

Digitalisation in the CPI: Canadian case studies

1. Making sense of online price data from Web scraping
2. A new approach to calculating the telephone services price index

Daniela Ravindra, Statistics Canada

Second Meeting of the UN Committee of Experts on Business and Trade Statistics

11-13 June 2019 - New York



Delivering insight through data for a better Canada



Statistics
Canada

Statistique
Canada

Canada

1. Making sense of online price data from Web scraping

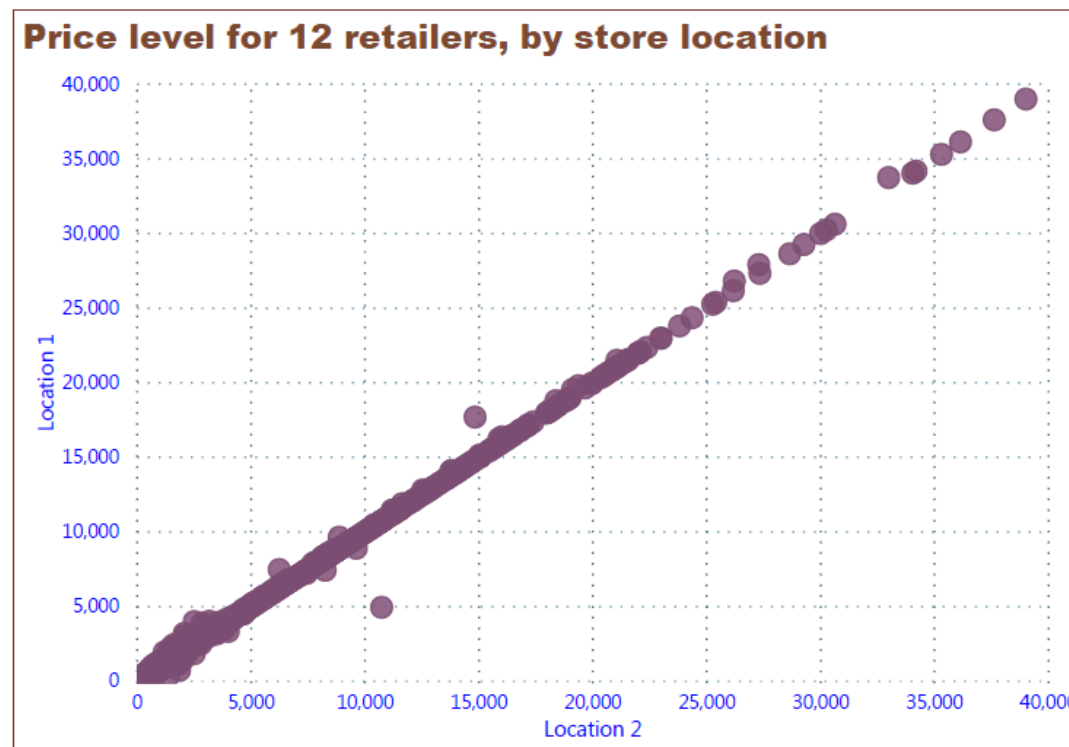
Retail e-commerce in Canada was valued at \$18 billion in 2018, a 15% increase in value from 2017

- With rapid growth in e-commerce, there is an urgency to quantify the growing importance of this sector on Canadian consumers.
- Web scraping represents an unprecedented opportunity to collect large amounts of price information from major retailers who are offering an increasing number of products online.
- At the same time, we need to take steps to ensure we truly understand web scraped data.
- We also need to determine how this data could be used within the CPI.



