Measuring intangibles at scale with the Cloud Computing Services Producer Price Index (CCSPPI)

1. Introduction

Statistics Canada developed the Cloud Computing Services Producer Price Index (CCSPPI) to measure price change for cloud computing services provided in Canada. It classifies according to three commonly-identified sectors: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). Cloud services are held primarily to be a digital data delivery service to substitute on-premises data hosting and other related IT services.

The purpose of the index is to improve the coverage of the digital economy and deflate nominal value of cloud computing services output in the Canadian System of Macroeconomic Accounts in the future.

Producer Prices Division (PPD) currently publishes the annual Informatics Professional Services Prices Index (IPSPI) covering four North American Product Classification System (NAPCS) codes:

- NAPCS 482: Software and Software Licensing
- NAPCS 643: Custom software design and development services
- NAPCS 774: Computer systems design and development
- NAPCS 75111: Data processing, hosting, and related services

The IPSPI sample of establishments (units) is drawn from NAICS[1] 54151, 51121, 58121 respectively, with 58121 representing units primarily engaged in Data processing, hosting, and related services. Units may however provide services classified according to the four NAPCS listed above. The result is a sample that covers the cloud computing industry indirectly, without proper demarcation. Furthermore, the disparity present in the structure of the larger and smaller cloud service providers (CSP) required a more flexible, judgemental approach in sampling.

Furthermore, in light of the flexible usage of 'cloud' terminology, and in the interest of a comprehensive, intuitive, and adaptable scope, we created an index that covers the IaaS, PaaS and Systems SaaS public cloud sectors, the latter being a sub-sector of the SaaS sector. In the absence of a unique NAICS code for the cloud computing services industry, the SaaS sector, in its entirety, is believed to include a multitude of CSP that may not necessarily fall under the scope of the index. As CSPs already offering IaaS and PaaS tend to offer System SaaS cloud services, this sub-sector was included to capture the full range of the industry in a common framework.

2. Data Source

The data used in index calculation are obtained on a monthly basis for each CSP's advertised service offerings associated with a data center in Canada. Each CSP has at least one data center located in Canada. For companies with multiple data center locations in Canada, prices observed for the larger data center were used. Prices are collected in Canadian dollars where available, whereas prices collected

[1] North American Industry Classification System

in U.S. dollars are converted using the exchange rate monthly average[2]. Price information was aggregated using revenue / market share information obtained from industry business intelligence. Initially, the aim was to collect as much data as possible and establish a methodological framework. Over time, automatic data collection methods were introduced as they became available or apparent, namely API querying and HTML scraping.

3. Methodology

3.1 Sample

The sample was determined in consultation with industry partners and data obtained from external providers, with the aim of measuring price changes primarily for "hyperscalers"[3]. In addition, the sample was selected to adequately cover provision of services in the sectors of Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Systems Infrastructure Software-as-a-Service (SaaS-SIS), in contrast to existing price index literature on the industry focusing primarily on the IaaS component. Generally, the relevant characteristics priced in the provision of IaaS services are limited to a stable set of variables conducive to conventional price index models.

3.2 Basket of Services

The basket of cloud services is defined using publicly-available information obtained from the website of each sampled CSP and is updated on a biennial basis. Typically, CSPs advertise pre-packaged solutions to their customers which consist of bundled cloud services, which are referred to as cloud products. Products advertised on CSP websites consist of a series of sub-products whose functions are defined by a set of sub-product instances.[4]

The basket was determined judgementally by consulting information found on CSP websites combined with external revenue information for market categories. A representative portfolio of cloud services provided by the company was created, by analyzing their frequency in advertised solutions packages, aided where necessary with industry knowledge. These are usually grouped according to their service use offerings, e.g. cloud storage solutions, cloud blockchain solutions and cloud internet-of-things (IoT) solutions. Because the structure is unique to each CSP, comparing similar products across CSP is not always possible. Products are therefore compared at a general market category level, e.g. "Data Management Software".

