Measuring intangibles at scale &

Report on experimental development of the Cloud Computing Services Producer Price

Jean-Philippe Tardif, Statistics Canada

Delivering insight through data for a better Canada

Why measure this industry?

- Overall importance of digital economy cannot be ignored:
  - 39% of Canadian businesses used cloud computing in 2019 (Statistics Canada Daily, Monday November 23, 2020)
  - Cloud spending as share of total IT spending projected to grow to 14.2 % in 2021, up from 9.1% in 2020 (Gartner, 2020)
  - COVID-19 pandemic and increase preference for remote working will increase the importance of the cloud for Canadian enterprises
- Currently Informatics Professional Services Price Indexes capture this industry to the extent that CSP are in an in-scope NAICS (Data hosting, Software Publishing, Computer Systems Design)
- Experimental index covering new sector in hopes of eventually deflating the industry

Fundamental questions that helped determine the scope and process

- Sample & Basket selection: **where to start?**
- Data collection of large, heterogeneous datasets: **how do we make this efficient?**
- “in the cloud” vs. “the cloud”
  - Software like Office 365, Gmail, Teams are pure “SaaS”, whereas Cloud Computing is the rentable infrastructure behind it, **so where to draw the line?**
- How should the **service be quality adjusted?**
  - Some collected information are quality measures, but how should it be used when adjusting the prices?

Scope & Methodology

- Sample targets “hyperscalers” with data centers in Canada
- Aim to cover IaaS, PaaS and SysSaaS sectors, in contrast to existing price index literature on the industry focusing primarily on the IaaS component
    - IaaS has a more stable set of characteristics than SaaS which complicates the price index calculation
- Use industry revenue data and iterative analysis of company product offerings to determine most relevant products for basket formation

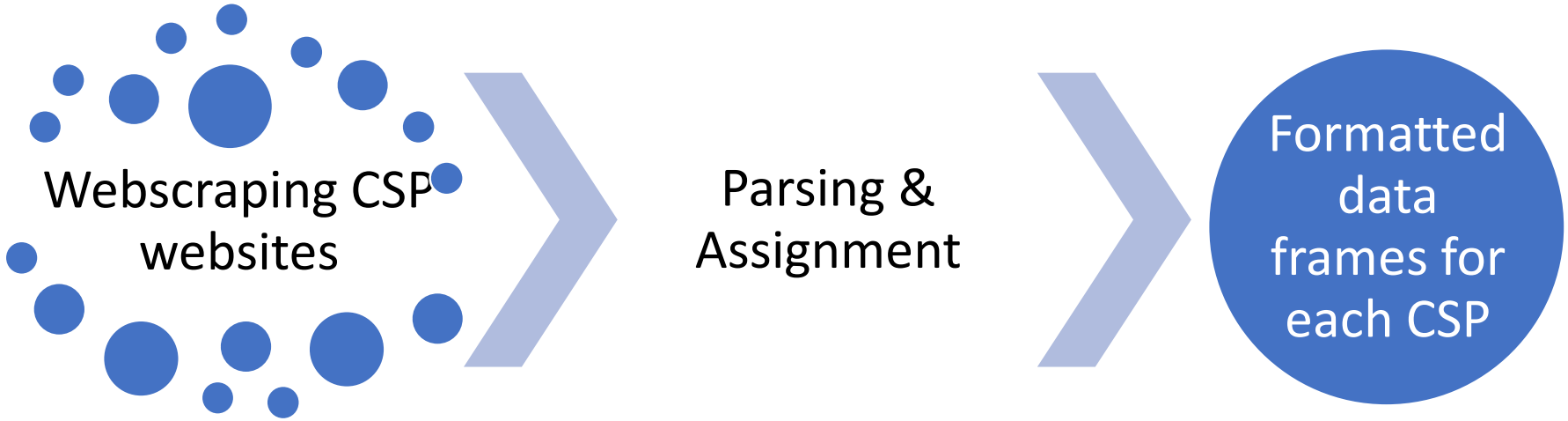
“Why don’t you just measure a typical bill over time?”

