Catalogue no. 62F0014MIB

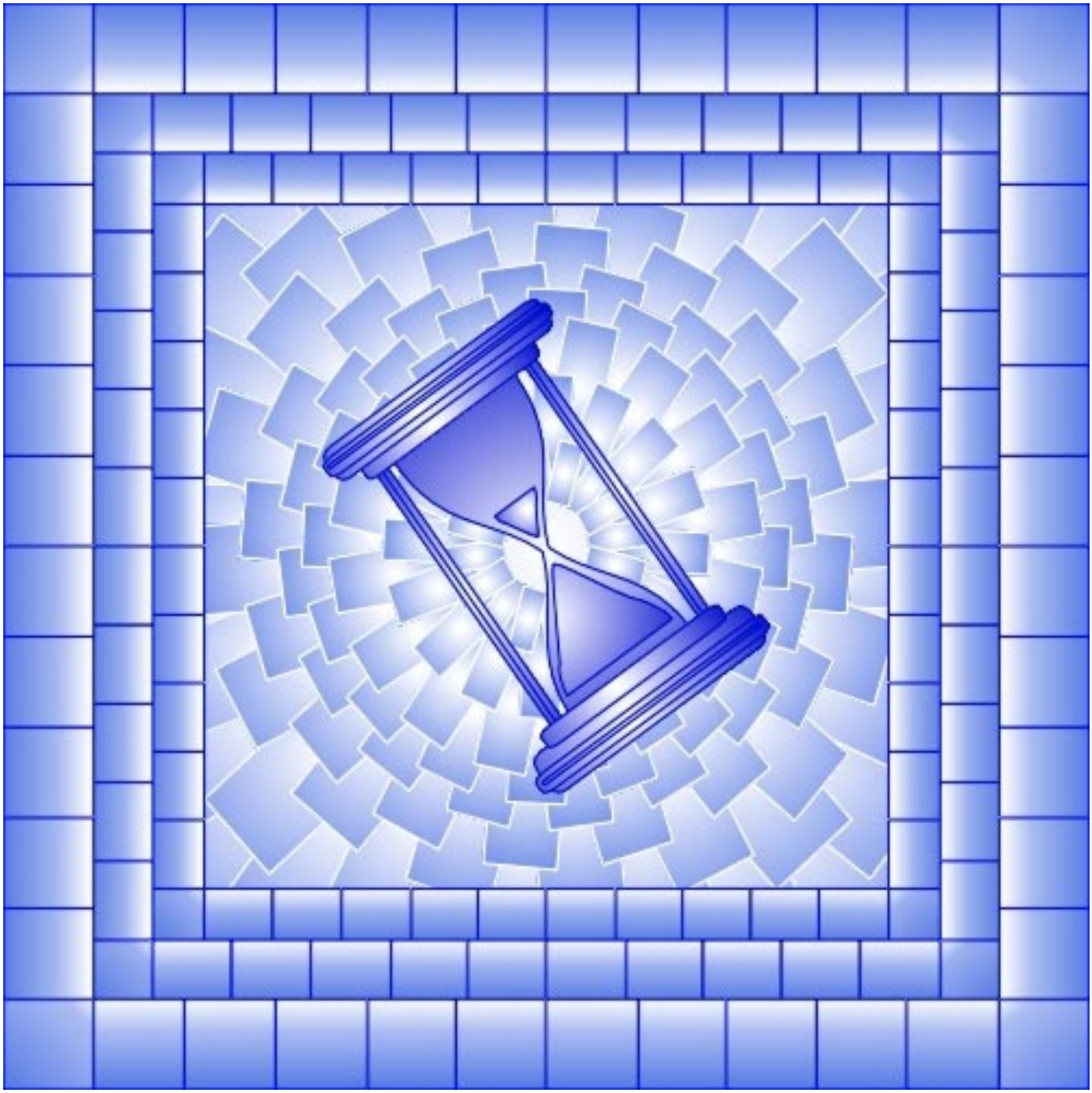
Analytical Series

Prices Division

Potential Uses of Scanner Data – A Case Study Using Coffee Data

By Hugh Scobie

No.6



Statistics Canada Statistique Canada

Canada

Potential Uses of Scanner Data – A Case Study Using Coffee Data

By Hugh Scobie
Prices Division, Statistics Canada

Catalogue No. 62F0014MPB, Series No. 6
ISBN: 0-660-59377-6
ISSN: 1206-2677

13-D7, Jean Talon Building, Ottawa, K1A OT6

Facsimile: (613) 951-2848

Telephone: (613) 951-2036

Email: ***scobieh@statcan.ca***

This paper is available at www.statcan.ca/english/freepub/62F0014MIB/free.htm

October 1998

Preface

Prices Division first started publishing the Analytical Series in December 1996 as a means to convey conceptual and applied research undertaken by its staff, and at times, by other persons from within or outside Statistics Canada on the subject of price indexes.

All papers are reviewed by a panel of experts from within Statistics Canada or outside the agency. Views expressed in the papers are those of the authors and do not necessarily reflect those of Prices Division or Statistics Canada.

The purpose of the series is to disseminate knowledge and stimulate discussion. Questions and comments on any aspect of the papers are welcome and can be forwarded to Louis Marc Ducharme, Director (Internet email: ducharl@statcan.ca; Telephone: 613-951-0688) or to Robin Lowe, Chief, Quality Assurance (Internet email: lowerob@statcan.ca; Telephone: 613-951-9495), Prices Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

Abstract

Many statistical agencies are currently investigating the use of scanner data for their own purposes. Interest has grown in the potential uses of these data to improve the quality of price indexes. This paper reports on initial research done in Prices Division at Statistics Canada. The paper looks at scanner data and the feasibility of its use to produce Consumer Price Index (CPI) estimates. It evaluates current CPI methodology and procedures, and examines the impact the use of scanner data would have on the CPI commodity indexes. The main focus of the study, however, is to explore the impact that scanner data would have on the CPI basic commodity indexes covered by scannable items. Since the CPI criteria relate to a limited selection of scanner data, an examination will be made of the impact of gradually relaxing the criteria to include more products and outlets from the scanner data. The initial subset was derived by applying the CPI criteria of volume selling brands and outlets. Each of these changes in criteria yielded a different subset of scanner data. Calculations were performed using these various subsets of scanner data, and their results were compared to the CPI. An analysis of the results will be used in determining the strengths and limitations of the CPI data, in detecting any deficiencies, and in providing information for the revision of the pricing selection.

TABLE OF CONTENTS

Tables

TABLE OF CONTENTS (continued)

Page

Charts

Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

1.0 Introduction[1]

Scanner data, particularly grocery store sales, has developed to a high degree and is now widely used. Many statistical agencies like Statistics Canada are currently investigating the use of scanner data collection systems for their own purposes. Standardised scanner purchase information is routinely collected and processed for large representative samples of outlets on a weekly basis. The data are available for most products at the level of individual brands and package sizes. Interest is growing in the potential uses of this data source with the objective of improving the quality of our price indexes. This paper reports on initial research done in Prices Division on the use of scanner data in the production of price indexes.[2] Three potential uses of scanner data are:

i) to produce better Consumer Price Index (CPI) estimates;
ii) to evaluate current CPI sampling methodology; and
iii) to shed light on the economic behaviour of commodities.

The CPI estimates are derived from price surveys that are based on judgmental samples of items and outlets. Strictly speaking, sampling errors in these surveys cannot be quantified. This approach is justified by the belief that competition leads to approximate uniformity in price movements of identical goods in any particular market. In practice, the CPI selects volume selling items and outlets that yield a limited number of price observations. Price observations are taken on specific days of the month.

The goal of the study is to use the wide variety of prices from the scanner data to examine the strengths and limitations of our data, to detect any deficiencies, and to provide information for the revision of the pricing selection. The study will focus mainly, however, on how scanner data can produce better CPI estimates. This will be done by exploring the impact that scanner data would have had on the basic CPI commodity indexes covered by scannable items for which scanner data is collected. Since the CPI criteria relate to a limited selection of scanner data, we will examine the impact of gradually relaxing the criteria to include more products and outlets from the scanner data. The initial subset was derived by applying the CPI criteria of volume selling brands and outlets. Each of these changes in criteria yielded a different subset of scanner data. Calculations were performed using these various subsets of scanner data, and their results were compared to the CPI.

The question of the quality of the CPI is of interest to many segments of society. The general public is impacted since many government programs are indexed automatically which affects the benefits that people receive. Sources of measurement error in the CPI have become quite topical with the increasing costs of government programs. As well, analysts are interested in this type of information because it helps interpret the reasonableness of the CPI. It also provides a yardstick for negotiations in salary and other money transactions.

---

1 *We would like to express our gratitude to A.C. Nielsen-Marketing Research who kindly provided the scanner data for this study on coffee.*

2 *A preliminary version of this paper was presented to the Price Measurement Advisory Committee of Statistics Canada on May 13, 1996.*

Research on scanner data is a relatively new area of study. With the growing use of scanners to record purchase transactions, statistical agencies such as Statistics Canada will need to develop ways to utilise this source of data. Whether in a research or production mode, scanner data remains an invaluable tool for analysis.

2.0 Organisation of the Study

The study is organised into four major sections:

i) scanner data and the feasibility of its use;
ii) evaluation of current CPI methodology and procedures;
iii) impact of using scanner data on the CPI commodity indexes; and
iv) conclusions and implications for the CPI programme.

The first section explores the nature of scanner data and the selection of representative commodities to research, the problems with scanner data and the techniques used to edit them, and their impact on the quality of the CPI data. The second section deals with certain aspects of the CPI methodology and procedures, such as the reliability of the criteria for item and outlet selection, the size of the price sample, the frequency of price collection, and the calculation procedures for basic commodities. The third section examines the potential impact of using scanner data on basic commodity indexes. The last section draws conclusions from the study and describes the implications on the CPI programme. The initial calculation duplicates the CPI methodology with a subset of the scanner data, but subsequently, expands it to include more observations or outlets to compare to the CPI.

3.0 Scanner Data and the Feasibility of its Use

3.1 Description of Scanner Data

Scanner data is raw purchase information on products derived through the reading of bar codes by optical scanners. For the purposes of this study, scanner data is limited to products purchased from supermarkets. For some products, sales in other types of outlets (e.g. specialty stores) may be quite significant. Each product is identified by a universal product code (UPC). In practice, a unique UPC describes a specific variety, size and package type. For example, a UPC would describe a variety of instant coffee as "NESCAFÉ, Rich Blend, Jar, 200 grams". Several UPC's may therefore be assigned, for example, for the same brand of coffee to distinguish between various sizes or types of packaging. Pre-packaged commodities are relatively standardised and therefore generate standardised information. Some products in a grocery outlet such as meat, fruit and vegetables, bakery goods, bulk food and food sold in case-lots, such as soft drinks, for which quantities and sometimes quality vary according to the purchased unit, do not provide standard units. As a result, other methods of data collection may be required to supplement the existing scanner data.

Coffee was selected as a typical standardised product for which a sufficiently large number of observations were available. A.C. Nielsen provided scanner data on weekly coffee sales for both instant and roasted coffee in Toronto and Montréal over a three-year period from September 1992 to August 1995. The market shares of instant and roasted coffee differ quite significantly between cities. In Montréal, instant coffee represents 58% of coffee sales and roasted coffee 42%. The sales in Toronto are the complete opposite with instant coffee representing 43% and roasted coffee 57%. Over 2 million observations made up the scanner data sample, representing approximately 20% of the major grocery outlets in the two cities. In comparison, the CPI sample over the same three-year period consists of around 3,000 observations.

Although the database is large, most of the revenue generated by coffee sales is concentrated in a reasonably small subset of products. A Lorenz curve constructed from observations on roasted and instant coffee in Toronto and Montréal showed that over 80% of sales were concentrated in one-third of all the items, and more than 99% of sales in two-thirds of the items.[3]

The scanner data on coffee is distributed over 573 separate universal product codes with total weekly quantities expressed in units sold and revenues in dollars from a total of 157 outlets in Toronto and Montréal. Each UPC is identified by the outlet in which the sale originates. Each outlet sells many varieties of coffee, but within chains, they do not necessarily carry the same selection. Unique codes were assigned to each outlet by A.C. Nielsen, but no information was released on either the name of the chain or its location within each city. As long as there was one sale associated with a particular UPC, an observed quantity and revenue were produced. By dividing the total revenue by the total quantity sold, a price per unit was obtained for each week.

Because the CPI only takes a "snapshot" of transactions during the month by selecting specified items and outlets at particular points in time, scanner data is more comprehensive than CPI data. Scanner data represents all commodity transactions that took place in the selected outlets throughout the full three years. In examining this data, it was noted that sales of specific varieties of coffee can vary considerably from week to week within a month and, in many cases, may be near or equal to zero for particular varieties in some outlets.

3.2 Problems with Scanner Data

Over the course of this study, a number of problems were experienced in the use of scanner data. These problems relate to the following three aspects: volume of the data, the frequency of gaps, and the extensive classification detail.