Offering service level-guarantees, such as discounts applied if a quality standard is not met, has become a norm for the industry, and influences service users' choice of service provider. As the index aims to

[2] "Monthly average foreign exchange rates in Canadian dollars, Bank of Canada" (2020), data for Canada by Statistics Canada & Bank of Canada (Table: 33-10-0163-01 (formerly CANSIM 176-0081)

[3] A hyperscaler is defined as a CSP that can scale provision of cloud services perfectly to corresponding increases in demand, with no limit. According to revenue data from external sources, these providers represent a majority of the industry.

[4] A sub-product instance is a classification subset found below the sub-product level such that the only difference in cost to the user across sub-product instances within a sub-product is the difference in the pure price component. For example, sub-products offered in multi-level volume-based pricing tiers incorporate several sub-product instances into one grouping.

maintain coverage of the range of services offered by each CSP, the terms set out by each sample CSP's service level agreement (SLA) were recorded and measured over time similarly to other products in the basket. Since the SLA represents quality of service for other products rather than a stand-alone product with its own revenue, the relative change in the numerical terms of the SLA (i.e., discount level if terms are not met) is applied to the set of products belonging to a given company.

3.2.1 Simplifying Multi-faceted products

Across CSPs collected, virtual machine (VM) products are included as part of the sample. However, there are many ways a VM can be configured in order to meet a user's needs. Users have the choice to purchase a VM instance with a specific operating system from a family of VMs with various pricing structures. The total number of prices obtained per CSP ranges from 1,000 prices observed to over 30,000. In order to analyze a simplified yet encompassing set of prices, a VM instance was randomly sampled from each VM family and an available operating system randomly assigned to each instance. Analysis of previously advertised price changes suggests that price change applies to all members of a VM family when a given CSP changes prices for their VMs. It is therefore important to exhaust the range of all VM families, but not necessary to include every VM instance and operating system within the sample. With automated collection techniques, the burden of large datasets is alleviated, and this technique is no longer prioritized.

3.3 Price Index Aggregation

To measure price changes in this "industry"[5], an index is constructed for each CSP at the product-level and then aggregated to form a CCSPPI on a quarterly basis.

In the absence of weight information at the lowest-level, as products within CSP often vary significantly from each other by the units of measurement and price-determining characteristics advertised, an approach that treats all sub-product prices as equal in aggregation was preferred over a transaction-based pricing scheme.

3.3.1 Limitations

Because the product mix and its observed characteristics within a CSP is heterogeneous, conventional price measurement techniques, such as elementary price indexes and hedonic quality adjustments face limitations. Over the first 12-month data collection cycle, there has been a comparable amount of observed nominal price changes (changes in advertised prices month-over-month) as changes to the service quality characteristics of each product. Under the framework of a Jevons price index, new sub-products added at the product-level will not affect a change in the nominal price of cloud services, though they may imply an improvement in the range of capabilities a client may face compared to the previous period. Existing hedonic imputation or adjustment techniques are not widely applicable due to the multitude of varying units of measure and the absence of a total price at the product level.

3.3.2 Calculation of Elementary Aggregates

At the first stage, an average price of sub-product instances is determined at the sub-product level. For example, the cost of $0.02 per GB of data transfer between two regions would be a sub-product for a generic object storage product. The associated quality in this case, for example, is the daily,

[5] While 'industry' is used here, notionally cloud services are treated under the index on a NAPCS basis.

unadjustable limit on the number of transfer requests. Sub-product level prices are then aggregated using the Jevons price index' formula to create a product-level price index for each CSP. For values in the collected data considered to represent qualities, the Jevons index formula is applied to the quality-adjusted relatives for a given product, with the inverse price relative formula applied to data classified as qualities. Because of the non-linearity observed in changes to quality-type data points, those greater than 1 are log-transformed to better quantify the diminishing impact of changes in quality-type variables in relation to the overall characteristics of the product[6]. Quality-adjusted relatives are then applied to the corresponding price index for that product by multiplication. These different initial level aggregations are referred to as elementary price indexes.

Let I_k^j represent the price for product j and CSP k,

$$I_k^j = \prod_{i=1}^{n} \left(\frac{p_{ik1}}{p_{ik0}}\right)^{\frac{1}{n}} \prod_{g=1}^{m} \left(\frac{q_{gk0}}{q_{gk1}}\right)^{\frac{1}{m}}$$

where there are *n* unique price-type sub-products i, and *m* unique quality-type sub-products *g* within each product j for CSP k.

