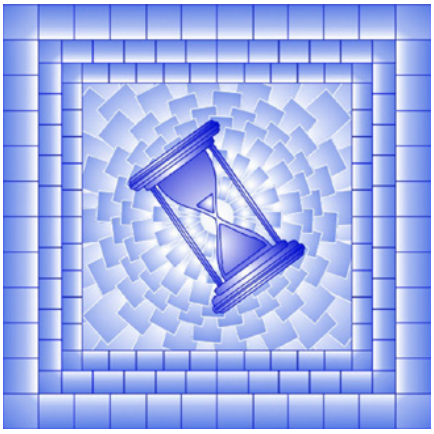
Catalogue no. 62F0014M
ISSN 1706-7723
ISBN 978-0-660-37282-2

A new approach for estimating the Computer Equipment, Software and Supplies Index in the Consumer Price Index



by Lance Taylor and Roobina Keshishbanoosy

Release date: February 17, 2021

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

A new approach for estimating the Computer Equipment, Software and Supplies Index in the Consumer Price Index

by **Lance Taylor** and **Roobina Keshishbanoosy**

The Consumer Price Index (CPI) measures the change in prices of consumer goods and services over time. To accurately reflect trends in the market and consumer behavior, Statistics Canada periodically updates the methods applied to various components of the CPI.

The release of the January 2021 CPI (published on February 17th, 2021) marks the implementation of enhancements to the calculation of the computer equipment, software, and supplies index.

The computer equipment, software, and supplies index represents 0.42% of the 2017 basket, and 4.1% of the recreation, education, and reading index, a major component of the CPI.

Enhancements to the Index

The computer equipment, software, and supplies index in the CPI measures the monthly changes over time in the price of laptops, desktops, monitors, and printers. Previously, Statistics Canada used the household component of the computer and peripherals price index[1,2] as a proxy for their price movements. In order to better measure price change for these items, enhancements have been made to the index including:

- A new, mainly web-scraped, data source covering more product brands;
- Automation of data cleaning to ensure an accurate and more timely index;
- A new statistical model which is updated monthly with new input data and parameters for the measurement of laptop and desktop price changes.

These enhancements will allow for improved coverage of computer equipment, software, and supplies products in the CPI through an increase in the number of observations used to calculate the index, as well as a timelier receipt of data to feed the index. In addition, prices will be collected weekly from retailer websites, supplemented with other surveys and/or sources such as Statistics Canada's retail sales data,[3] market data on brand and product expenditures, and supplementary data collected online.

The New Methodology

The nature of the consumer electronics industry and its rate of technological advancement cause the frequent replacement of outgoing items with new and improved products. Therefore, in order to estimate pure price movement and to control for quality changes,[4] an updated and improved statistical method[5] is used for desktops and laptops to impute monthly prices of incoming and outgoing items.[6] For monitors and printers, which tend to change less frequently, a pure matched model price index will remain in use.

1. Computers and peripherals price indexes (CPPI): Detailed information for November 2020.
2. Table 18-10-0208-01 (formerly CANSIM 331-0013).
3. Monthly Retail Trade Survey (MRTS): Detailed Information for November 2020.
4. Infographic: Measuring Pure Price Change in a Constantly Changing World.
5. Typically, approaches similar to this are considered hedonic methods and use an Oridnary Least Squares (OLS) regression. For example, the Computers and Peripherals Price Index uses an OLS regression to model prices. However, hedonics generally involve the analysis of the contribution of individual product characteristics to price, whereas for price imputation only the predicted price is relevant. As such a parameterized hedonic model is not necessary, and the new methodology will use a random forest.
6. Canadian Consumer Price Index Reference Paper: Chapter 7.

The log of monthly prices for laptops and desktops are modelled as a function of a set of explanatory variables using a random forest[7] algorithm. While each product has a separate model, the explanatory variables used are mostly the same. The variation in log price is thus explained by characteristics such as storage space, storage type, total RAM, type of RAM, display size (for desktops this variable is set to zero if the desktop in question is not an all-in-one desktop), number of CPU cores, CPU speed, CPU brand, GPU brand, product weight, the presence of a touch screen (laptops only), item manufacturer, and item retailer. For categorical variables, categories with low counts or observations with unknown values are grouped together as 'other'.