[3] *SeeChart 1 - Lorenz Curve for Observations on Coffee Universal Product Codes on page 16. The curve was constructed based on total sales over the three-year period for all varieties of coffee.*

3.2.1 Volume of Data

Problems associated with the volume of data should not be underestimated. The scanner data for coffee contained 2 million observations for sales of all varieties over the three years. This sample represents an average of 50,000 observations per month over the study period. Sales of many varieties of coffee are infrequent which are reflected in their lower sales figures. Some aggregation of detail is required in order to simplify index calculation. Aggregation of varieties should be done in some cases when they are essentially the same good. On the question of aggregating over stores, Reinsdorf[4] states, "Whether to aggregate over stores is a more difficult question. CPI policy has been to treat items in different stores as different goods, which in effect assumes that all price differences between stores reflect quality differences". Aggregating outlets within the same chains therefore would appear both desirable and feasible.

3.2.2 Frequency of Gaps

The frequency of gaps in data is high – even for volume selling varieties. For sales of a given UPC in a given outlet in any given week, there is a 40% chance that the reported sales will be zero. One problem is determining whether these records reflect zero sales or zero availability. The discontinuities in the data complicate the comparison of prices over time. Table 1 illustrates these discontinuities by aggregating records according to their frequency of sales reported. Over the full three years, less than 5% of records, accounting for 17% of total coffee sales, contained observations in every week. Those with two or fewer years of observations in any continuous period accounted for 70% of records but only 30% of sales. Approximately 40% of records had continuous data for a year or less, but represented less than 10% of sales. Weekly revenues for varieties of coffee can vary considerably. These revenues can be influenced by promotional sales or local buying habits. New varieties were also introduced over the three-year span. The large volume of replaced varieties and outlets with such a large number of records involves considerable work in maintaining a matched sample over time for the purpose of reflecting pure price movement.

Table 1: Total Coffee Sales by Frequency of Observations (September 1992 to August 1995)

Frequency of Sales Reported	Percentage of Records	Percentage of Total Sales
52 weeks or less	41.6	9.9
104 weeks or less	68.6	30.8
130 weeks or less	80.4	46.9
144 weeks or less	89.1	64.2
150 weeks or less	92.8	73.7
155 weeks or less	96.6	82.8
156 weeks or less	100.0	100.0

[4] *See Reinsdorf, M. (1995), "Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee" by the U.S. Bureau of Labor Statistics.*

3.2.3 Classification Detail

Data are classified at a very fine level of detail. Many varieties of the same coffee are available in numerous sizes. As well, multiple varieties of the same coffee are recorded separately. For example, specialty coffees such as Maxwell House's Café au lait, Cappuccino, Espresso or Vanilla Cappuccino are all recorded separately. Combining UPC's with identical price-determining characteristics[5] helps to reduce the number of gaps in the data. The varieties are, in effect, the same good. To a great extent, the large volume of data is due to the detailed classification.

3.3 Techniques for Overcoming Discontinuities in Scanner Data

A number of techniques were used to overcome the discontinuities in the scanner data and to facilitate efficient use of all of the available data.

3.3.1 Aggregation of Weekly Data to Monthly Data

Because weekly data are not necessarily available for infrequently purchased brands, weekly revenues and units sold were summed to create monthly data. Also, a monthly average price is more consistent with what is used in the CPI, where prices are collected over a three-week period. An average price per unit for the month was then derived by dividing monthly revenues by the quantity (unit) sold in that month. This conversion to a unit price facilitated the use of quantities of any specified period to aggregate prices of different sized packages. By aggregating data monthly, only 13.3% of the cells were empty.

3.3.2 Imputation for Unreported Months from Existing Data

Quantities and revenues for each month in which the item was unreported (e.g. out-of-stock) were imputed from the previous month's data. The assumption was that the same price remained. This rule was performed for a maximum of three successive months after which time the item was dropped. In those periods in which prices were rising, the imputation method would have had a dampening effect on the price change, while the reverse effect would be true when prices were declining. The effect of these imputations should not be significant because they represent a limited number of observations in any month.

3.3.3 Exclusion of Low Sales Items

Based on their revenues over three years, the least important 10% of items were excluded on the premise that their market share was small. In terms of their share of total sales, these items represented an insignificant portion (0.1% maximum). Their exclusion removed a large number of gaps in monthly data.

---

[5] *Ibid.*

For some items, observations were very sporadic. The database was further edited. Records that contained less than 6 months of continuous data over the three years were excluded. Although a significant number of these items were excluded, their share of total sales was negligible, preventing them from adversely affecting the calculations. The impact of editing on the scanner data sample in terms of percentage of records excluded with their share of total sales are as follows.

Table 2: Impact of Editing on Scanner Data Sample[6]

Product	Market	Percentage of Records Excluded	Percentage of Total Sales
Roasted Coffee	Toronto	19.0	3.7
Instant Coffee	Toronto	15.9	2.0
Roasted Coffee	Montréal	28.3	5.2
Instant Coffee	Montréal	21.4	2.4

The coffee data indicate that there may be, presently, many operational problems in using scanner data for the production of CPI price series. The sheer volume of data would require an overhaul of existing computer processing and editing systems. New editing procedures would need to be devised to ensure accuracy of the final results. Coffee is a relatively standardised commodity which may not be the case for other products. Non-standardised commodities may require tailor-made treatments. The availability and the timing of data may make scanner data unsuitable to meet actual rigid CPI production time constraints. With the advances in technology, many of the operational problems may be overcome in the future.

4.0 Evaluation of Current CPI Methodology and Procedures

The current CPI methodology and procedures can be evaluated in two ways: first by analysing the raw data from scanners, and then by calculating sub-indexes. The scanner data contains both instant and roasted coffee sales in Montréal and Toronto.

For an easy reference to the charts relating to the variants for roasted and instant coffee, please refer to Table 3 on the next page.

[6] *Table 2 indicates that there is a higher percentage of records for which observations are discontinuous in Montréal than in Toronto.*

Table 3: List of Chart Numbers for Variants of Roasted and Instant Coffee for Toronto and Montréal

Variants	Toronto		Montréal	
	Roasted Coffee	Instant Coffee	Roasted Coffee	Instant Coffee
	Chart Numbers*			
Standard Brands, Standard Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted	4	15	25	35
All Standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted	5	16	26	36
All Non-standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted	6	17	27	N/A
All Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted	7	18	28	N/A
All Brands, All Sizes, Top 14 Outlets, Weighted and Unweighted	8	19	29	37
All Brands, All Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted	9	20	30	38
All Standard Brands, Standard Sizes, All Outlets, Weighted	10	21	31	39
All Non-standard Brands, Standard Sizes, All Outlets, Weighted	11	22	32	N/A
All Brands, Standard Sizes, All Outlets, Weighted	12	23	33	N/A
All Brands, All Sizes, All Outlets, Weighted	13	24	34	40

* Charts are located on pages 16 to 35.

Initially, instant and roasted coffee were viewed from the perspective of their distribution of sales by package size. Chart 2 on page 16 shows the distribution of sales of roasted coffee in Toronto and Montréal representing sales over the three years of data. The chart confirms that the 300-gram size represented in the CPI is the volume seller with over 50% of sales in both cities. Sales for the 300-gram size are, however, considerably higher in Montréal. The sales for various other sizes such as 369 gram, 1 kg. and 1.1 kg. are significant in Toronto while those for 225 gram are notable in Montréal. For instant coffee (shown in Chart 3 on page 17), the CPI representative size of 200 grams leads sales in both cities. However, 150-gram sales are quite prevalent for instant coffee in Toronto and Montréal.

The CPI methodology tends to represent top selling brands in volume selling outlets. The outlet sample is determined in terms of overall grocery sales. The criterion of largest sales of groceries in determining top selling stores is more practical than using sales of particular grocery commodities. Applying the CPI criteria of selecting the top brand in each of the leading 14 outlets in terms of coffee sales can yield a different selection of items. Such was the case for roasted coffee in Toronto where the top selling brands in the top 14 outlets were generally "store brands" rather than "national brands", and in the 1.1 kg. size rather than the 300-gram size.

The CPI price samples are chosen using criteria based on judgement, rather than probability sampling techniques. Judgement is involved in the selection of brands, sizes and outlets included in the CPI sample. By using scanner data that takes into account all transactions in a larger sample of outlets, an indication of the possible sampling errors inherent in using judgmental sampling can be derived. A number of questions can be addressed concerning how representative the CPI sample is using scanner data. How representative are i) standard brands of all brands within a commodity class, ii) standard sizes of all sizes, iii) top selling brands of all brands and iv) volume selling outlets of all outlets? Since scanner data also provides information on the various brands, the impact of weighting using sales on the overall price movement can be measured. In order to shed light on these questions, various sub-indexes were calculated using scanner data and these were compared to the CPI. The methodology for these calculations is explained below.

The analysis is done by comparing sub-indexes calculated in various ways derived from the total scanner data. The sub-indexes were calculated as month-to-month ratios of the sum of the unit prices for the items (UPC's) collected in two consecutive months. The unit prices used to calculate the sums were obtained from *matched samples*, which means they referred to items of the same quantity and the same (or equivalent) quality in both months reflecting, therefore, *pure price movement.*[7]

The unit prices and units sold were standardised to the CPI sample (e.g. 300 grams for roasted coffee). The indexes were calculated either unweighted, for which the unit prices in each month's sum were equally weighted, or weighted with fixed weights. The fixed weights were derived for each observation by averaging the monthly units sold over the number of months for which observations were available.

For sub-indexes calculated as the sum of unweighted price observations for the representative commodity, the calculation was as follows:

$$P_{t/b} = \frac{\sum_{i=1}^{n} p_t}{\sum_{i=1}^{n} p_b}$$

[7] *Movement that is not attributable to changes to either the quality or quantity of the given commodity.*

where

$P_{t/b}$ is the price relative for the commodity,

$\sum_{i=1}^{n} p_t$ is the sum of the standardised prices of all brands in the observed period (t), and

$\sum_{i=1}^{n} p_b$ is the sum of the standardised prices of all brands in the base period (b).

For sub-indexes calculated as the weighted sum of price observations for the representative commodity the calculation was as follows:

$$P_{t/b} = \frac{\sum_{i=1}^{n} p_t q_c}{\sum_{i=1}^{n} p_b q_c}$$

where

$P_{t/b}$ is the price relative for the commodity,

$\sum_{i=1}^{n} p_t q_c$ is the sum of the standardised prices of all brands in the observed period (t), weighted with fixed weights; and,

$\sum_{i=1}^{n} p_b q_c$ is the sum of the standardised prices of all brands in the base period (b), weighted with fixed weights.

The initial calculation of the sub-indexes started by selecting from the whole database a subset for coffee in a particular city using the same criteria used for constructing the CPI. The application of these criteria resulted in a smaller subset of all available data. Then, the choice of selected commodities was gradually expanded to see the impact on the results.

As previously stated, coffee is available in a wide variety of products and sizes. There is a standard size (SS) that represents the volume selling size. Many products are also sold in non-standard sizes (NSS). Within the range of coffee products, there are brands of standard quality coffee (SB) and others, including premium brands, flavoured coffee, and so on. These may be referred to as non-standard brands (NSB).

For the CPI, only standard brands and sizes of coffee products are considered (SB, SS). The sample is further restricted by only selecting one brand per outlet, which should be the best selling brand in that outlet. This is a matter of judgement, and the choice is made according to the importance of their sales. The CPI outlet sample is composed of 14 outlets in each of Toronto and Montréal, chosen because they are the leading sellers of groceries.

The observations are equally weighted and the average price of the 14 observations in each period is compared with the average price in the preceding period, or with the subset of observations that are present in both periods. It is also important to note that a missing observation may be imputed as last month's price for up to three months.

In order to compare to the CPI, we selected in the first phase the top selling 14 outlets with the largest sales. The top selling outlets were selected on the basis of their overall sales of coffee over the three year period. For each outlet, we selected the individual brand (within SB, SS) that had the largest sales over the period. These 14 observations were averaged each period. Charts 4 and 15 (for Toronto) and charts 25 and 35 (for Montréal) in each case shows the results.[8]

In the second phase, we expanded the selection, to include all standard brand and standard size products (not just the best selling one) in each outlet. In the third phase, non-standard brands were included, and finally in the fourth phase, we included non-standard sizes. One should note that the outlet sample still covers only the largest 14 outlets, and the weighting is done with both fixed and equal weights. It is clear that equally weighting all brands, including some with very low sales, can lead to unrepresentative results. The fixed weights used in the calculation reflect the total sales over the three year period divided by the number of months for which non-zero sales were posted.

The calculations reflected in the various charts that are outlined in Table 4 on the next page show the impact that the expansion of commodities within the top 14 outlets would have on the CPI.