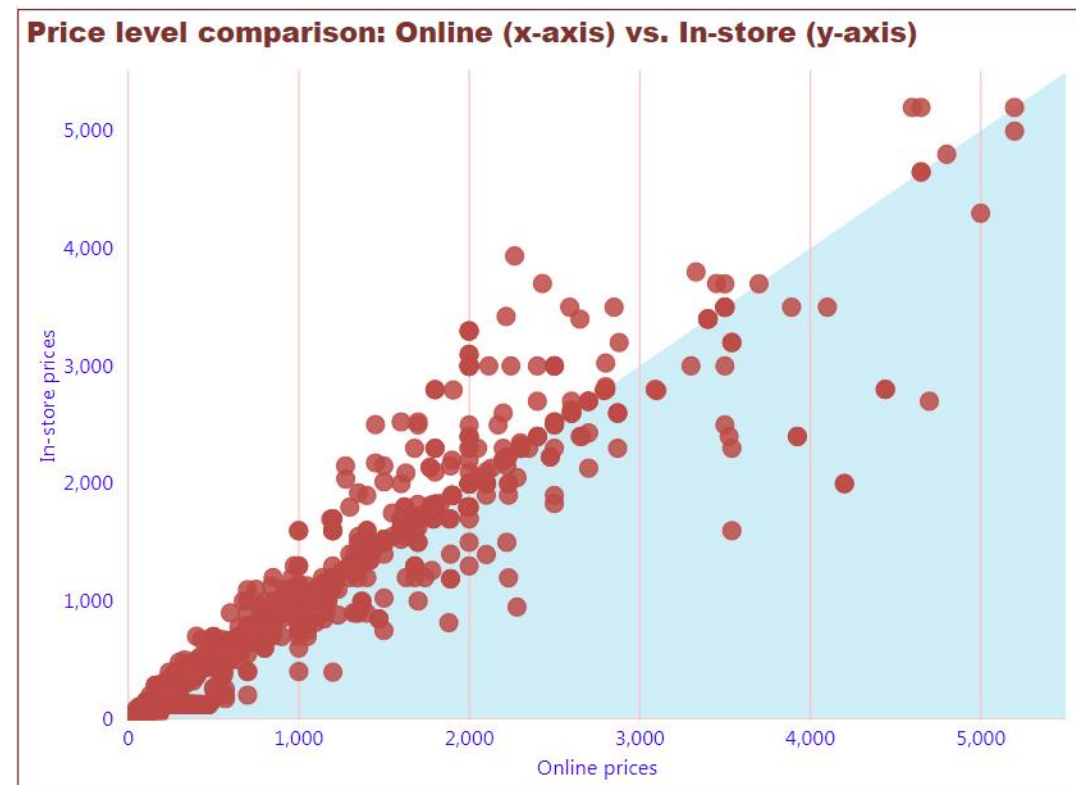
Finding 1: Retailers seem to apply national pricing to goods sold online

- Online prices don't appear to change based on the physical location of a consumer.
- One retailer had a small percentage of products priced differently across geography, but this could be due to lengthy scraping times.
 - Some sites take multiple days to scrape
- Different studies including Cavallo (2018) also found that online retailers, including Amazon, tend to post the same price for each good, regardless of the location of the buyer
- If national pricing exists online, would we miss geographical differences present only in physical stores?



Finding 2: No apparent systematic price difference between online and in-store prices for matched products