3.3.3 Compilation of upper level aggregated indexes

The elementary price indexes are then aggregated to the product category[7] in which they are classified using an unweighted geometric average.

$$ProductCategoryRelative_k^m = \prod_{j=1}^{Jtot} I_k^j$$

where *m* designates the product category.

The product-category indexes are then aggregated across CSP using the Young index formula to create an overall index. The weights used for this aggregation are derived from the proportions of revenue / market share for each cloud market category and for each CSP, as obtained from external providers.

For this case, the total monthly CCSPPI is calculated by:

$$CCSPPIMonthlyRelative = \sum_k \sum_m w_k^m * ProductCategoryRelative_k^m$$

where w_k^m is the relative importance for product category *m* and CSP *k*. The monthly relative is then averaged on a quarterly basis to form the quarterly CCSPPI.

[6] For example, the limit imposed on the number of transfer requests increased from 10,000 at time t_0 to 50,000 at t_1, this would represent a quality relative at time t_1 of (ln 10,000)/ln(50,000) = 0.851.

[7] A product category corresponds to a level of classification where similar cloud services across CSP can be grouped according to their end use

4. Results

Table 1 CCSPPI Results

Quarter	Index, 2020Q1=100	CAD/USD	Index, 2020Q1=100 (no FX)
2020 Q1	100	100	100
2020 Q2	100.68	101.76	99.61
2020 Q3	97.78	97.77	99.12
2020 Q4	96.02	95.67	98.61
2021 Q1	94.82	93	99.03
2021 Q2	93.02	90.17	98.88

The price trend observed for the period was heavily influenced by the exchange rate between Canadian and American dollars. When removing the impact of the exchange rate, we observed a price trend that lined up with expectations formed prior to the development of the index, that of prices lightly falling over time.

While observed price or quality changes at the sub-product level were often[8] markedly higher than the overall index trend (+/- 50%) for that product, approximately two thirds of all price or quality changes represented a price change smaller than 10%, many of which can be attributed to frequent incremental VM hourly price changes. Keeping in mind that a product's price index is composed of price and quality indexes, out of 45 products measured from February 2020 to June 2021, an average of 6.4 products showed some total price index change on a monthly basis. Furthermore, price changes for products classified as 'SaaS' tended to exhibit fewer price changes than those classified as 'IaaS' or 'PaaS'. In Table 2 when virtual machine products were excluded from this analysis, only 4.7 products exhibited price index (price or quality) change on average each month.

In terms of the concentration of product price index changes, the top third of products classified according to price change frequency accounted for approximately 70% of all product price index changes on a monthly basis. The frequency and concentration of observed product price index changes led to the production of the index on a quarterly basis.

Table 2 Frequency of product-level price index changes

Reference Period	Total products showing overall index change	Total products showing a price change	Total products showing a quality change	Products showing an index change (excluding VMs)	Total number of products
2020 Q1	5.7	4.7	3.0	3.7	45
2020 Q2	7.0	4.7	3.0	5.0	45
2020 Q3	6.0	5.0	3.0	4.7	45
2020 Q4	6.0	4.7	3.0	4.3	45
2021 Q1	10.0	8.0	2.3	7.7	45

[8] 16.8% all price or quality changes

5. Next steps

The CCSPPI was created with three objectives. First, it was designed to be an experimental index that would cover a new range of services.

Second, we hope that it will be used as a deflator by the Canadian System of Macroeconomic Accounts (CSMA). It remains to be seen which role the CCSPPI will have in deflation as discussions with the CSMA in regards to their needs in deflating the digital economy are in early stages.

Third, given the overlap between CCSPPI and IPSPI, a portion of the IPSPI index may be replaced by the new index and hence reduce response burden.

Development work on this index is not complete however. As cloud computing is continually evolving, the goal, then, is to "future-proof" the CCSPPI as much as possible to ensure it can serve as a deflator down the road. Regular environmental scans aim to keep notions of industry movements up to date. Currently, work is underway to improve gaps in automated data collection to further improve timeliness and coverage of possibly additional CSPs.