[8] *Charts 4 and 15 (for Toronto) and charts 25 and 35 (for Montréal) represent Standard Brands (SB) and Standard Sizes (SS),with one item selected, and with equal weights used for coffee.*

Table 4: Impact of the Expansion of the Sample, Top 14 Outlets

Toronto		Montréal		Products	Size	Outlets	Weights
Roasted Coffee	Instant Coffee	Roasted Coffee	Instant Coffee				
Chart Numbers							
5	16	26	36	SB	SS	14	Equal and average monthly sales
6	17	27	N/A	NSB	SS	14	Equal and average monthly sales
7	18	28	N/A	SB+NSB*	SS	14	Equal and average monthly sales
8	19	29	37	SB+NSB*	SS+NSS	14	Equal and average monthly sales

* Non-standard brands not available for instant coffee in Montréal

SB = Standard Brand SS = Standard Size

NSB = Non-standard Brand NSS = Non-standard Size

We could also select the best selling brand from the whole selection, in each of the 14 outlets. Many of them may be the same as selected in Charts 4 and 15 (Toronto) and Charts 25 and 35 (Montréal), but many may not. Therefore, Charts 9 and 20 for Toronto and Charts 30 and 38 for Montréal are (SB+NSB), (SS+NSS), select 1 item, 14 outlets, equal and average of monthly sales as weights.

We can also expand the selection to include observations from all outlets. As in dealing with products with little sales, there is not much value in dealing with outlets with little sales and giving them equal weights. The next set of tests repeats the previous ones:

Table 5: Impact of the Expansion of the Sample, All Outlets

Toronto		Montréal		Products	Size	Outlets	Weights
Roasted Coffee	Instant Coffee	Roasted Coffee	Instant Coffee				
Chart Numbers							
10	21	31	39	SB	SS	All	Average monthly sales
11	22	32	N/A	NSB*	SS	All	Average monthly sales
12	23	33	N/A	SB+NSB*	SS	All	Average monthly sales
13	24	34	40	SB+NSB*	SS+NSS	All	Average monthly sales

* Non-standard brands not available for instant coffee in Montréal.

SB = Standard Brand SS = Standard Size

NSB = Non-standard Brand NSS = Non-standard Size

Charts 13 and 24 (Toronto) and charts 34 and 40 (Montréal) use all the data, with fixed weights for the period, and this measure should be the most authoritative estimate of price change. The other measures, including the actual CPI, should be compared to it.

5.0 Impact of Using Scanner Data on Basic Commodity Indexes (September 1992 to August 1995)

The results of the sub-index calculations for roasted coffee are shown in charts 4 to 13 (for Toronto) and charts 25 to 34 (for Montréal). As for the results of the sub-index calculations for instant coffee, they are shown in charts 15 to 24 (for Toronto) and charts 35 to 40 (for Montréal). Some charts have been omitted for instant coffee in Montréal due to the lack of non-standard brands in the A.C. Nielsen sample. An additional chart (14) has been added for roasted coffee in Toronto to illustrate the varying price trends for non-standard sizes.

In reviewing coffee price movements from September 1992 to August 1995, their price movements were relatively stable in both Toronto and Montréal prior to June 1994. However, from June to December 1994, coffee prices accelerated quickly because of crop failures due to poor growing conditions in Brazil and in other coffee growing countries. After December 1994, coffee prices remained, for the most part, stable to the end of the study.

In reviewing the results of the various calculations, the following general observations can be made concerning the scanner data for roasted and instant coffee in both Toronto and Montréal. Similar results were obtained for both sets of data except as indicated for Toronto.

In the initial period of price stability, all of the scanner sub-indexes yielded essentially the same result as the CPI. It is only when prices accelerated due to market fluctuations that differences were noted subsequent to June 1994.

The CPI tends to move most like the scanner sub-indexes for the subset of observations selected using criteria followed by the CPI for brand and size. Charts 4, 5, 10, 15, 16 and 21 (for Toronto) and charts 25, 26, 31, 35, 36 and 39 (for Montréal) show the incremental impact of enlarging the coverage in both commodity and outlet dimensions. All the sub-indexes tend to be very similar to the CPI prior to October 1994 but higher afterwards. Although these scanner sub-indexes can be influenced more by short-term monthly fluctuations, the general trend is similar to the CPI.

When the selection is made from the top selling brand in the volume selling outlets, the scanner sub-indexes can be influenced by month-to-month fluctuations (e.g. see Chart 9 for Toronto roasted coffee). Interestingly, the criteria of the top selling brand in the top selling outlet does not yield the same subset of observations in the scanner data as the CPI for roasted coffee in Toronto. The top selling brand in the top selling outlets yields a non-standard brand in a non-standard size. In Chart 9 for Toronto, the top selling brand for the 14 top selling outlets reflects movements of the 1.1 kg. size. This size is for a store brand rather than a national brand and its movements reflect frequent price change over the period. The standard size is, however, the most representative size for the overall market as illustrated by Chart 13. Within sizes for Toronto roasted coffee, however, the application of the criteria used in the CPI would not yield the same subset of observations as in the CPI survey. The non-standard sizes also exhibit quite a different price change pattern. Toronto roasted coffee in these non-standard sizes tends to be subject to more monthly fluctuations as shown in Chart 14.

The prices of all standard brands in standard sizes tend to change more during the period when prices are accelerating than the top selling brands in the volume selling outlets. Since top selling brands would tend to be more "price-sensitive" due to the competition between grocery chains, this factor may influence the rate of price change.

Non-standard brands tend to be less influenced by rising prices than standard brands. In the period September 1994 to December 1994, the prices of non-standard brands for roasted coffee, for example as shown in Charts 6 and 27, rose less, in most cases, than the standard brands. Non-standard brands of instant coffee in Toronto, however, shown in charts 17 and 22, exhibited a similar price change pattern to that of the CPI.

When non-standard brands are combined with standard brands for standard sizes in Charts 7, 12, 18 and 23 (for Toronto), the moderating price change of the non-standard brands draws scanner sub-indexes closer to the CPI. The addition of non-standard sizes in Charts 8, 13, 19 and 24 (for Toronto) draws the CPI and scanner sub-indexes even closer.

Within the aggregation of standard or non-standard brands, the use of equal or unit sales weights to combine varieties produces similar rates of price change. This is not surprising for standard brands, but is for non-standard brands where there is a diversity of varieties. When standard and non-standard brands are combined, sales weights should be used due to the dispersion of price change between them.

Scanner sub-indexes for the top 14 outlets are very similar to those for all outlets for the same set of observations.

6.0 Conclusions and Implications for the CPI Programme

This study provided an initial investigation of scanner data but further research is required into other commodities that are available in grocery outlets and in other types of outlets. With the increasing availability of automated data, its use for direct pricing of commodities covered in the CPI, rather than interviewer price collection, should be studied. This is particularly pertinent for most grocery outlet items, which are pre-packaged and therefore generate standardised information.

Although this study refers to only one commodity, coffee, it demonstrates the usefulness of this type of data in assessing and improving the quality of the CPI commodity indexes. From this study, several conclusions can be arrived at that have implications for the CPI programme.

1) Although the prices of both standard and non-standard brands move similarly in periods of stable prices, standard brands tend to be more influenced by rising prices than non-standard brands. In order to be more reflective of overall price change, the CPI commodity sample should be expanded to include more diversity of brands, particularly, non-standard brands.

2) Although the CPI standard size of 300 grams is the volume seller for roasted coffee in both Toronto and Montréal, supermarket sales of other sizes in Toronto, in particular, are significant. Since these sizes seem to reflect fluctuating short-term price change, they may have an impact on a monthly CPI. The sample should, therefore, be expanded to include non-standard sizes.

3) The price change derived from volume selling outlets gives similar results to that for all outlets. This result would tend to confirm the validity of sampling volume selling outlets in the CPI.

4) The top selling brand in volume selling outlets may not always be representative of the universe of price change for a commodity. For example, for roasted coffee in Toronto, the top selling brand was a non-standard brand sold in a non-standard size. In most cases, however, the top selling brand in volume selling outlets is representative.

5) The use of sales weights does not appear to be required within the aggregation of standard or non-standard brands. Sales weights however should be used when combining standard and non-standard brands due to the dispersion of price change between them.

CHARTS

Chart 1: Lorenz Curve for Observations on Coffee Universal Product Codes (September 1992 to August 1995)

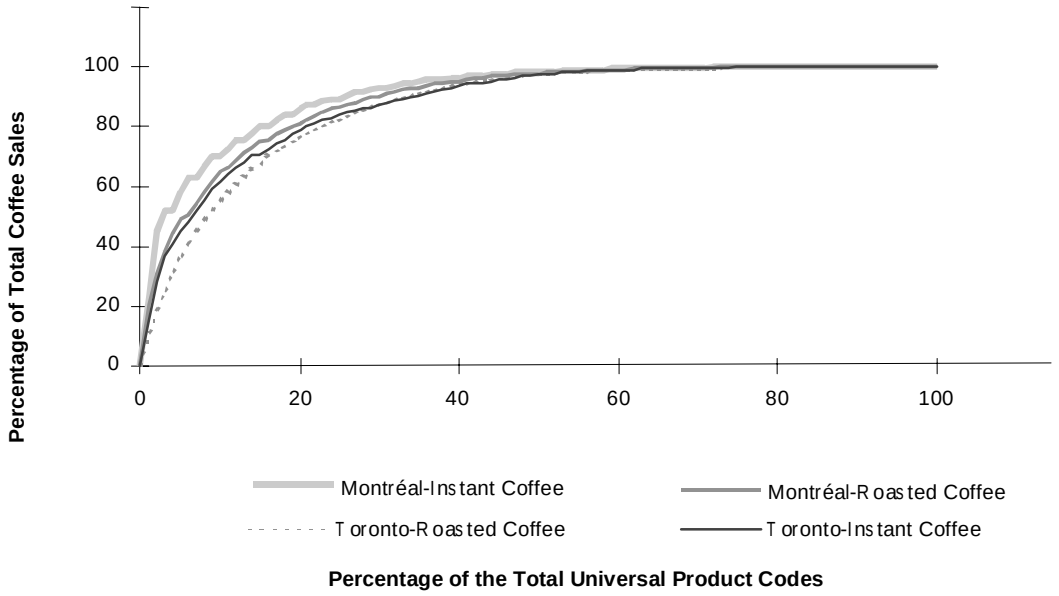


Chart 2: Distribution of Supermarket Sales of Roasted Coffee by Package Size, Toronto and Montréal

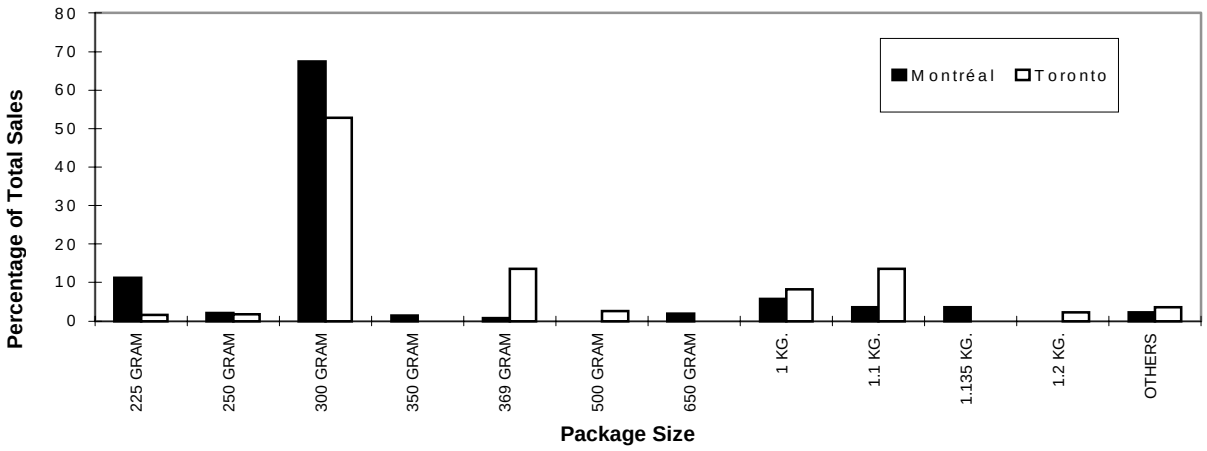
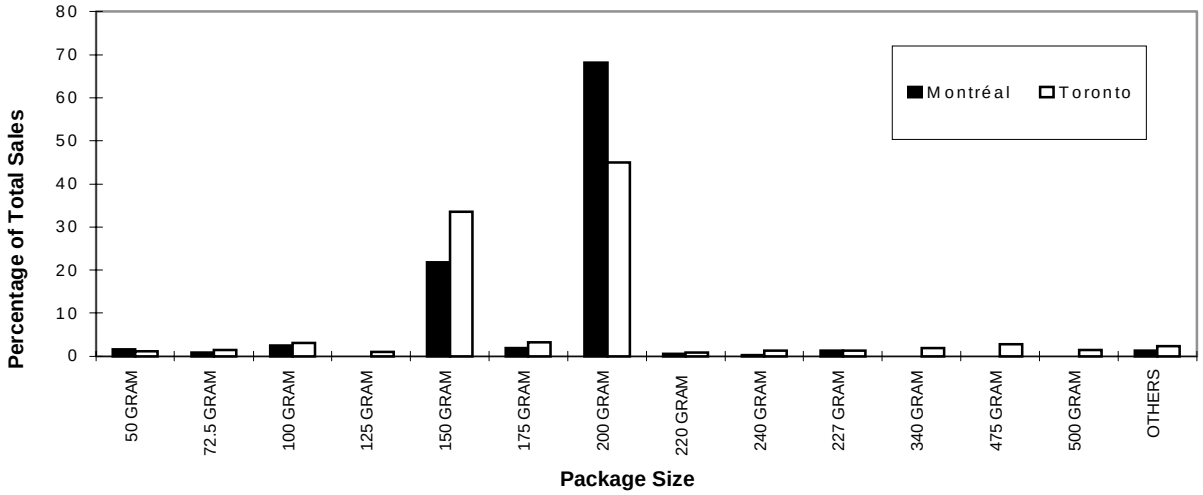
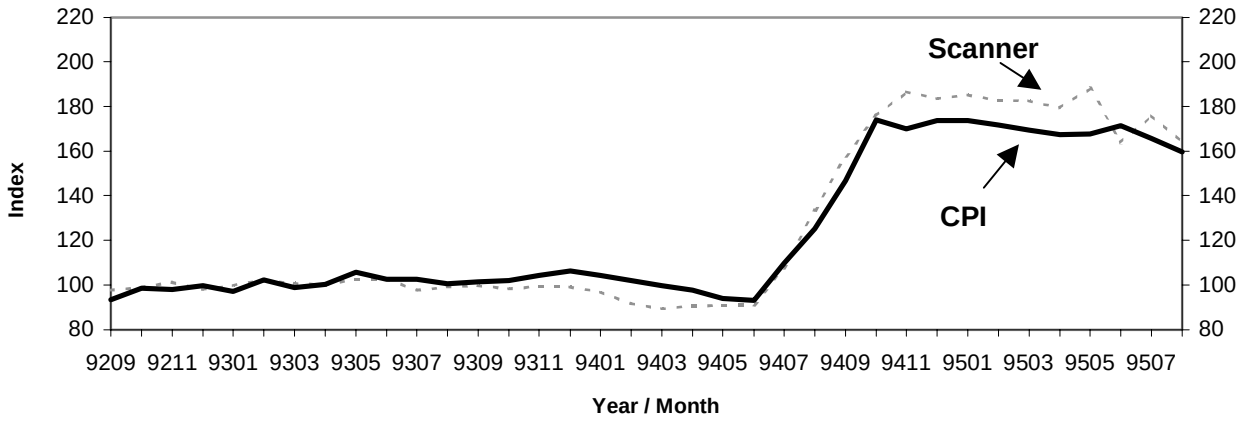