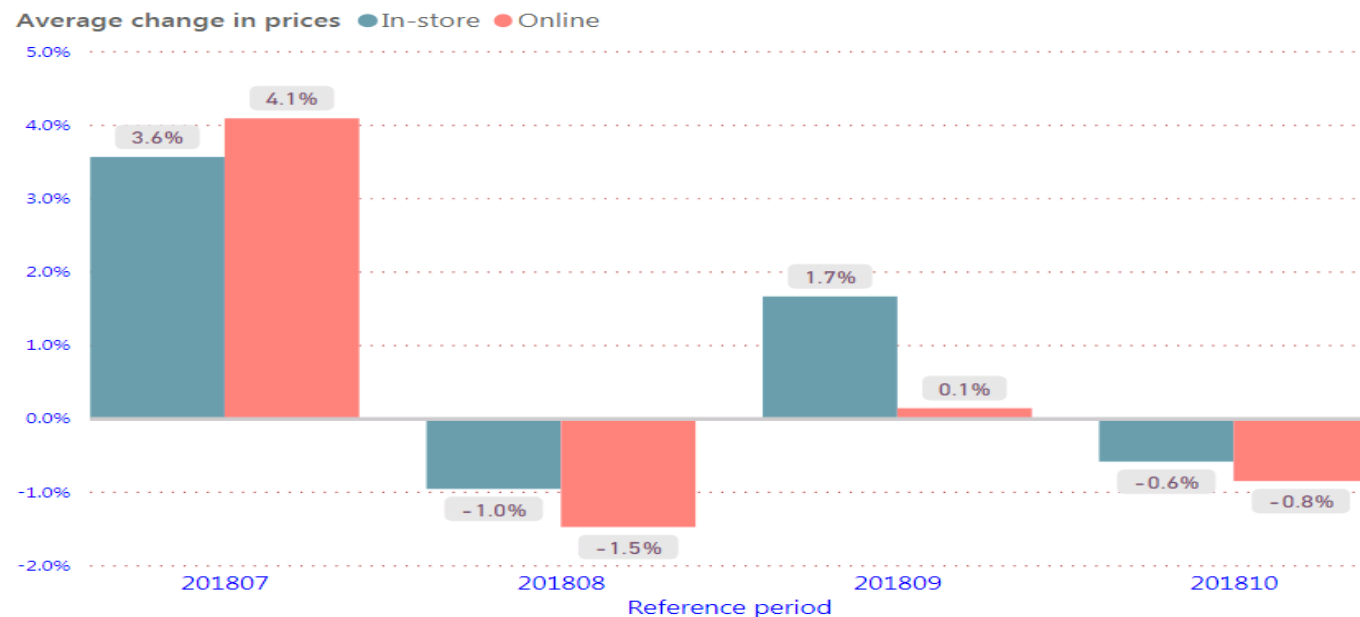
- Identifier SKUs are used to match products online and in-store.
- The comparison collected prices for 1,816 matched products and controlled for week of collection.
- While variation exists, results tend to show that prices for goods sold online are similar to those captured in-store.
- This is confirmed by other studies, such as Cavallo (2017), which find little difference for multi-channel retailers between their online and in-store prices.



Finding 3: Average price movements observed online are generally similar to those observed in physical stores for matched products

- For the most part, price movement magnitudes and directions for the 1,816 matched products are very close.
- A different study conducted by Statistics Canada using prices collected online and in store has reached a similar conclusion.

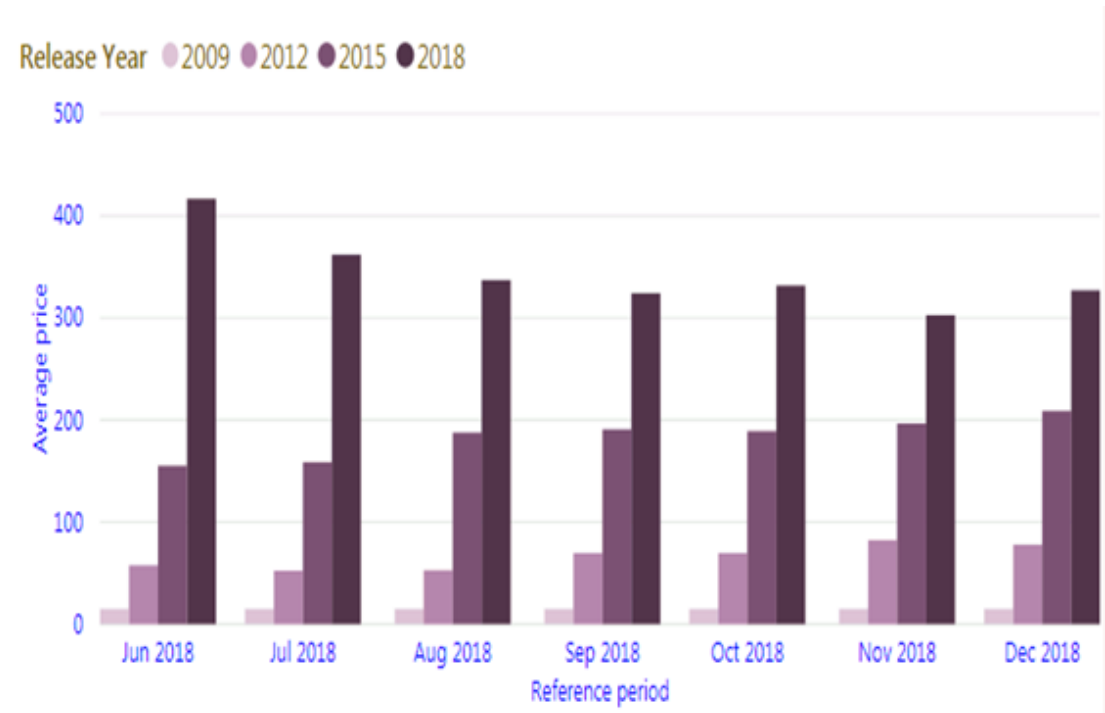
Average change in prices comparison: Online vs. In-store