- Multitude of possible billing combinations may lead to bias; treating sub-products under a Jevons requires weaker prior knowledge to proceed with making an index

Concept: Prices & Quality measures

- Pricing pages contain two types of variables: Prices & Quality measures
- **Prices** can be understood conventionally
- Examples of **Quality measures:**
  - *Allowable memory usage*
  - *Number of transactions/queries (Service quotas)*
  - *Computing power*
  - *Dimensions of a virtual machine*
- More generally, the guiding principle for classification during data collection:
  - *If the number is positively correlated with overall price of the product, then classify as price, otherwise the data point is a "quality" of the service*
- Log-transformation of qualities + US-CAD exchange where necessary

Processing



- Approximately 200 products, with +300 sheets of product information collected
  - API utilization/HTML scraping is the predominant method currently to collect data
- Only the most representative products are selected to create a ‘basket’ for each CSP
- Each selected product is scanned and catalogued; every sub-product is tracked over time

Total index

weighted aggregation

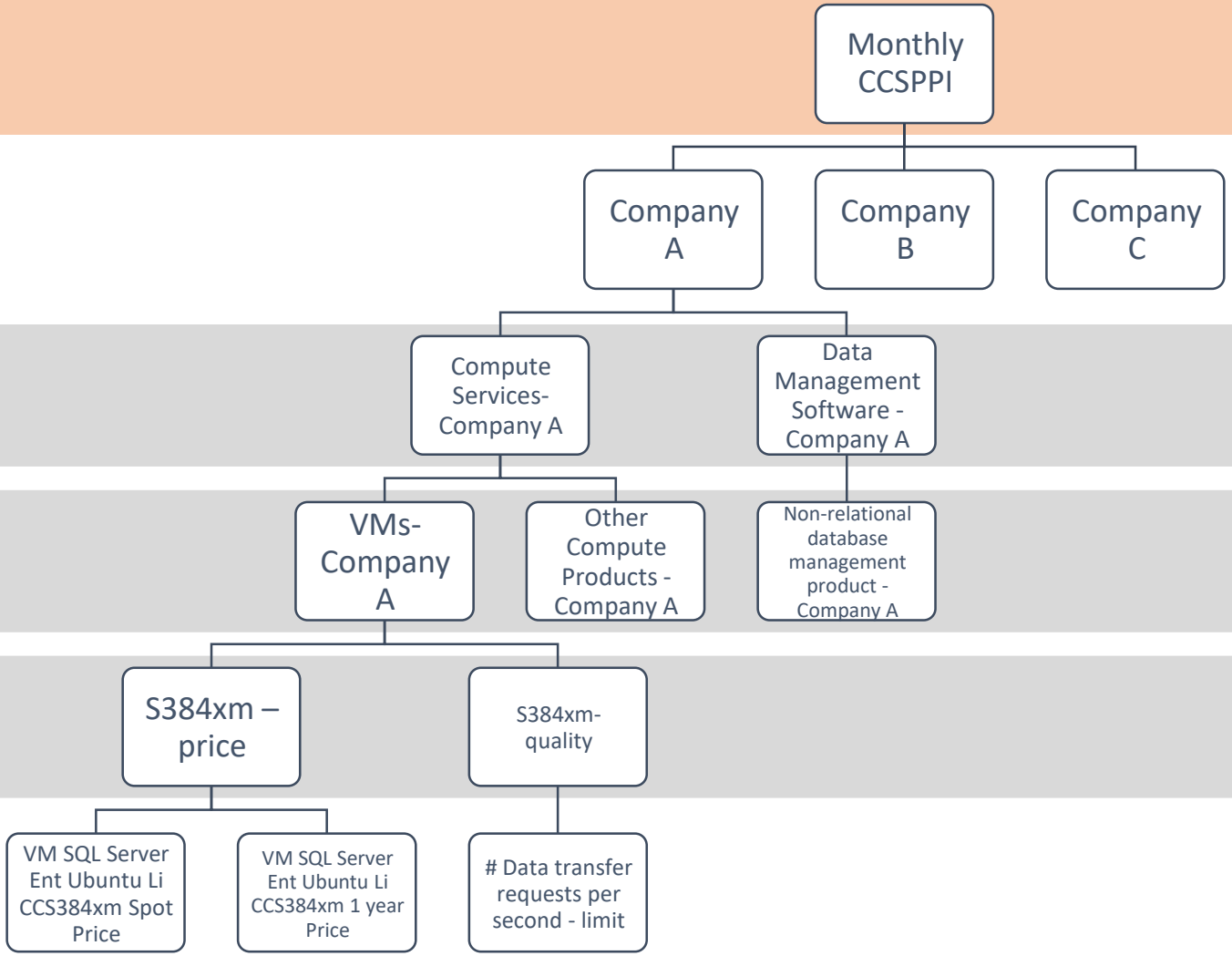
unweighted aggregation

Product-Category Relatives (*m1A, m2A...mk*)

Product-level indexes (Elementary Aggregates *j1A, j2A...jk*)

Sub-product(*1,2..j*)

SKU (ID)



Estimation

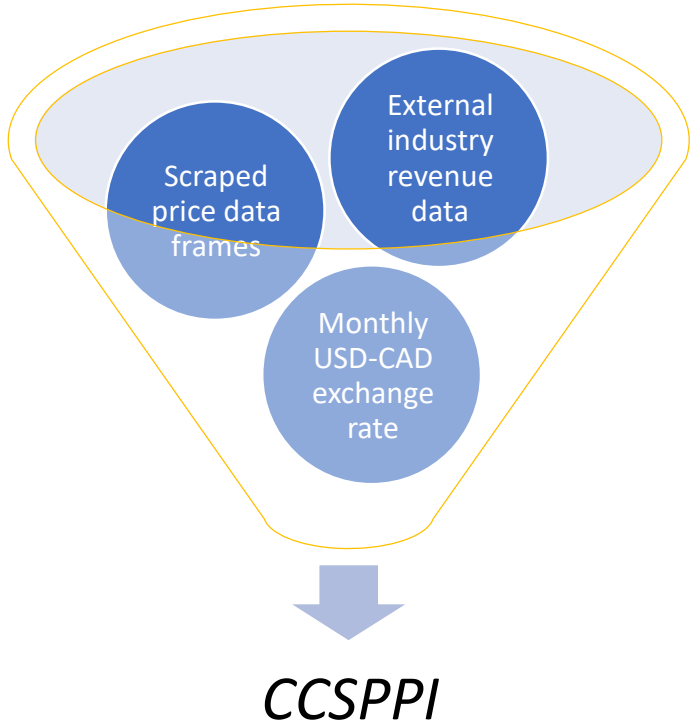
- Instead of taking the geometric average of price-type indexes & quality-type indexes, both are multiplied as a form of quality-adjustment, such that the price index *I* for product *j* and CSP *k is:*