Chart 3: Distribution of Supermarket Sales of Instant Coffee by Package Size, Toronto and Montréal



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100)

Chart 4: Standard Brands, Standard Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 5: All Standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted

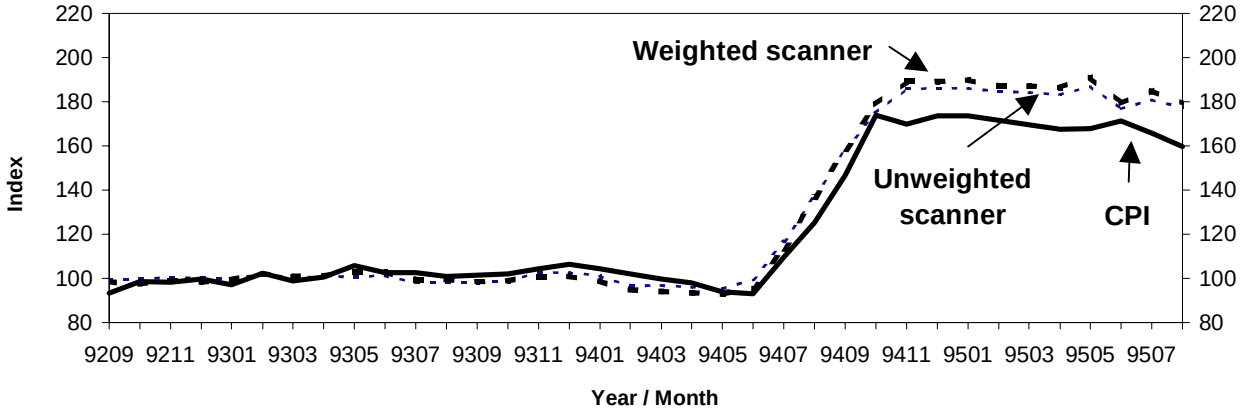
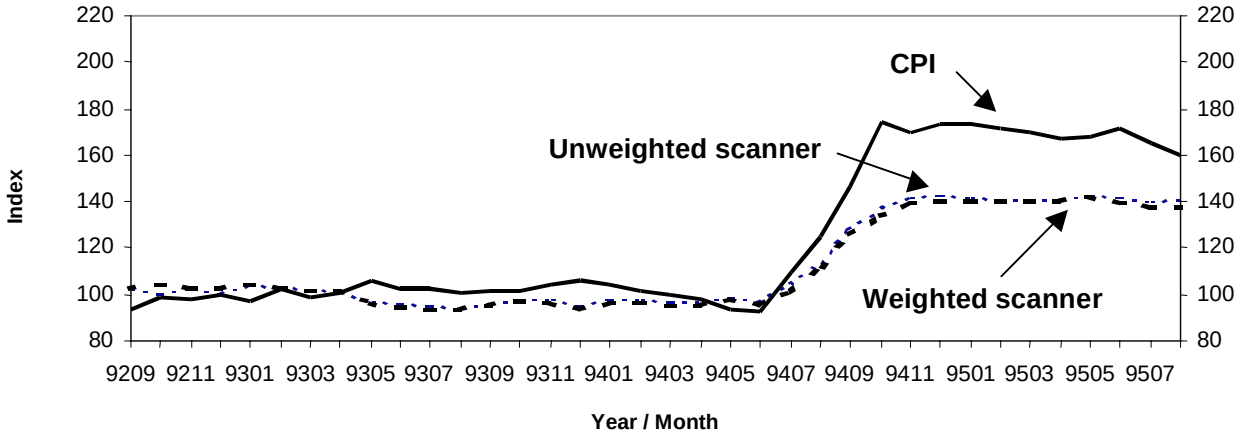


Chart 6: All Non-standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 7: All Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted

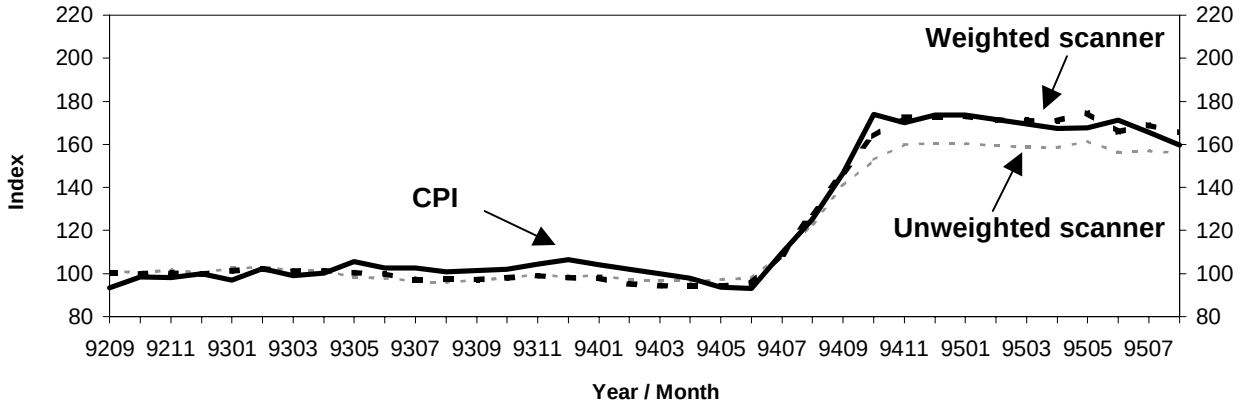
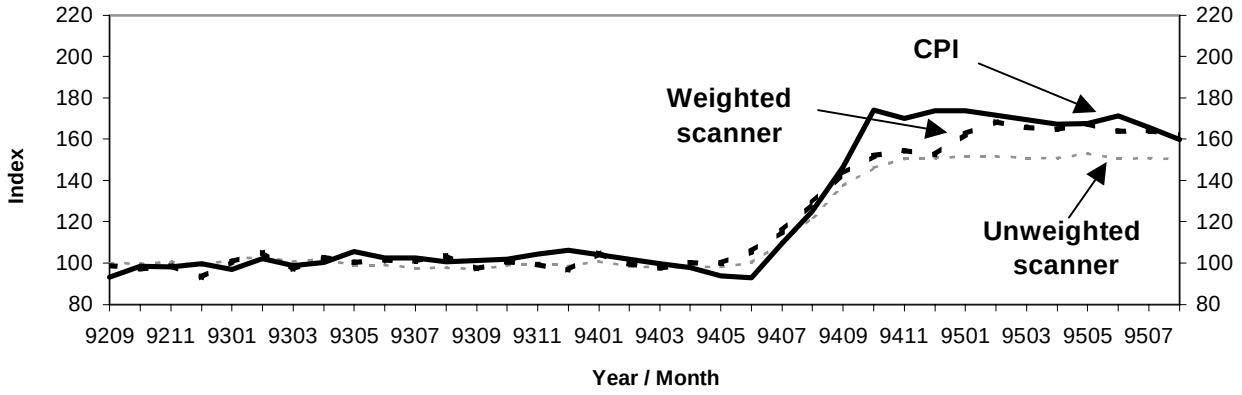


Chart 8: All Brands, All Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 9: All Brands, All Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted

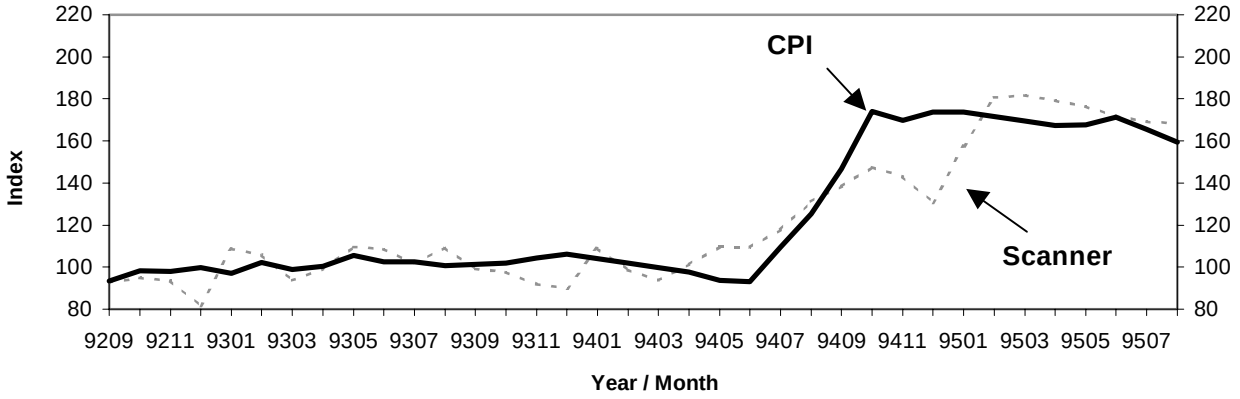
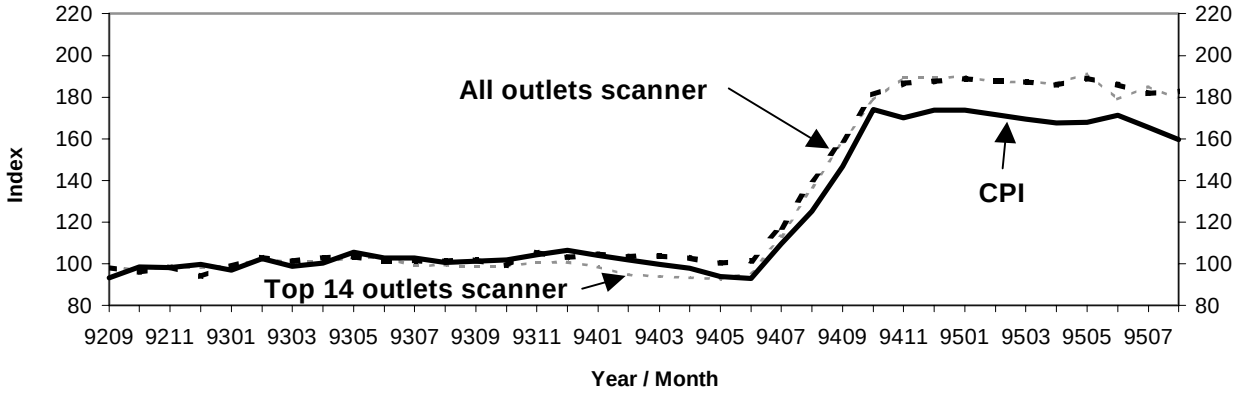


Chart 10: All Standard Brands, Standard Sizes, All Outlets, Weighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 11: All Non-standard Brands, Standard Sizes, All Outlets, Weighted

