve the CPI. But implicitly, improving the CPI suggested some sort of guarantee that the improvements were going to reduce the rate of increase. However, if you look back at recent improvements in the CPI, some of them reduced the rate of increase but some of them added to the rate of increase. It is very hard to say that the next two or three improvements that come up are necessarily going to reduce the index. If you can't say that, then Congress has lost interest in funding improvements in the CPI.

Second, though I think my BLS friends would disagree with this, there was no real BLS agenda on the table. The BLS Commissioner asked for an outside commission on the CPI, so the Boskin Commission was not unexpected from the BLS perspective. Yet, there was no comprehensive agenda for improving the CPI that could be hauled out in the middle of the big splash. Jacob Ryten made the remark that, in Statistics Canada, you would regard this as an opportunity, not a threat. Heads of statistical agencies and heads of price statistics programs in other countries have made the same point to me in conversations. There is an opportunity to improve economic measurement. Well, there is no question there is a threat. The threat arises because public discussion of measurement issues, discussions of the kind that have been in the newspapers in the U.S., can be very damaging to a statistical agency's professional reputation. There is little doubt that the discussion in the U.S. has damaged the BLS reputation in some quarters. But there is also an opportunity. And perhaps I shouldn't say this so bluntly, but I think there was no opportunity grasped, or maybe even grasped for, except in a marginal way.

I will give one example. The U.S. Consumer Expenditure Survey has a sample size of 5,000 households or "expenditure units." In the past, averaging three years of consumer expenditure surveys has been necessary for CPI weights. The BLS proposes to increase the

survey to 7,500 households to get from two years' data the same level of accuracy as before — not improved accuracy, but the same level. And so far as I can see, there is no program for addressing the substantial problems that exist in that survey now. Sample attrition is one, reporting accuracy is another. These are problems that affect all the other uses for the survey, uses that ought also to weigh in on the case for improving the expenditure survey. But even considering CPI uses, the detailed weights for the CPI (the expenditure totals for the CPI's 200-odd individual expenditure classes) are not accurate now, they rely on the law of offsetting errors to assure that the weighting error of the *overall* CPI is small. It is hard to see how one can have a program for improving the CPI without addressing the database on which its weights are based. The enhancement that has been proposed will only increase the sample size from 5,000 to 7,500, which looks almost minuscule compared with the scheduled 35,000 sample size of the Canadian household expenditure survey. It is such a small step.

The third reason this Round Five didn't happen — to repeat what I have said before — was all the flack and debate over the Commission's 1.1 number. At the end, we lost sight of the recommendations, mainly because of the controversy over the 1.1% estimation.

The CPI Commission's report certainly made a lot of news. When a statistical initiative hits the big time, one would like to find that when you're done with it you've got something positive out of it. It seems to me that we are back in a pool that is only a little bit more active than it was before. There haven't been many big gains out of several months' turmoil.

Statistics USA

Someone asked about "Statistics USA." As most people know, the U.S. statistical system is fragmented into many pieces — nobody knows how many actually — so that economic statistics are split about six ways and not even Paasche and Laspeyres know what is where. The proposal for unifying the statistical system, or for unifying the economic statistics agencies, had probably not a lot of political things going for it in the first place. When you change the structure of the statistical system, you have also got to change the structure of the way the budget is handled in Congress. For example, the committee that handles the BLS budget is not the committee that handles the Bureau of the Census budget. So, if you change the structure of the organization, you've also got to change the jurisdictions among those committees. There isn't anything politically to be gained by any member of Congress going out and agitating for this, and there are possibly some losses.

Some members of Congress are interested in the issue, nevertheless. I talked with one several months ago. He wanted to hold hearings, to find expert witnesses, to build the case for putting the statistical system together. I have believed for a long time (since my work on the Bonnen Report on the U.S. statistical system in the late 1970s) that the statistical system would be much better off if it were unified. But finding examples that could be used in hearings, examples of innovations that could be made, where the improvement would occur, that's harder. If you give an example, someone will disagree and there goes your consensus.

I'll give one example to show what I mean. The U.S. collects what used to be called SIC[1] codes from different places. The BLS does it with one survey, the Census Bureau does it as part of the economic censuses. There is duplication of effort just to find out which establishment is in which SIC. I tried, when I was chair of the Economic Classification Policy Committee in the U.S., to stop that waste. It was roughly $5 million a year, plus a lot of extra reporting burden plus, when you were done, there were large numbers of workers in manufacturing, for example, who were misclassified by one or the other survey. How could you possibly justify $5 million, extra reporting burden, and bad data? Well, people came in and argued "No, no, no, we want to keep doing what we've always been doing." And, of course, the arguments they gave, it wasn't that they were invalid, it's just that if you had a central statistical agency with a head like Ivan Fellegi of Statistics Canada, he would say: "Look there's a broader issue here that needs to be addressed, and although I understand these little problems, let's work this out." In the U.S. decentralized system, it just got stalled, and nothing happened. But you can't use something like that in hearings, because you bring in this example and somebody's going to say: "No, that's not so, there are all these problems we told you about." And then you've got a big argument about what is so or isn't so.

Conclusion

I can think of many cases in the U.S. system where unifying economic statistics would create a net gain. But I can think of few cases where everyone in the system says the change is a gain and no one objects (those get done, even in a decentralized system). As soon as you're in that situation, you've got a controversial hearing and there are no politics in it.

So, I don't think that Statistics USA is going anywhere. But I continue to have hopes for the future. A centralized statistical agency could, if it attracted a strong professional as its head, greatly improve the performance of the U.S. system.

[1] *SIC - Standard Industrial Classification.*

Bias in the CPI: Experiences From Five OECD Countries

Edited by Louis Marc Ducharme
Prices Division, Statistics Canada

Catalogue No. 62F0014MPB, No. 10
ISBN: 0-660-59374-2
ISSN: 1206-2677

13-A8, Jean Talon Building, Ottawa, K1A OT6

Facsimile Number: (613) 951-2848

Telephone: (613) 951-0688

Email: ducharl@statcan.ca

September 1997

Texte français au verso

Bias in the CPI: Experiences From Five OECD Countries

Edited by Louis Marc Ducharme
Prices Division, Statistics Canada

Catalogue No. 62F0014MPB, No. 10
ISBN: 0-660-59374-2
ISSN: 1206-2677

13-A8, Jean Talon Building, Ottawa, K1A OT6

Facsimile Number: (613) 951-2848

Telephone: (613) 951-0688

Email: ducharl@statcan.ca

September 1997

Texte français au verso