Finding 5: For some categories, there seems to be a relationship between product age and price

- For categories such as electronics and appliances, newer products tend to be pricier than old products.
- Might the lower cost of “digital shelf space” explain the presence of older products in the online marketplace?

Examples of Average Prices by Release Year for Certain Electronics and Appliances

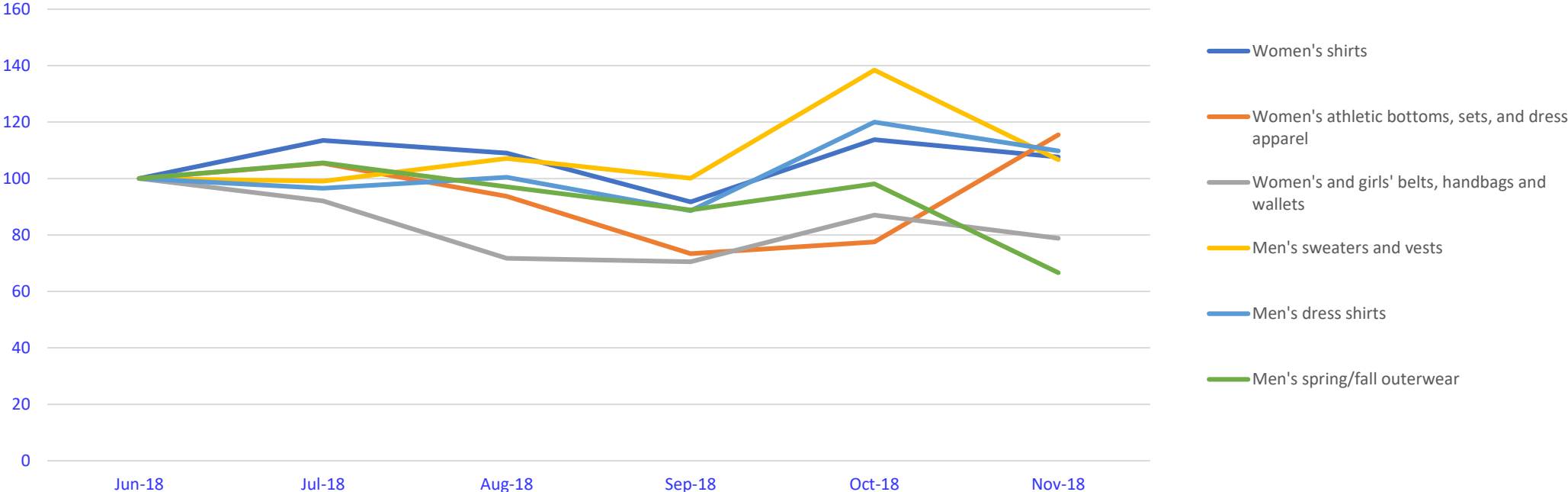


Clothing is our starting point for implementation of web scraped data

- Current methodology involves matching items over time and adjusting for quality of individual products.
- The new approach will leverage rich descriptive data from web scraping to classify clothing products into groups of similar products.
 - Individual products may enter or leave a group over time, but group definitions remain constant.
 - An average price is then the geometric mean of all products in a particular group.
- For now, product groupings are coarse, but Machine Learning will eventually be used to classify large numbers of web scraped products into finer groupings.

Preliminary examples of web scraped indexes for clothing items

Jevons Index,
June 2018=100



Note: the graph portrays indexes for a single retailer in a single Canadian geography.

