## LESSONS AND EXPERIENCES FROM AUSTRALIA, JAPAN, AND NEW ZEALAND

## INCORPORATING NON-TRADITIONAL DATA SOURCES INTO OFFICIAL STATISTICS

# THE CASE OF CONSUMER PRICE INDEXES





## **Table of contents**

1.	Intro	oduction	2
1	.1	Background	2
1	.2	Purpose of this report	3
1	.3	Traditional and non-traditional data sources for price measurement	3
1	.4	Why use non-traditional data for price measurement?	5
1	.5	Potential issues with adopting non-traditional data sources	7
2.	Cas	e Studies	11
2	2.1	Introduction to the case studies	11
2	2.2	Australian Bureau of Statistics (ABS)	12
2	2.3	Statistics Bureau of Japan	14
3	8.4	Statistics New Zealand (Stats NZ)	16
3.	Pric	rities in pursuing non-traditional data sources	18
4.	Sun	nmary lessons learnt from experiences so far	19
5.	Cor	nclusion	24
6.	6. References and resources		

### Acknowledgements

The author would like to thank and acknowledge the contributions from the Australian Bureau of Statistics, the Statistics Bureau of Japan, and Stats NZ. A special thank you to Leigh Merrington, and Andrew Tomadini (ABS), Naruhiko Yamagata (SBJ), Fiona Smillie, and Paul Pascoe (Stats NZ), as well as Gemma Van Halderen and Tanja Sejersen (UNESCAP), who all made valuable contributions.

### 1. Introduction

### 1.1 Background

Consumer Price Indexes (CPIs) are one of the highest profile official statistics produced by any National Statistical Office (NSO). CPIs measure the changing price of the goods and services that households buy. While internationally there some differences in the concepts and methods used to compile CPIs, they are generally used for similar purposes in most countries. These purposes include: being a tool for monetary policy, an economic indicator, and being used for the indexation of contracts and government transfer payments.

The possibilities offered by scanner data have captured the imagination of price statisticians and academics for decades. Price statisticians have dreamed of what they might be able to achieve if they had access to timely and comprehensive scanner data. However, until quite recently very few NSOs have successfully implemented scanner data into their CPIs.

Statistics Netherlands has been a leader in the use of scanner data for price measurement and was one of the first to implement scanner data into their CPI. In more recent years, the Statistics Bureau of Japan, Stats NZ, and the Australian Bureau of Statistics have implemented scanner and other forms of non-traditional data sources into their CPIs

Traditionally, the high cost of obtaining prices data meant that NSOs enjoyed a monopoly in the production of regular, nontrivial, price indexes. However, the launch of MIT's Billion Prices Project in 2008 proved that relatively comprehensive price indexes could be produced by other organisations<sup>1</sup>. The Billion Prices Project makes use of millions of prices automatically collected from retailer's websites through a process known as web-scraping. The Billion Prices Project grew to a point where MIT are collecting millions of prices every day from websites around the world and producing daily price indexes. The success of the project resulted in the initiative being expanded into a commercial venture called PriceStats. Furthermore, in 2016, the software company Adobe launched their Digital Price Index as part of their Digital Economy project<sup>2</sup>. Adobe's indexes initially covered a smaller range of products, but grew over time to a more comprehensive range including, but not limited to; consumer electronics, food, airfares, furniture, and non-prescription medicines. The Adobe Digital Price Index made use of transactions data obtained from a selection of Adobe's clients using their digital marketing tool.

<sup>&</sup>lt;sup>1</sup> Billion Prices Project

<sup>&</sup>lt;sup>2</sup> Adobe's Digital Economy Project

#### **1.2** Purpose of this report

The Asian and Pacific statistics community have a shared vision that by 2030 national statistical systems, including NSOs, are enabled and empowered to lead the development of innovative, trusted and timely statistical products and services, for the evolving statistical requirements of the 2030 Agenda for Sustainable Development. This report documents and draws on the experience of three NSOs in the Asia-Pacific region who have successfully implemented non-traditional data sources, specifically scanner data, into their CPIs. Lessons can be learnt from the experience of the Australian Bureau of Statistics, the Statistics Bureau of Japan, and Stats NZ that will help other NSOs in their efforts to incorporate similar approaches within their CPIs or other price indexes.

A great deal of work has been put into researching and developing methods that can be used to work with non-traditional data sources. This work has been well documented in many papers which have been published (e.g. on the Global Working Group on Big Data for Official Statistics website) or presented at international conferences (e.g. the UN Ottawa Group on Price Statistics), some of which have been referenced at the end of this report.

This report focuses more on the practical side of implementing non-traditional data sources into CPIs rather than on the technical side. The goal of this report is to help NSOs with decisions on whether and how non-traditional data should be implemented into their CPIs, and to help NSOs do this more efficiently. Additionally, this report might help manage the expectations of how far along with this work NSOs are and should help raise awareness of how much work is required to implement non-traditional data sources into CPIs.

#### **1.3** Traditional and non-traditional data sources for price measurement

Steering clear of the definition of 'big data', this report focuses on the use of non-traditional data sources in CPIs.

Traditional data for CPIs is usually manually collected directly from retailers via personal visits, phone calls, or manually from websites. The traditional data collection methods usually result in relatively small, but well targeted, samples of products and services. Because these samples are small and relatively expensive to collect, it is often hard for these samples to reflect the impact of promotional sales and new varieties of goods and services entering the market. Traditional price data collection can also be burdensome on retailers. This is because price collection staff will be regularly visiting stores capturing the prices of goods and services. Furthermore, in situations where the product or service being sampled is no longer available,

the price collector will often need to talk to staff in the retail store to understand why that product or service is no longer available and what a suitable replacement should be.