In the next steps, price relatives[8] for each brand within each product are calculated (for example, laptops from brand X). These are then aggregated to product relatives, which are then aggregated to the final relative. The brand relatives for each of laptops and desktops are calculated as the exponent of the arithmetic mean of differences in log price between periods (using an imputed log price where necessary):[9]

$$I_{t,brand,product} = \exp\left(\sum_{i}^{n_{t,brand,product}} \Delta\, \tilde{p}_{t,i,brand,product} * w_{t,product,i,retailer}\right)$$

where,

$$\Delta\tilde{p}_{t,i,brand,product} = \begin{cases} p_{t,i,brand,product} - p_{t-1,i,brand,product}, \text{ if } i \text{ is in both periods, or} \\ \hat{p}_{t,i,brand,product} - p_{t-1,i,brand,product}, \text{ if } i \text{ isn't in current period, or} \\ p_{t,i,brand,product} - \hat{p}_{t-1,i,brand,product}, \text{ if } i \text{ isn't in previous period.} \end{cases}$$

where,

$p_{t,i}$ is he log observed price of the i^{th} observation in period t,

$\hat{p}_{t,i}$ is the log price imputed using the model estimated in period t,

$n_{t,brand,product}$ is the number of observations in period t,

This can be expressed as:

$$w_{t,product,i,retailer} = \frac{s_{y-1,retailer}}{n_{t,retailer,product} * \sum_{retailer} s_{y-1,retailer}}$$

where $s_{y-1,retailer}$ is the previous year's computer and computer peripherals sales by a given retailer.

This is done so that the total weight of a retailer is held constant for the year, preventing unwanted index movements caused by composition effects.[10]

7. Random forests are a statistical prediction method that operates by constructing a set of decision trees and uses the average result of the individual trees to form the prediction.
8. A measure of the price change from one period to the next, given by the current period constant quality price over the previous period constant quality price.
9. This is equivalent to a weighted geometric mean.
10. In the context of price change measurement, a composition effect would be a change in average prices caused by changes in the sample as opposed to changes in the prices of observations within the sample.

Accurate and reliable price index estimates for monitors and printers are possible without statistical modelling due to lower product churn and a slower pace of technological change for these products. Thus, the brand price relatives for these goods are calculated simply as the exponent of the arithmetic mean of differences in log observed prices between periods:

$$I_{t,brand,product} = \exp\left(\sum_{i}^{n_{t,brand,product}} \Delta p_{t,i,brand,product} * w_{t,product,i,retailer}\right)$$

where:

$$\Delta p_{t,i,brand,product} = p_{t,i,brand,product} - p_{t-1,i,brand,product}$$

For the estimation of each product level index, the geometric mean of each product's brand price relatives is then taken, with their corresponding expenditure weights, to obtain a product level index $I_{t,product}$, i.e.:

$$I_{t,product} = \prod_{brand} I_{t,brand,product}^{w_{t,brand,product}}$$

where the weight of the brand of a product in is the previous period share of price updated expenditures,[11] i.e.,

$$w_{t,brand,product} = \frac{s_{t-1,brand,product}}{\sum_{brand,product} s_{t-1,brand,product}}$$

and,

$$s_{t,brand,product} = s_{t-1,brand,product} * I_{t,brand,product}$$

Up to this point, geometric means are used at the product and product-brand level aggregations in order to account for potential substitution effects[12] between individual product models and between different brands of the same product. However, the final index aggregation is calculated as the sum of current period constant quality expenditures over the sum of previous period constant quality expenditures. The current period constant quality expenditures are calculated as the sum of previous period component expenditures multiplied by the corresponding price relatives. This can be done since it is unlikely that there is substitution between the different product groups (for example, a printer cannot be substituted for a laptop). The ratio of constant quality expenditures defined above can be simplified as the weighted arithmetic mean of product indices, so the final computer equipment, software, and supplies index movement at time t, $I_{t,71010301}$ can be expressed as:

$$I_{t,71010301} = \sum_{product} I_{t,product} * w_{t,product}$$

11. Canadian Consumer Price Index Reference Paper: Chapter 6.
12. As an item gets more expensive relative to another comparable item, consumers may switch to this substitute.

where $w_{t,product}$ is the product's share of computer equipment, software and supplies expenditures during time t so that:

$$w_{t,product} = \frac{s_{t-1,product}}{\sum_{product} s_{t-1,product}}$$

and $s_{t,product}$ is given by the sum of price updated product-brand expenditures;

$$s_{t,product} = \sum_{brand} s_{t,brand,product}$$

for all t.