Current approaches, challenges and the future

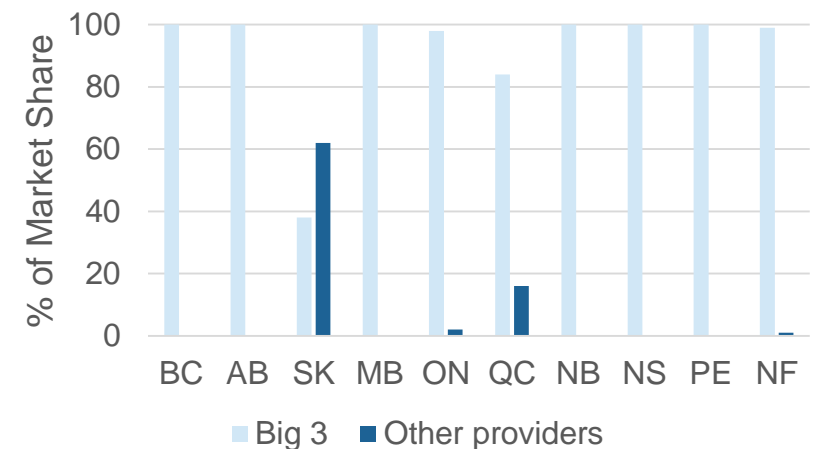
- Web scraping represents a growing source of price and product information that is increasingly relevant to the CPI.
- Statistics Canada adopted an “inform and proceed” approach to web scraping without resorting to enforcement via the Statistics Act.
- There are technical challenges associated with this informal relationship
 - Chiefly automated blocking and unannounced site changes
- And further work is required on methodological issues, such as sample selection and product classification.
- To successfully integrate web scraping into the CPI, we also need to ensure simultaneous progress in other tools, such as Machine Learning, in order to classify large amounts of data.

2. A new approach to calculating the telephone services price index

The Canadian wireless market has 3 main providers controlling 92% of the market

- The wireless market in 2017:
 - Total retail revenues of \$24.5 billion
 - Average revenue per user of \$65.33
 - LTE networks covered 99% of the population
 - 92% of Canadians had access to LTE-Advanced network in 2017, compared to 83% in 2016
 - Average mobile data subscriber used over 2GB of data per month in 2017, an increase of 30% from 2016.
- According to the Canadian Radio-television and Telecommunications Commission (CRTC), total revenue of Canadian cellular services has grown by an average of 5% since 2013

Market Share: Big 3 versus other providers



Source: Communications Monitoring Report 2018, Canadian Radio-television and Telecommunication Commission (CRTC)

The bundling of cellular services and devices is common practice for Canadian telecommunications providers