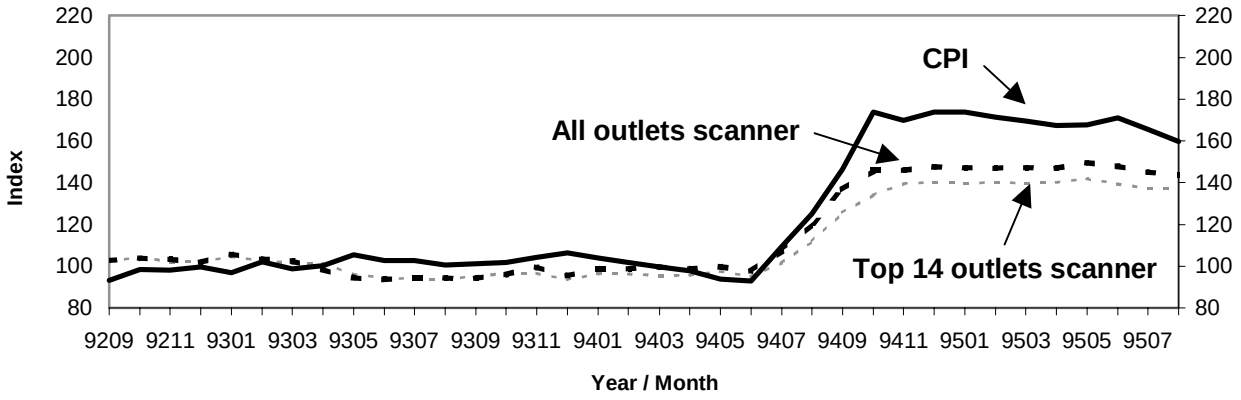
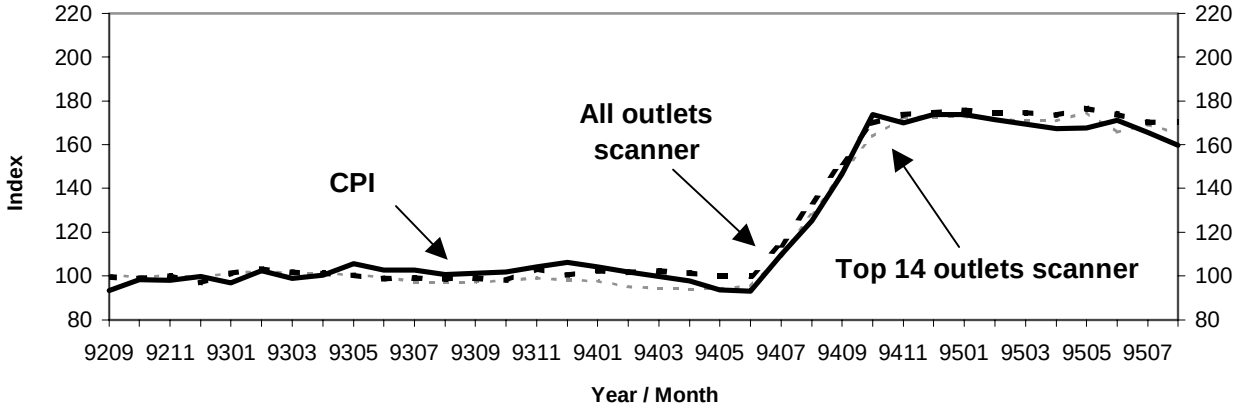


Chart 12: All Brands, Standard Sizes, All Outlets, Weighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 13: All Brands, All Sizes, All Outlets, Weighted

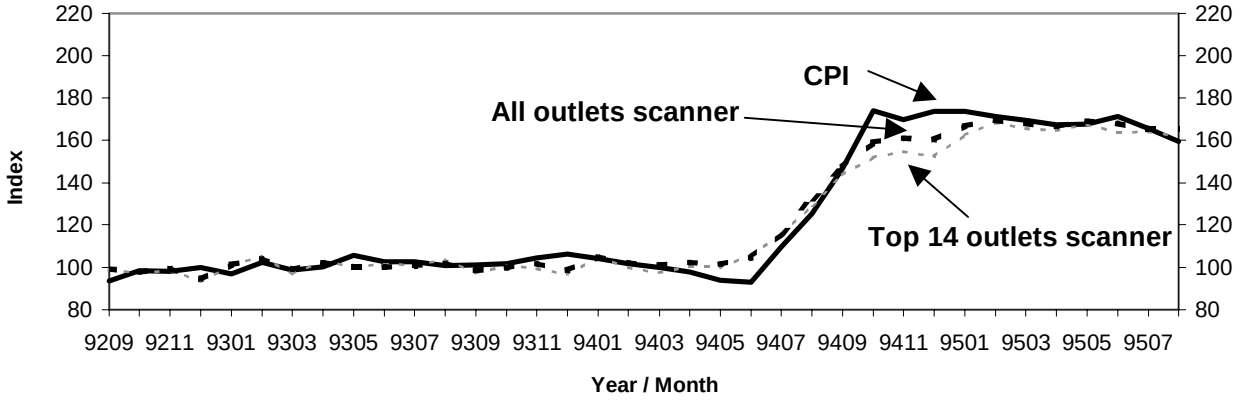
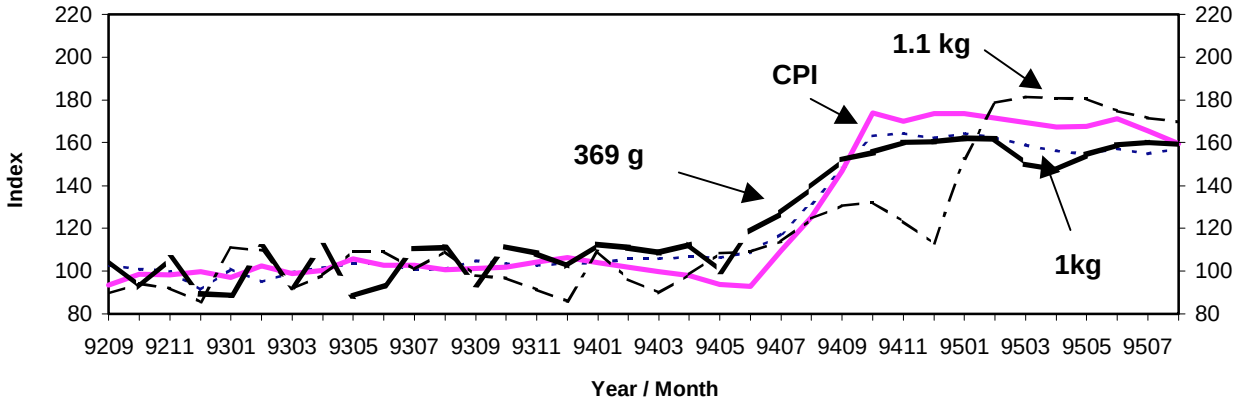


Chart 14: Comparison of Scanner Sub-indexes for Non-standard Sizes and CPI for Roasted Coffee for Toronto (September 1992 to August 1993=100)



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Toronto (September 1992 to August 1993=100)

Chart 15: Standard Brands, Standard Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted

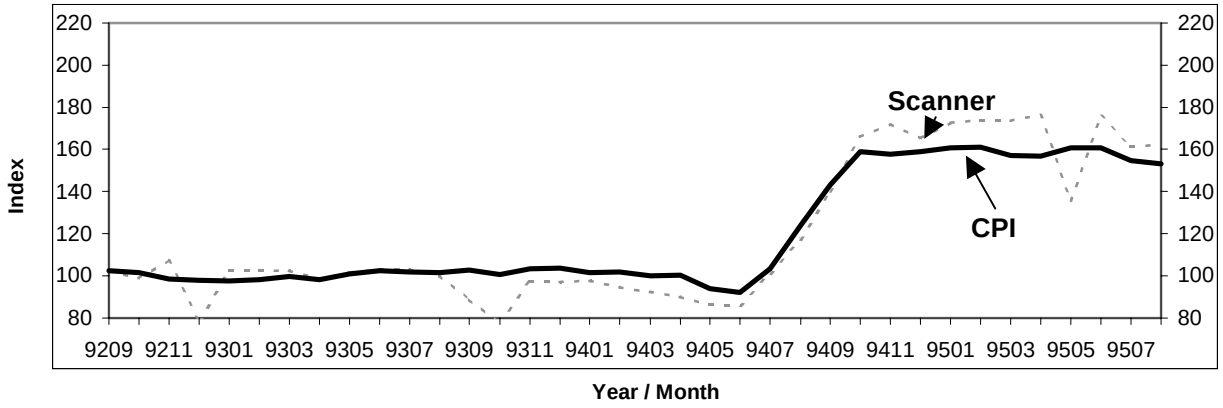
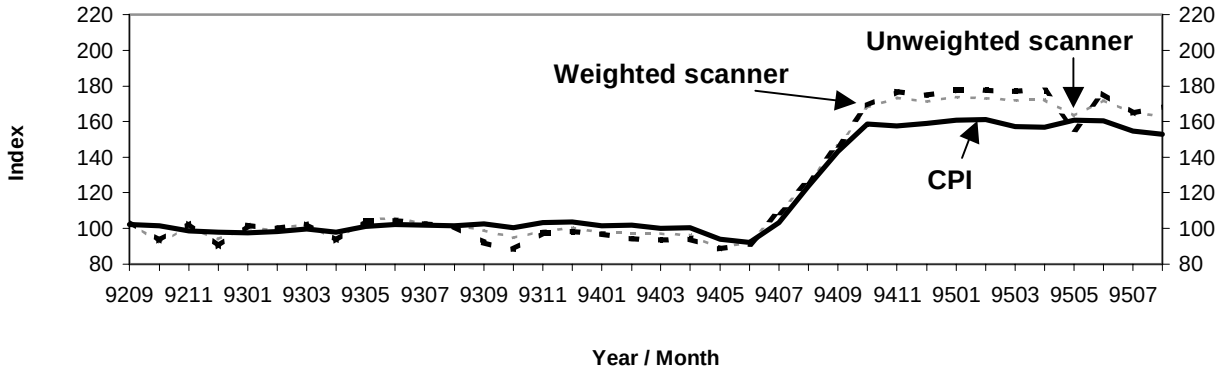


Chart 16: All Standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 17: All Non-standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted

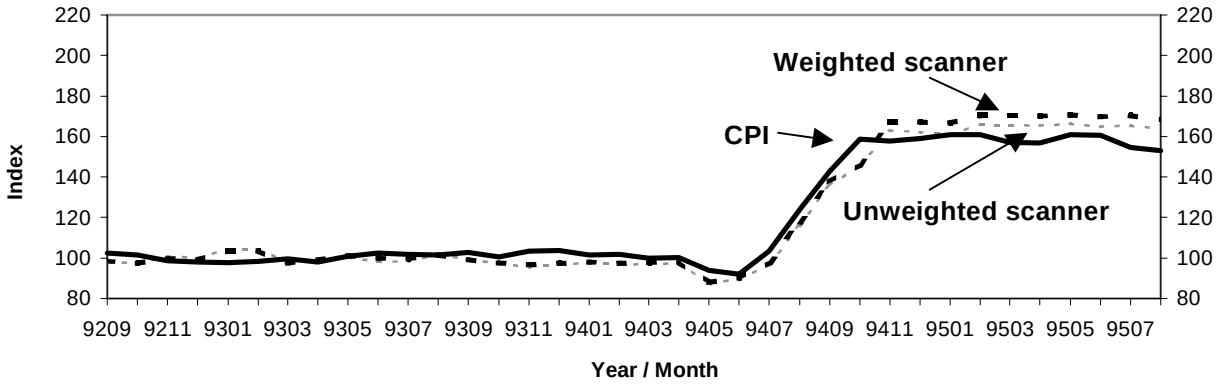
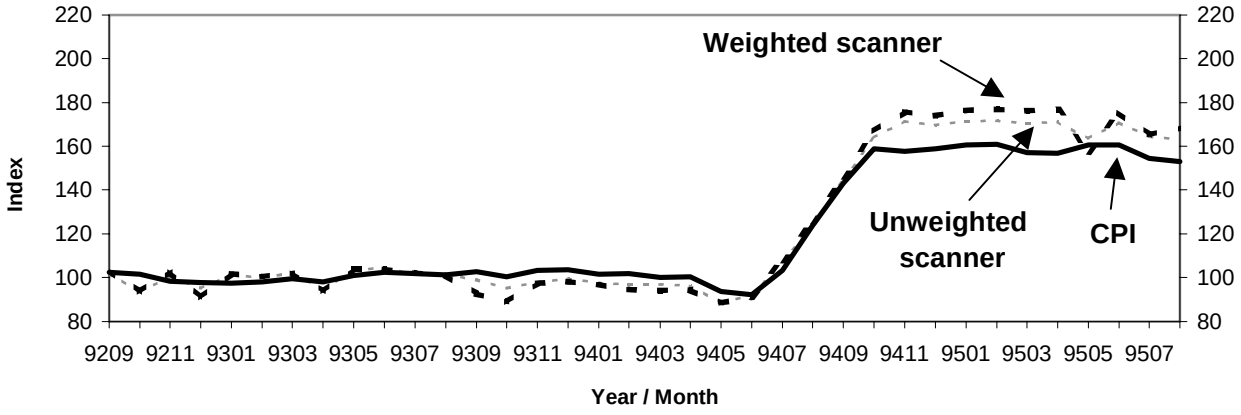


Chart 18: All Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 19: All Brands, All Sizes, Top 14 Outlets, Weighted and Unweighted

