The Integration of Web-Scraped Data into the Clothing and Footwear Component of the Consumer Price Index

Release date: February 19, 2020
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.
The Integration of Web-Scraped Data into the Clothing and Footwear Component of the Consumer Price Index

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 in consumer behaviour, Statistics Canada periodically reviews and updates the data sources and methods applied to various components of the CPI.

The release of the January 2020 CPI (published on February 19, 2020) marks the integration of web-scraped data and sample enhancements into the sub-indexes of the clothing and footwear component.

Web scraping is a process through which information is gathered and copied from the web.¹ There are numerous advantages to replacing prices collected in the field² with publicly available data from official websites. This approach reduces response burden on businesses, saving valuable time and resources for businesses while continuing to provide high-quality data in a cost-effective manner. It is also an efficient means of acquiring large volumes of information at greater frequency, and to produce timely and accurate statistics.

The clothing and footwear component represents 5.17% of the 2017 CPI basket (at basket link month), and is comprised of four sub-component indexes: clothing; footwear; clothing accessories, watches and jewellery; and clothing material, notions and services. Beginning with the January 2020 CPI, for selected retailers, some prices for these indexes are now being web-scraped and are no longer collected in the field. As a result of this change, Statistics Canada interviewers are collecting approximately 10% fewer prices in clothing and footwear retail stores every month. At the same time, coverage of the products sold by these retailers has improved, and the overall number of prices collected on a monthly basis has increased.

More frequent price collection

Field-collected price quotes for the clothing and footwear component are recorded once a month by interviewers at various locations across the country over a period of two weeks. The new approach will allow for weekly price collection over the entire month. Web-scraped prices may now be combined with field collected data, resulting in a more robust data set.

Greater variety of products and number of prices

While the retailers in the CPI sample remain the same, the accessibility of price information on the Web and the efficiency of web scraping will make it possible to collect near census-level price data for a greater variety of products and will improve product coverage across retailers.

The release of the January 2020 CPI is an important milestone for the CPI as it marks the first use of web-scraped data in the calculation of the clothing and footwear price indexes. Over time, field collected prices for the clothing and footwear sub-indexes will continue to be replaced by web-scraped data on a per-retailer basis. This transition will take place in stages. As the CPI evolves, Statistics Canada will continue to explore the use of alternative data sources to meet Canada’s ongoing information needs.

¹ For more information of how Statistics Canada collects and uses existing data including web scraping please visit Using new and existing data for official statistics: Where does data come from? and Frequently asked questions on using new and existing data for official statistics.
² Field collection is the practice of sending Statistics Canada interviewers to a specific geographic setting to collect and record data (in this case, brick-and-mortar retail stores).

Statistics Canada – Catalogue no. 62F0014M