The Use of Alternative Data for Price Indexes: The Canadian Experience

Heidi Ertl, Statistics Canada
Heidi.Ertl@canada.ca

Overview
While the majority of Statistics Canada’s published price indexes are still based on traditional survey data, massive efforts are being made to leverage administrative and alternative data sources to reduce response burden and program costs, while maintaining or improving data quality. However, this reliance on alternative data is not without challenges.

The Canadian experience
Of the 25 published price indexes, more than half use administrative/alternative data, either solely or in combination with survey data. This extensive use of alternative data has led to the development of innovative tools and new ways of managing, integrating, capturing, processing and estimating price data.

Why use alternative data?
Besides making use of existing data and saving on collection costs, using alternative data can:

- Reduce response burden
- Maximize number of price observations
- Minimize sample and survey bias

The challenges with alternative data
Alternative data as a replacement for survey data has many advantages, however there are also challenges and risks that may come into play:

- Prohibitive cost
- New ways to evaluate/certify data quality
- Supply interruptions, changes to source data
- Limitations on intended use
- Conceptual differences (e.g. list prices)
- IT transmission and storage
- Capture and processing of unstructured data

Alternative data: types

<table>
<thead>
<tr>
<th>Public data</th>
<th>Administrative data</th>
<th>Data from other STC programs</th>
<th>Data from third parties in the private sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet websites, subscriptions to catalogues or journals</td>
<td>Acquired under the Statistics Act or other formal agreements</td>
<td>Data from other surveys programs used as inputs into another</td>
<td>Data compiled by the private sector and acquired through contract or data sharing agreements</td>
</tr>
<tr>
<td>Unstructured price data for specific products/services</td>
<td>Structured data containing financial data, response tax data etc.</td>
<td>Structured data from other price indexes, and surveys</td>
<td>Commercially available structured data</td>
</tr>
</tbody>
</table>

Alternative data: tools and work flow

Price index program | Type of source data | Data capture | Processing | Estimation | Main challenges |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment Banking Services</td>
<td>Data from private sector Other STC programs</td>
<td>Custom loader to Repository</td>
<td>Generalized processing system</td>
<td>Weighted average of annual price changes by product, chained together to form an index series</td>
<td>Product heterogeneity (within period and over time) Non-repeating transactions Shortage of transactions or missing data in some periods</td>
</tr>
<tr>
<td>Industrial Product and Raw Materials</td>
<td>Monthly electronic survey Other STC programs</td>
<td>Survey Admin Data Interface to Repository Custom loader to Repository</td>
<td>Generalized processing system</td>
<td>Price relatives weighted by sales of goods manufactured. Indexes are produced by commodity and industry</td>
<td>Multiple data sources, price concepts</td>
</tr>
<tr>
<td>Non-residential and Apartment Building Construction</td>
<td>Quarterly electronic survey Other STC programs Public data</td>
<td>Survey Custom loader to Repository</td>
<td>Generalized processing system Secondary processing system to handle multiple dimensions (product, building type, city)</td>
<td>Quarterly price movements for weighted representative products applied to building models to produce indexes by building, city and trade groups</td>
<td>Survey collects price movements, while alternative data collected is unit price. Additional processing in secondary system required to accommodate multiple dimensions</td>
</tr>
<tr>
<td>Commercial Software and Computers and Peripherals</td>
<td>Data from private sector</td>
<td>Admin Data Interface to Repository</td>
<td>Generalized processing system</td>
<td>Fixed sample for desktop, laptop, tablet, servers and smartphones. Matched sample for computer peripherals (printers, monitors) and commercial software. Also known as ‘maximum overlap’ method</td>
<td>Completely dependent on external data source Changes to data can impact weighting, number of price observations/products and index methodology Fixed sample does not fully maximize data from the supplier and requires hedonic quality adjustment to link discontinued and replacement models</td>
</tr>
</tbody>
</table>