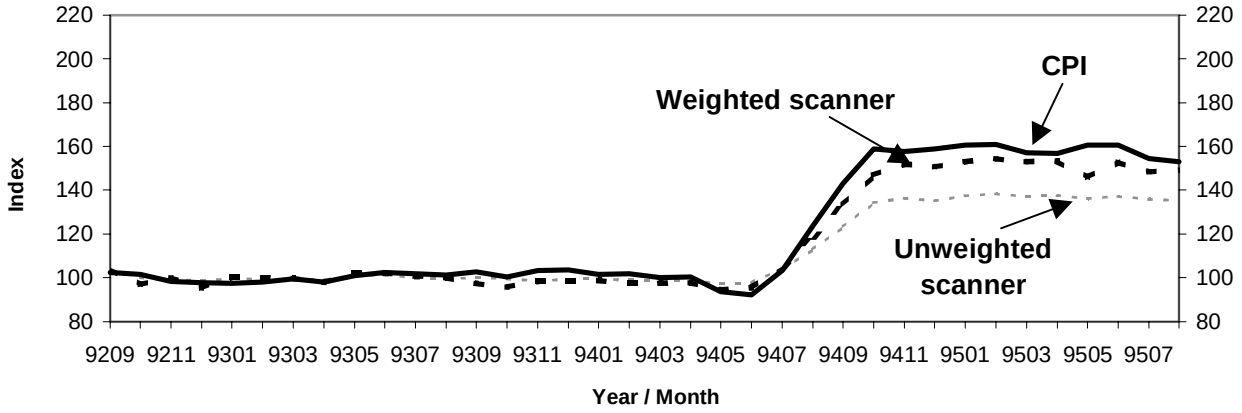
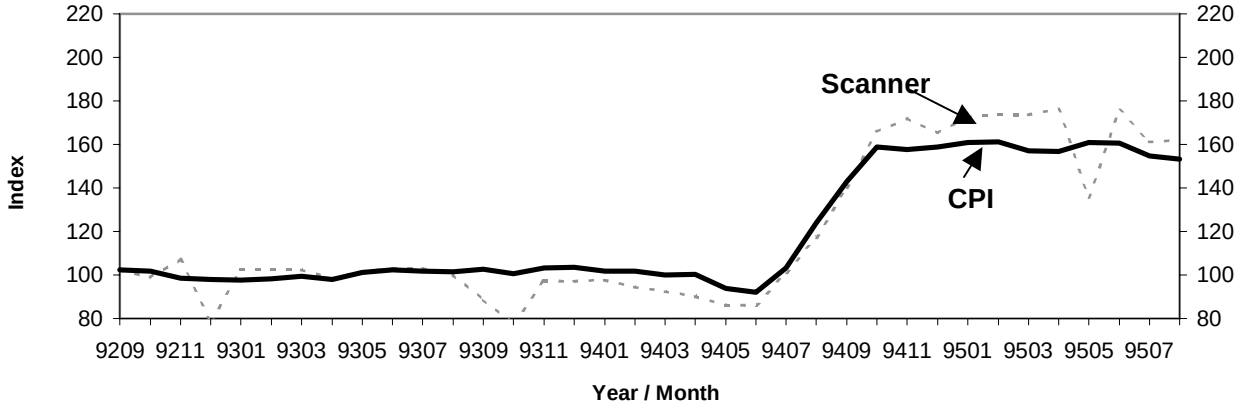


Chart 20: All Brands, All Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 21: All Standard Brands, Standard Sizes, All Outlets, Weighted

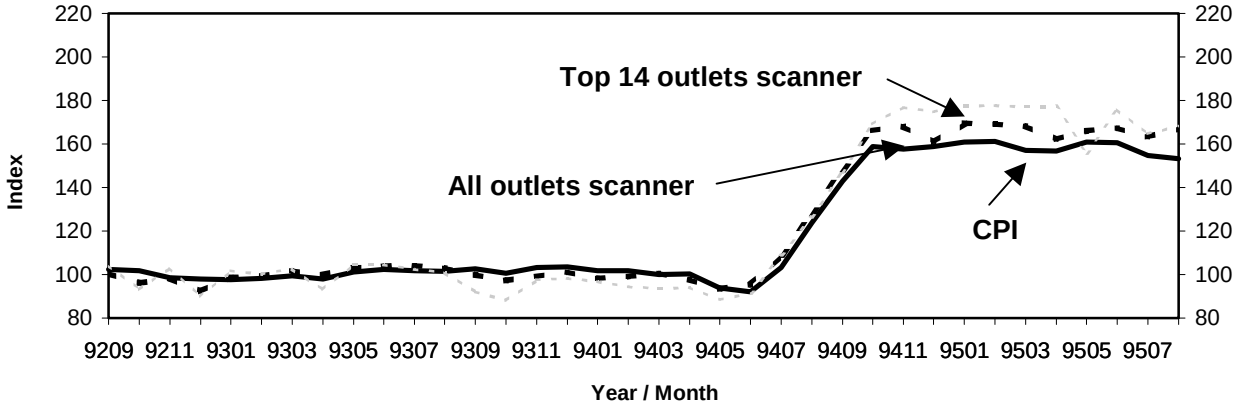
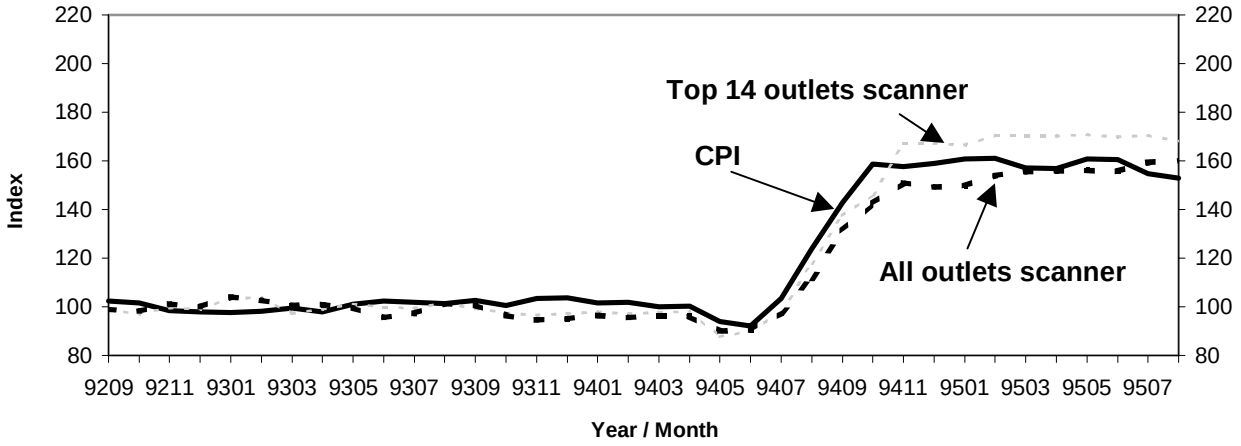


Chart 22: All Non-standard Brands, Standard Sizes, All Outlets, Weighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Toronto (September 1992 to August 1993=100) (continued)

Chart 23: All Brands, Standard Sizes, All Outlets, Weighted

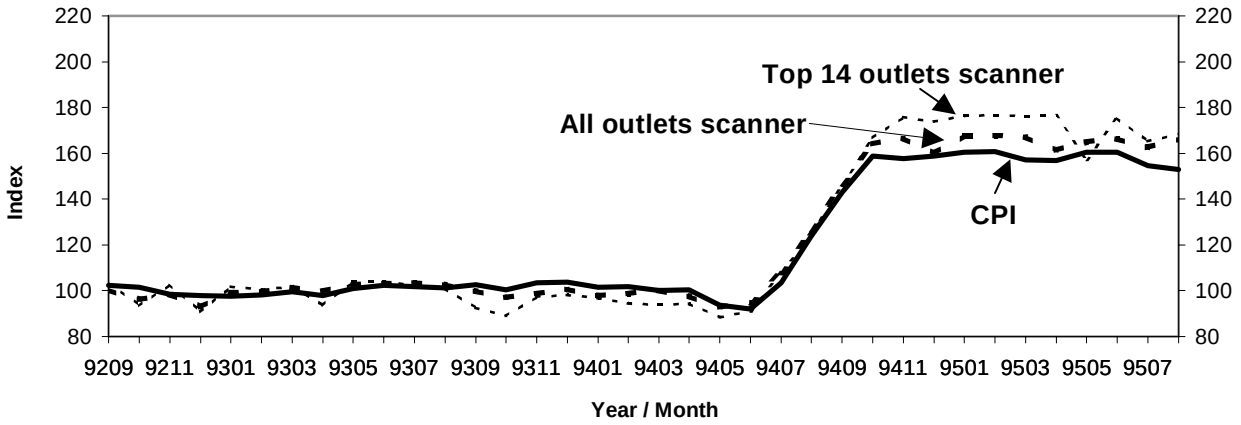
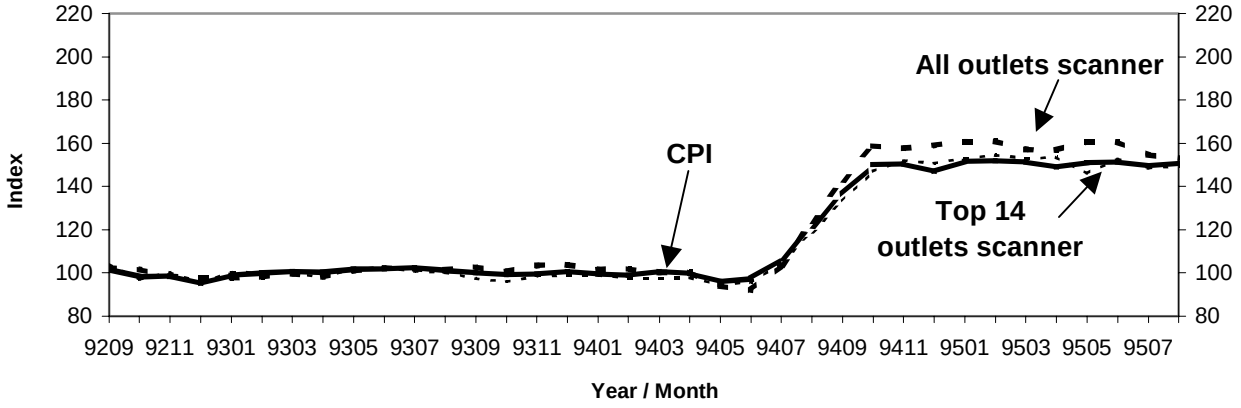


Chart 24: All Brands, All Sizes, All Outlets, Weighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Montréal (September 1992 to August 1993=100)

Chart 25: Standard Brands, Standard Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted

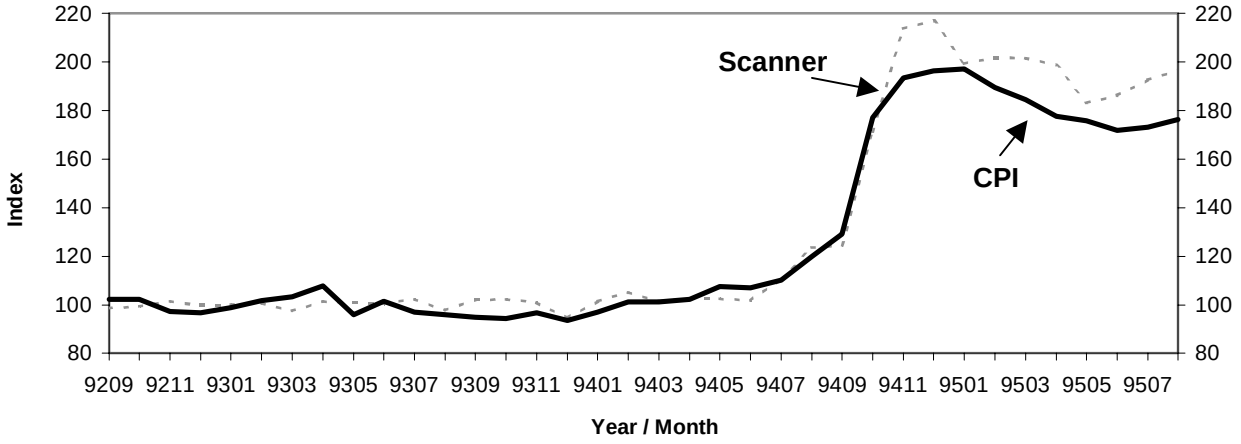
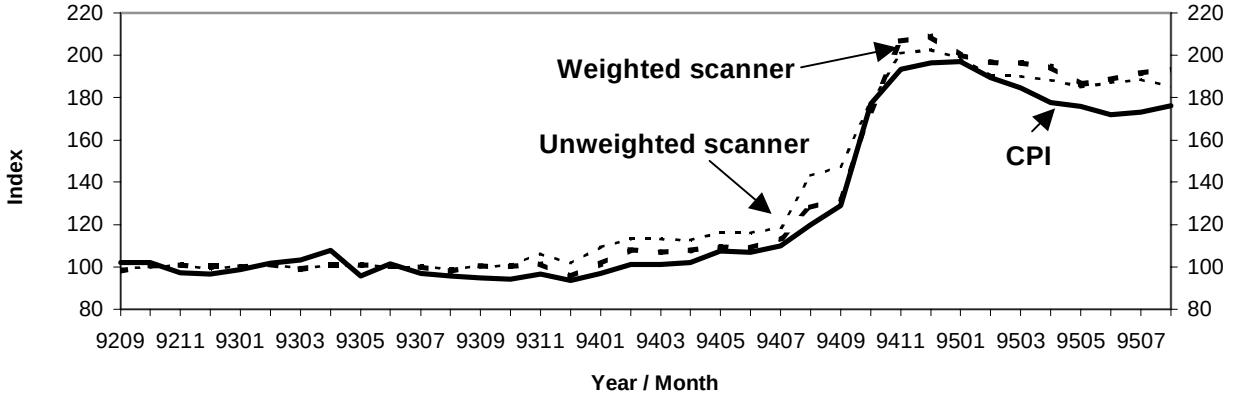