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

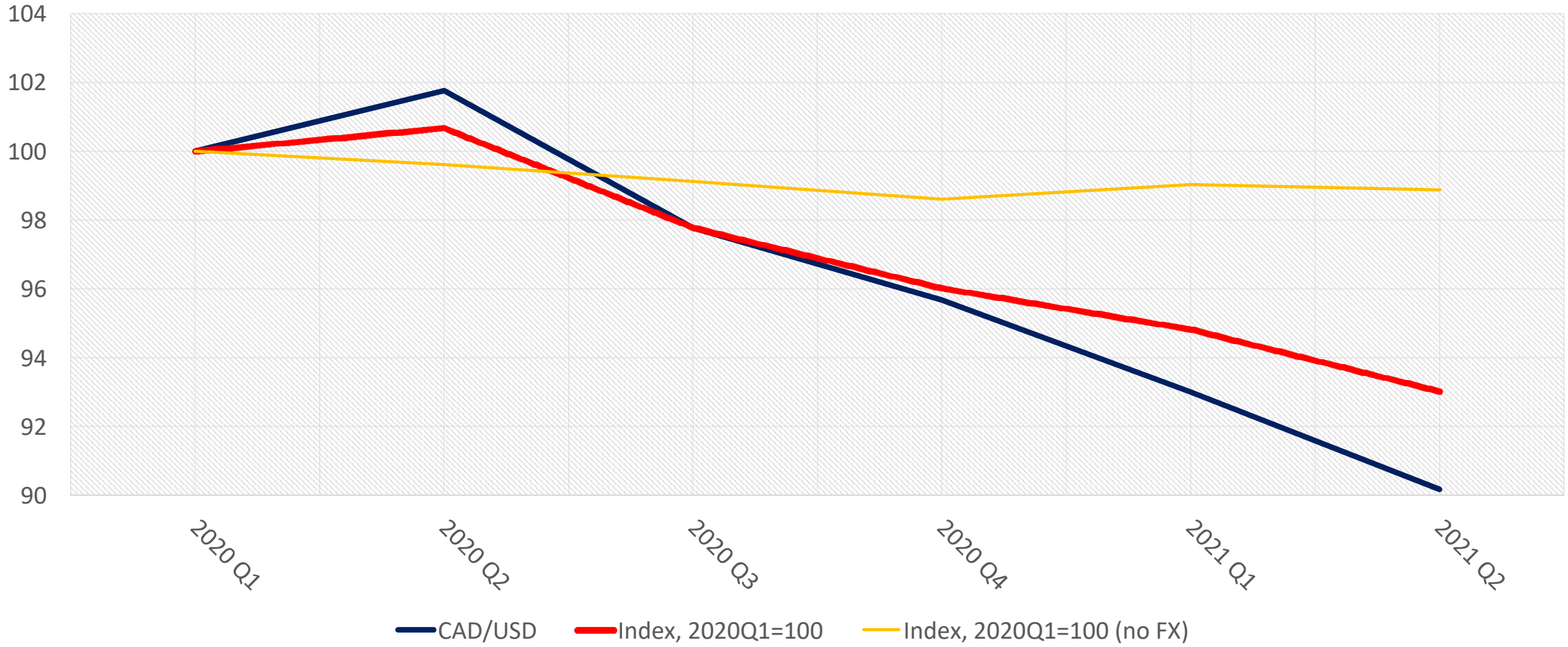
- Each product index is grouped according to product-category by geometric average, where things are generally comparable
- At product-category level, weights are applied by company, and aggregated via arithmetic average:

$$CSPPIMonthlyRelative = \sum_{k}\sum_{m} w_k^m * ProductCategoryRelative_k^m$$

  where w_k^m is the relative importance for product category *m* and CSP *k*.
- Currently, basket duration is two years, and index frequency is quarterly



Quarterly Index Results



Microdata analysis

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	**6**	**5**	**3**	**4**	**6**
2020 Q2	**7**	**5**	**3**	**5**	**7**
2020 Q3	**6**	**5**	**3**	**5**	**6**
2020 Q4	**6**	**5**	**3**	**4**	**6**
2021 Q1	**10**	**8**	**2**	**8**	**10**

Next Steps – expanding scope

- Where needed, ensure index structure is adaptable to industry changes
- Continue preliminary discussions with Canadian System of Macroeconomic Accounts to determine proper use in deflation of digital economy
- Continually integrate new CSP with diverse pricing models; fine-tune collection techniques where possible
- Evaluate overlap with existing price indexes for overlapping industries