Example

In order to illustrate the index calculation process described above, we present an example using a set of fictitious data.

Say at time *t* for brand *m* of product *p* there are 3 observations *a*, *b*, and *c*. Observations *b* and *c* are observed in both the current and previous period, whereas *a* enters the CPI sample this period. Observations *a* and *b* are sold by retailer *r1*, and *c* is sold by retailer *r2*. The sales of computer products by *r1* is equal to the sales of computer products by *r2*. The current and previous period prices of the observations are given by table 1.

Table 1
Fictitious price observations for brand m of product p

Observation	Previous period log price	Current period log price	$\Delta p^*_{t,i,b,p}$
a	7.1 (imputed via model)	7.0	-0.1
b	6.0	6.0	0.0
c	6.9	7.0	0.1

If we use a model to impute the prices of non-continuities for product *p*, and the weight of *a*, *b*, and *c* are 0.25, 0.25, and 0.50, then the product-brand index movement $I_{t,b,p}$ is calculated as $\exp(-0.1 * 0.25 + 0.0 * 0.25 + 0.1 * 0.50) = 1.025$. If we rely on a pure matched-model approach for product *p*, then *a* has no price relative and is thus excluded, and the weight of *b* and *c* are 0.50, and 0.50. The product-brand index movement $I_{t,m,p}$ would then be calculated as $\exp(0.0 * 0.50 + 0.1 * 0.50) = 1.051$.

Now, say at time t the set of product-brand indices are calculated for product *p*, which include movements for brands *m1*, *m2*, and *m3*. The current and previous period price movements, price updated expenditures,[13] and weights are given by table 2.

Table 2
Set of fictitious expenditures, weights, and price movements for some brands

Brand	$I_{t-1,b,p}$	$I_{t,b,p}$	$s_{t-2,b,p}$	$s_{t-1,b,p}$	$w_{t,b,p}$
m1	0.900	1.000	50	$0.900 * 50 = 45$	45/365
m2	1.000	1.025	100	$1.000 * 100 = 100$	100/365
m3	1.100	0.950	200	$1.100 * 200 = 220$	220/365

13. Canadian Consumer Price Index Reference Paper: Chapter 6.

The product index movement $I_{t,p}$ is then calculated as $1.000^{45/365} * 1.025^{100/365} * 0.950^{220/365} = 0.976$.

Lastly, for final index movement aggregation, say at time there is a movement for each of laptop, desktops, monitors, and printers. The current period price movement, previous period expenditures, and weights used are given by table 3.

Table 3
Fictitious weights and price movements for computer equipment, software, and supplies products

Product	$I_{t,p}$	$s_{t-1,p}$	$w_{t,p}$
Laptops	0.976	365	365/600
Desktops	1.010	120	120/600
Monitors	0.980	55	55/600
Printers	1.050	60	60/600

The final index movement $I_{t,71010301}$ is calculated as

$$0.976 * \frac{365}{600} + 1.010 * \frac{120}{600} + 0.980 * \frac{55}{600} + 1.050 * \frac{60}{600} = 0.990.$$

In summary

With the incorporation of web-scraped data and methodological updates, the new computer equipment, software, and supplies index constitutes an important enhancement towards the measurement of price change for goods which are critical to the digital economy. As of the release of the January 2021 CPI this new index replaces the use of the computers and peripherals price index as a proxy. Due to the changes in methodology and data, users are advised not to make year over year index comparisons until a full year has passed since implementation. If users need more information they may contact the authors: Lance Taylor (lance.taylor2@canada.ca), or Roobina Keshishbanoosy (roobina.keshishbanoosy@canada.ca), or the info unit (statcan.cpddisseminationunit-dpcunitedediffusion.statcan@canada.ca).