Chart 26: All Standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Montréal (September 1992 to August 1993=100) (continued)

Chart 27: All Non-standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted

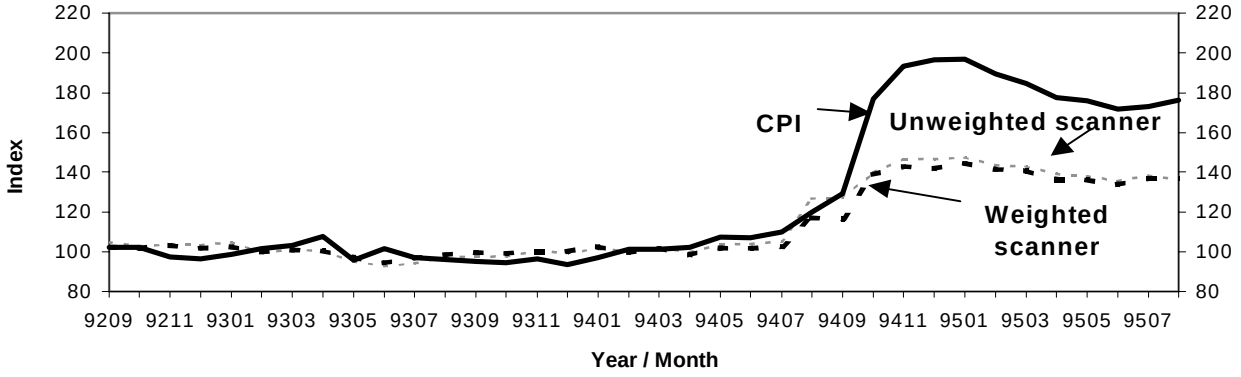
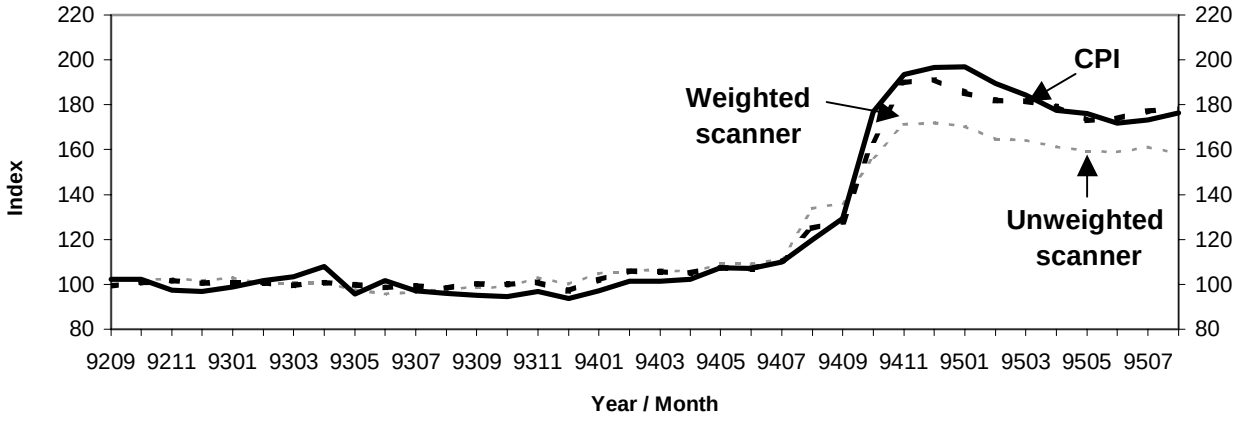


Chart 28: All Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Montréal (September 1992 to August 1993=100) (continued)

Chart 29: All Brands, All Sizes, Top 14 Outlets, Weighted and Unweighted

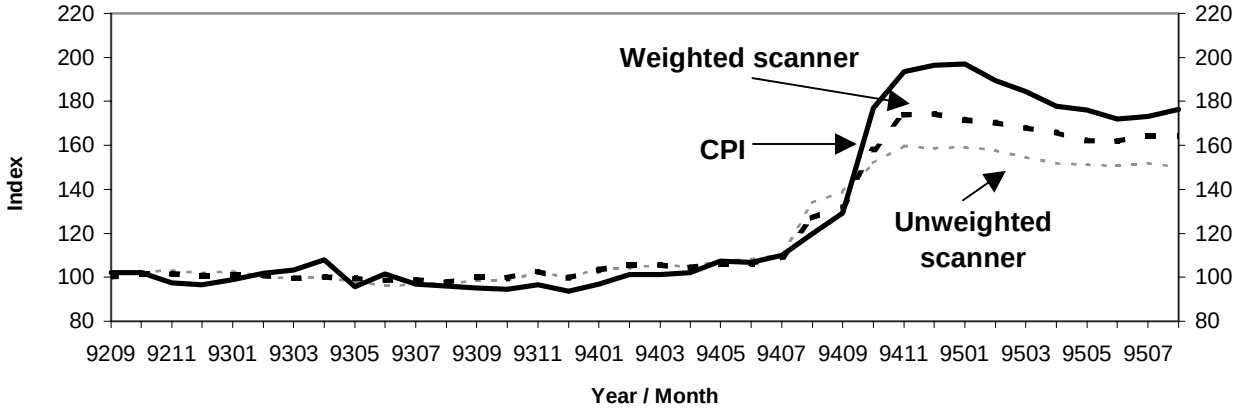
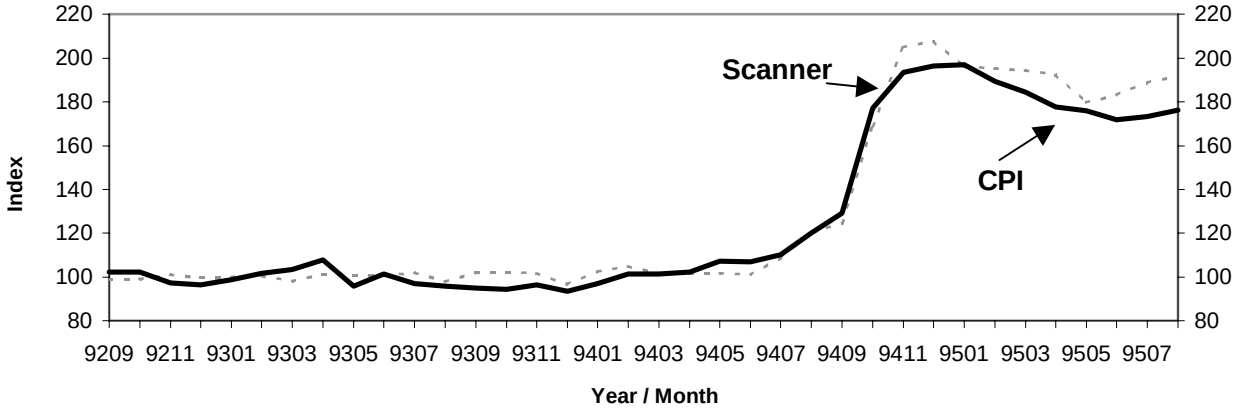


Chart 30: All Brands, All Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Montréal (September 1992 to August 1993=100) (continued)

Chart 31: All Standard Brands, Standard Sizes, All Outlets, Weighted

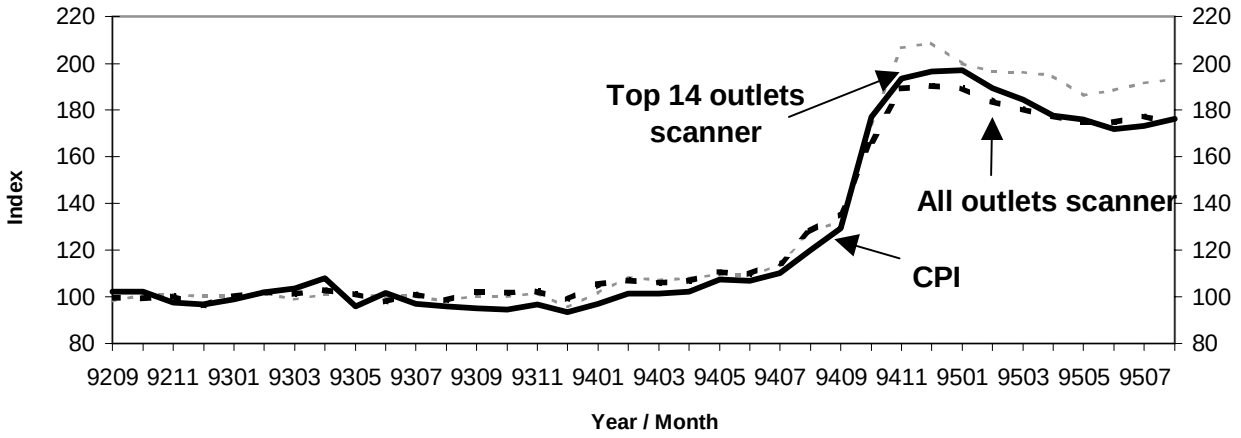
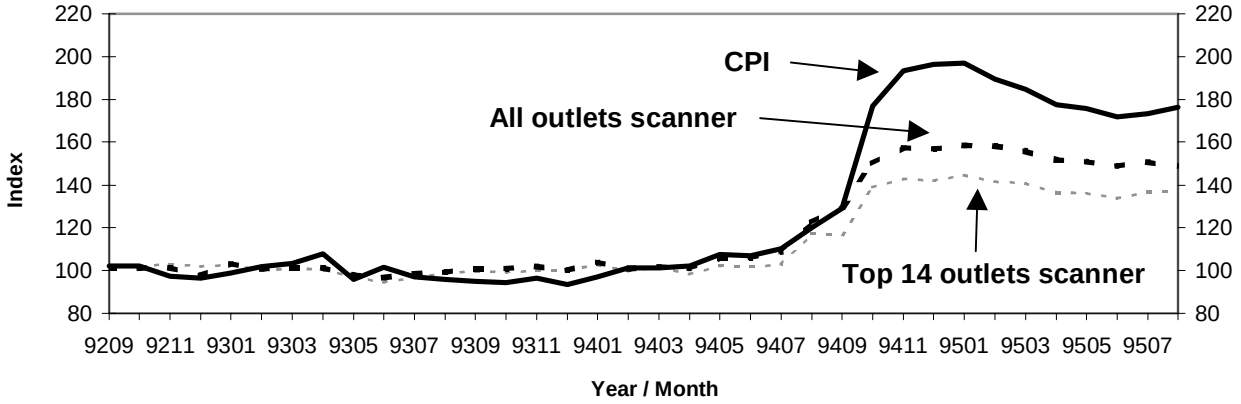


Chart 32: All Non-standard Brands, Standard Sizes, All Outlets, Weighted



Comparison of CPI and Scanner Sub-indexes for Roasted Coffee for Montréal (September 1992 to August 1993=100) (continued)