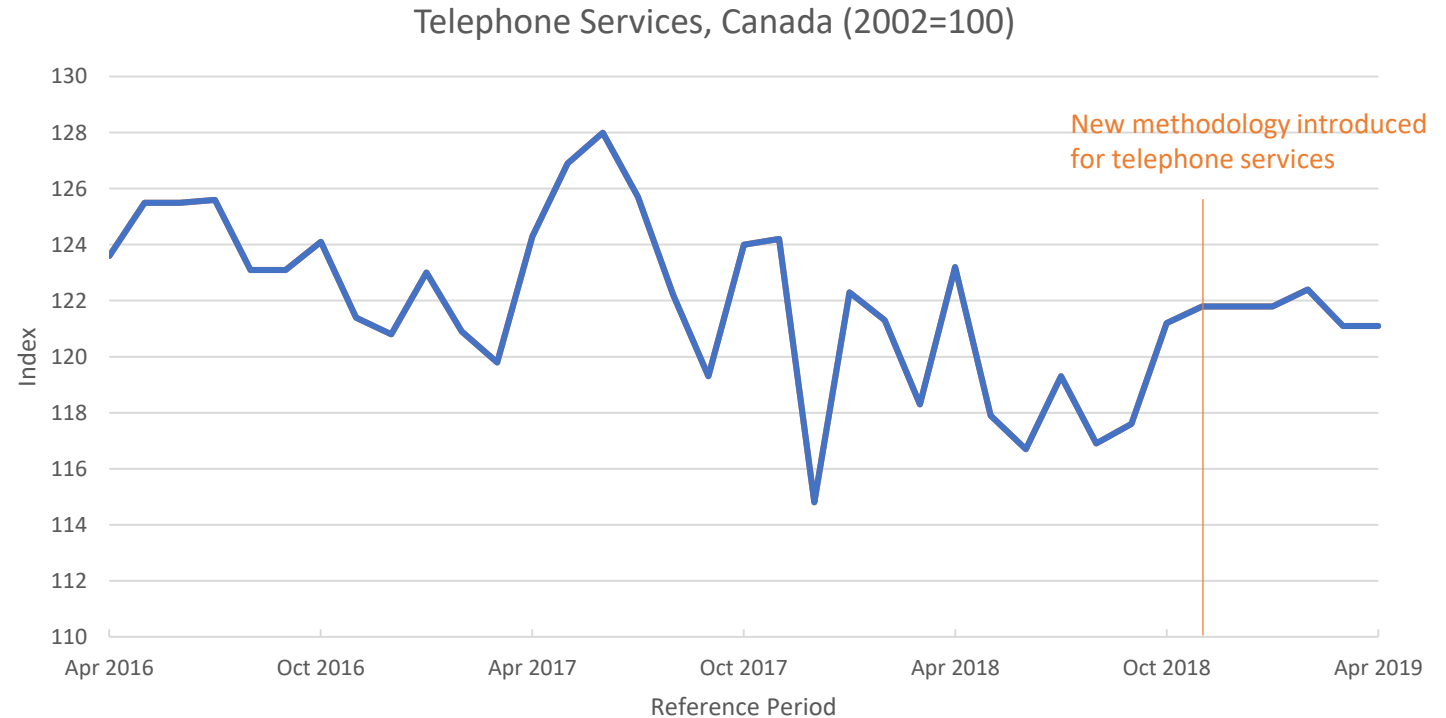
- Unlike most countries, the majority of Canadians don't purchase their phones outright from the service provider.
 - Customers may pay for part of the phone, while the other part is built into the contract (monthly payment for services)
 - Customers may pay for bundled services offered by the big 3
- Until recently, the Canadian CPI bundled cellular services and phone prices.
- This bundling led to:
 - volatility in the cellular services index, driven by changes in on-contract smartphone prices;
 - price change in the cellular services index even if the service plan prices did not change.

A new approach introduced in November 2018 separates the device from the cellular service package:

- Pricing cellular service plans:
 - user-profile approach – matching plans to pre-set customer profiles
 - include only the service plan prices, as advertised by the provider
 - device subsidy – or portion of the monthly plan price that goes towards paying for the phone – is now removed
- Pricing devices:
 - matched model approach – sample of 4 phones from each provider in each province
 - accounts for the lower up-front cost and regular monthly payment for the device
 - quality adjustment is based on a substitution approach, whereby half of the price difference between the current and previous phone model is attributed to price change and the other half is attributed to change in quality
- For information consult: [Methodology - Estimating the telephone services index of the CPI](#)

The impact of ‘unbundling’ on the telephone services price index:

- Improved stability and interpretability of price movement in cellular services
- More accurately captures price movements of smartphone hardware
- Release of new phone model no longer has any impact on cellular services



Looking ahead: questions and challenges

- How are other countries dealing with bundling in telecommunications? Canadian approach is similar to UK and New Zealand.
 - Both apply a user-profile approach for cellular services and have a separate index for hardware prices
 - New Zealand also uses a matched model approach for pricing hardware
- How to deal with non-contract plans?
- Methods of quality adjustment for products/devices with rapid technological improvements.
 - Should similar quality adjustment substitution methods be applied to other products such as smartwatches, tablets, etc. (i.e. half of the price difference between the current model device and previous version is attributed to price change and the other half is attributed to change in quality)
 - What about hedonics – which quality characteristics should be included in the model?

THANK YOU!

For more information,
please visit
www.statcan.gc.ca



#StatCan100