For this report, non-traditional data sources for price measurement include scanner data, online data, and administrative data.

Scanner data is one of the best-known forms of transaction data used in CPIs. Scanner data are electronic records of transactions, usually where a barcode has been scanned. This data generally contains a weekly, or monthly, summary of all transactions for a retail chain, or location, at the most detailed (individual) product level. The data usually contains total value of sales as well as the total quantity of sales for the reference period. Using this information, unit prices can be calculated, which account for the different quantities of products sold at different prices over the reference period. As a result, CPIs using scanner data have the opportunity to better reflect the true impact of promotional sales than they would using traditional, manually collected data.

Scanner data can be obtained either directly from retailers, or in some cases it can be purchased from market research companies. However, it is good to be aware that, in some cases, data from market research companies might not be as timely as data sourced directly from retailers. Having said that the delay in getting data from market research companies could be due to them being busy adding value to the data, like aggregating some of the data which makes the volume of data received more manageable.

*Online data* is data that is electronically collected from retailer's websites. For this report, online data includes price and product/service data that has been automatically collected from websites by software through a process known as web-scraping, or data collected from website's Application Programming Interfaces (APIs). Accessing a website's APIs allows NSOs to more efficiently access the price and product/service data that underlie the website. Web-scraping is the same process used by the MIT Billion Prices Project to collect the online data they are using in their price indexes. Accessing data from websites via APIs, when they are available and access is granted to them, can be seen as preferable to web-scraping because a web-scraping tool doesn't need to navigate the retailer's website hence the process for collection is more efficient and creates less traffic on the retailer's website.

Administrative data is another important non-traditional data source that can be used for CPI. Administrative data is data that has been compiled by organisations for administrative purposes, for example, price lists or fee schedules for prescription drugs. Administrative data has been included within many country's CPIs for some time. In recent years, some NSOs have looked to make more use of administrative data sources to limit burden on respondents, improve the accuracy of their CPIs, or reduce the costs of data collection. As NSOs have worked to develop new methods that are more suited to large datasets, this has expanded the potential use of administrative data sets, as these methods can be applied to the administrative datasets as well.

#### 1.4 Why use non-traditional data for price measurement?

User expectations are being influenced and changed in a number of different ways. Since the Global Financial Crisis, official statistics, especially economic and financial statistics, have faced greater scrutiny from the user community. Stakeholders are expecting more from NSOs at a time when these organisations are often being asked to operate with less funding. Additionally, because inflation has been so flat in many countries for an extended period, some users are starting to ask more detailed questions about how CPIs are constructed in an effort to understand why inflation has been so low.

These influences on user expectations have meant some users are looking for more granularity in official statistics, including CPIs. Non-traditional data sources can potentially help NSOs deliver more granular CPI series by significantly increasing sample sizes and the geographical coverage of indexes. However, when dealing with non-traditional data sources, more granular data might not always be available, especially when it comes to the geographical granularity. This is true for online data, when a retail chain might offer national pricing online, but allows individual stores to set their own prices. Also, when NSOs are negotiating for the supply of scanner data they might need to make some trade-offs with the level of granularity available within the data. For example, some retailers might only agree to supply scanner data that has been aggregated to a broad regional level rather than supplying data for specific locations. This might be done to maintain the confidentiality of individual stores or franchise partners.

User and stakeholder expectations are also being influenced by what has been achieved by initiatives like the MIT Billion Prices Project and Adobe's Digital Price Indexes. NSOs no longer enjoy the monopoly on producing the type of information they once did. Additionally, because these initiatives are already making use of non-traditional data sources, they are constructed using a very large number of data points which in many cases dwarf that of the official CPIs. This can lead users to question the quality of the CPI and ask why similar methods are not being used for the official CPI.

Non-traditional data sources allow NSOs to increase the coverage of their CPIs by including nearly all the products sold by sampled retailers rather than just sampling a small number of products. This will significantly increase the same size, capture new products as they enter the market, and in the case of scanner data, will allow weights to be used more dynamically or updated more frequently.

While online shopping has been around for many years, it is still growing at a fast rate. Most major retail chains have a growing online presence, and more is being invested in sophisticated methods where these retailers monitor each other's websites and adjust their prices accordingly. In an environment where prices are more dynamic and change more frequently, traditional manual data collection methods could fail to capture this and risk being unrepresentative. This means that sticking with the status quo could be both a quality and a reputational risk that NSOs need to be aware of.

CPIs tend to be some of the timeliest statistics produced by NSOs. Some have suggested that non-traditional data sources will help CPI's become even more timely, pointing to the speed and frequency with which the Billion Prices Project produce their indexes. However, comprehensive measures of inflation cannot be published before the least timely price component of the index is captured, checked, and processed. This means that CPIs cannot really be any timelier than the data for the least timely good or service. This means that most non-traditional data sources will not help make CPIs any timelier than they are now.

Additionally, official statistics must be of fit-for-purpose quality. Given the importance of the CPI the significant policy uses of them (e.g. large sums of transfer payments indexed by them) and the non-revisable nature of most CPIs, there should always be a strong focus on the accuracy of the index. Generally, important stakeholders like central banks prioritise improvements to the accuracy and granularity of the CPI ahead of improved timeliness.

# 预览已结束, 完整报告链接和二维码如下:

https://www.yunbaogao.cn/report/index/report?reportId=5\_744