Chart 33: All Brands, Standard Sizes, All Outlets, Weighted

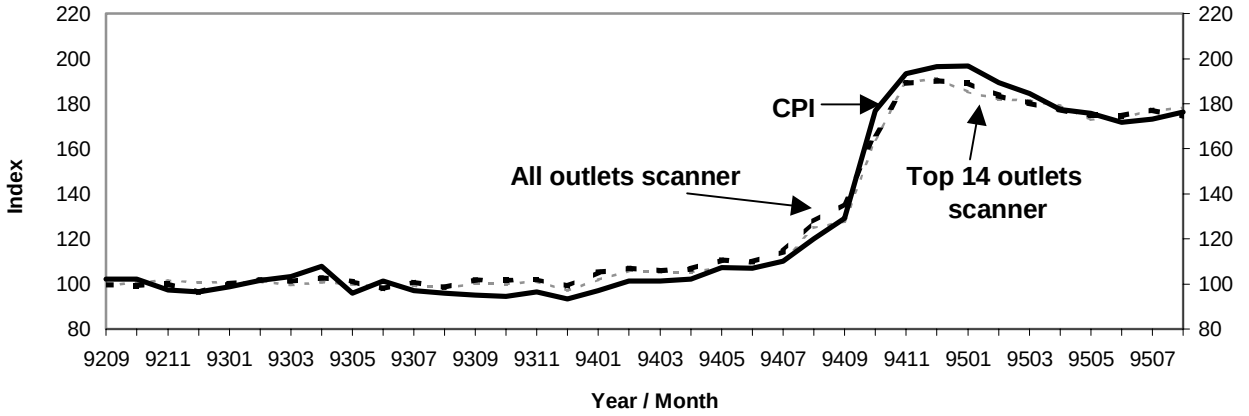
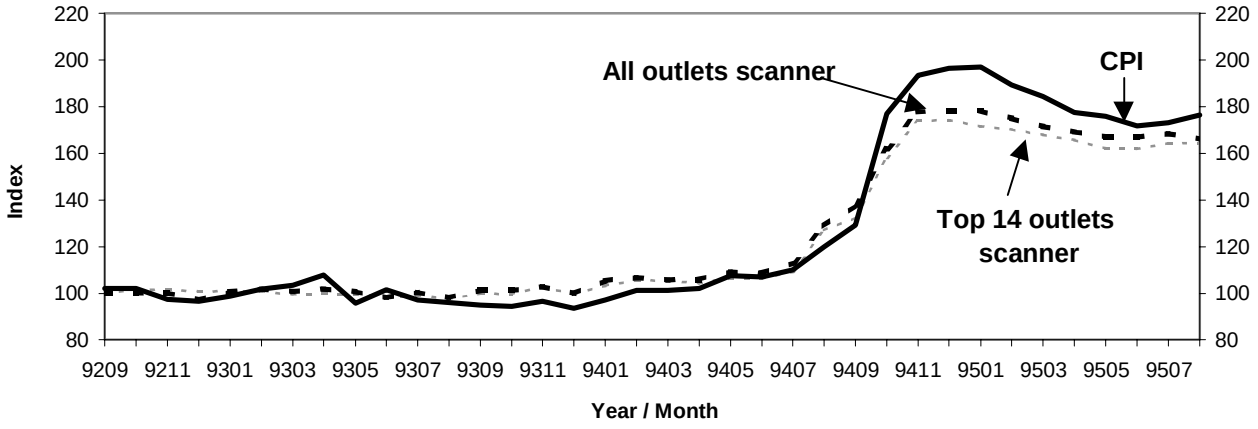


Chart 34: All Brands, All Sizes, All Outlets, Weighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Montréal (September 1992 to August 1993=100)

Chart 35: Standard Brands, Standard Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted

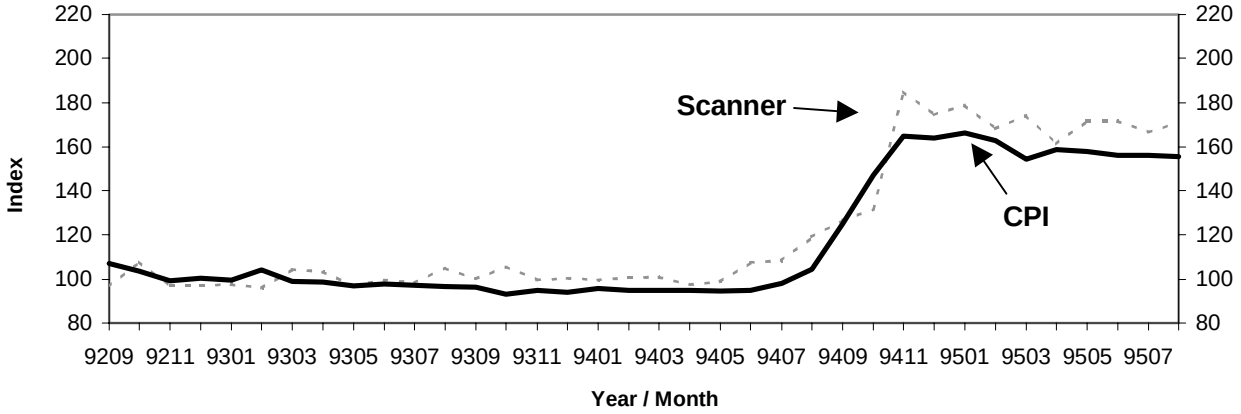
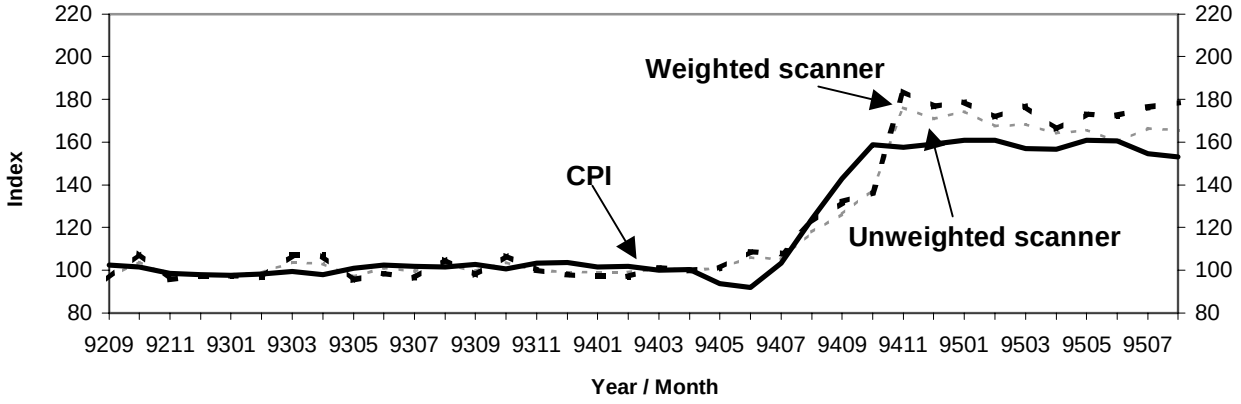


Chart 36: All Standard Brands, Standard Sizes, Top 14 Outlets, Weighted and Unweighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Montréal (September 1992 to August 1993=100) (continued)

Chart 37: All Brands, All Sizes, Top 14 Outlets, Weighted and Unweighted

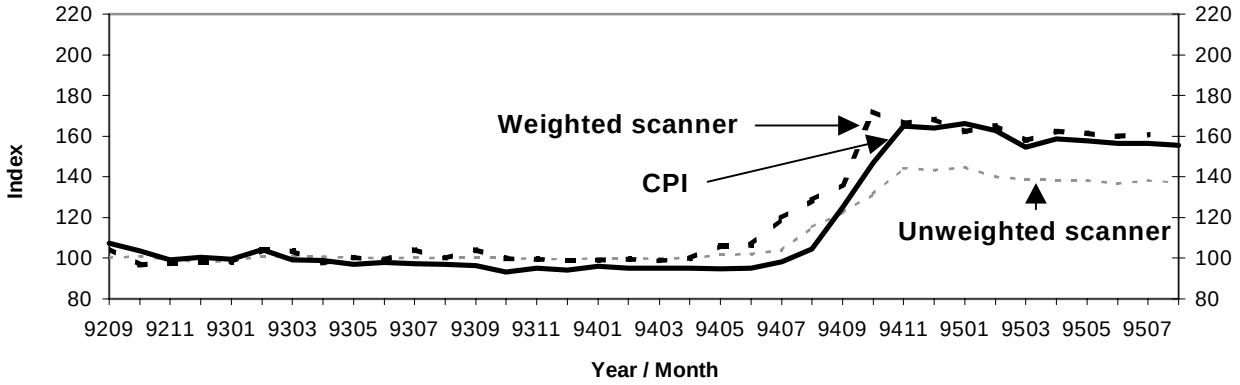
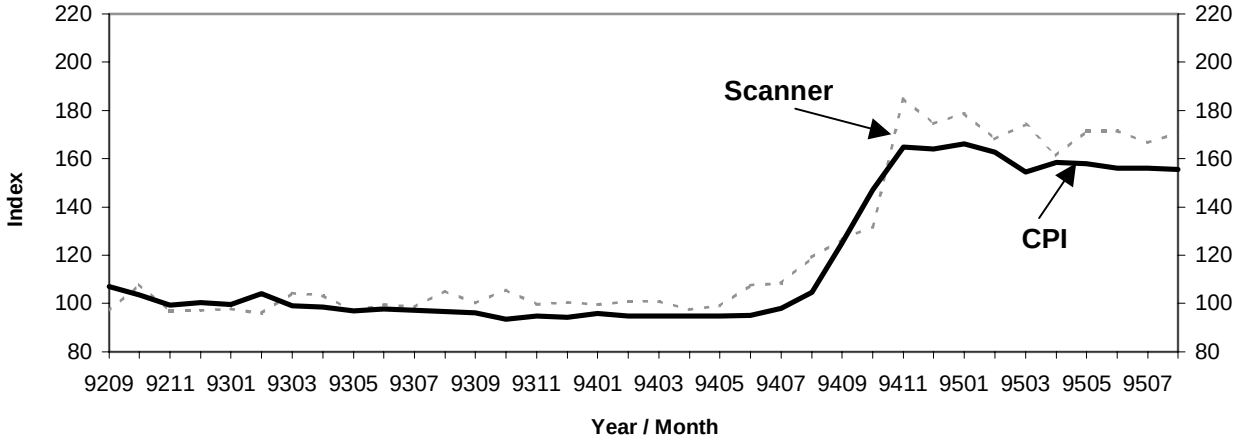


Chart 38: All Brands, All Sizes, Top 14 Outlets, Top Seller per Outlet, Unweighted



Comparison of CPI and Scanner Sub-indexes for Instant Coffee for Montréal (September 1992 to August 1993=100) (continued)

Chart 39: All Standard Brands, Standard Sizes, All Outlets, Weighted

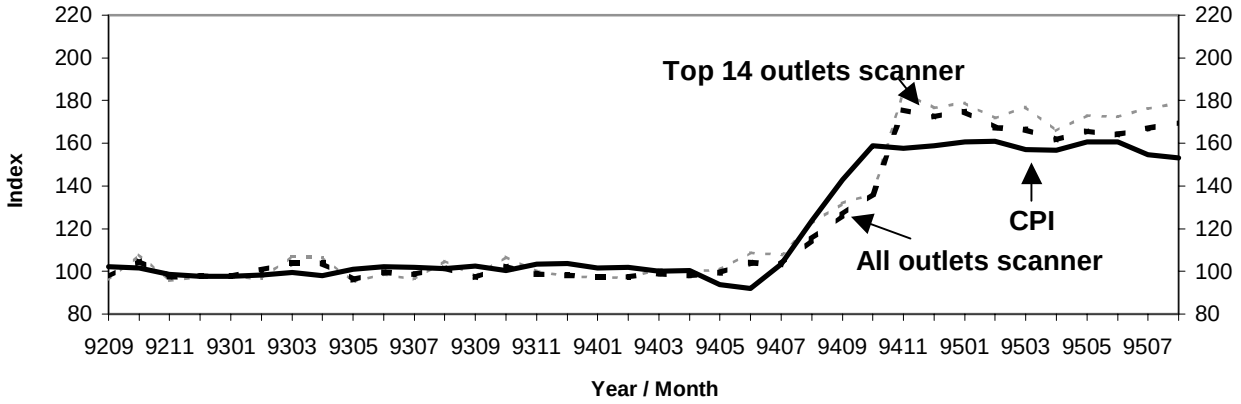
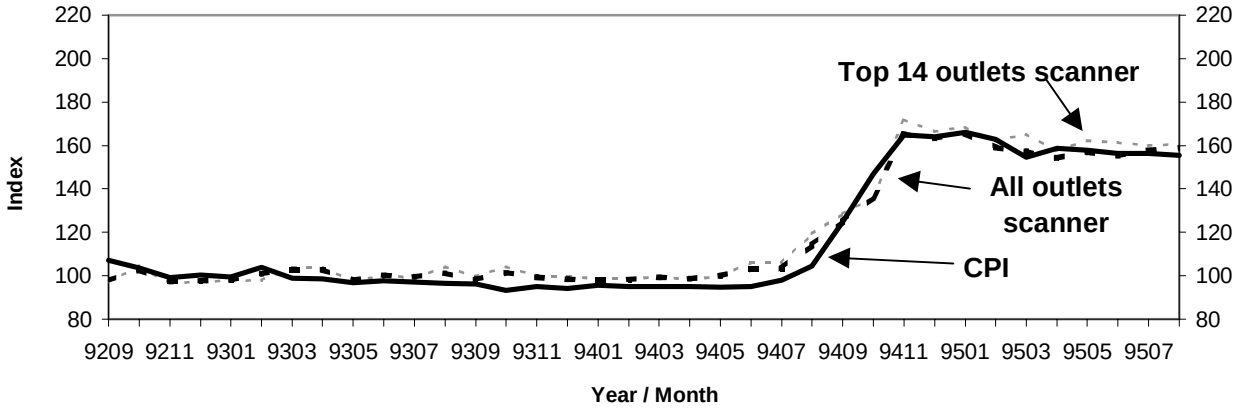


Chart 40: All Brands, All Sizes, All Outlets, Weighted



References

Nakamura, A. and Popkowski-Leszczyc, P. (1995), "Scanner Data: Its Use in Marketing and Potential Relevance for Index Number Making".

Reinsdorf, Marshall (1995), "Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee", U.S. Bureau of Labor Statistics.

Silver, Mick (1995), "Elementary Aggregates, Micro-Indices and Scanner Data: Some Issues in the Compilation of Consumer Price Indices", *Review of Income and Wealth*, Series 41, Number 4, December 1995, pp. 427-438.