



An
Phríomh-Oifig
Staidrimh

Central
Statistics
Office

Assessing the Quality of Transaction Data for Use in the CPI

Anthony Dawson, Calvin O'Brien

Central Statistics Office, Ireland

Transaction/Scanner Data

- Transaction/Scanner data is generated by point-of-sales terminals in shops and provides information at a barcode level
- Transaction datasets contain product quantity and retail information for the products over a period.



Grocery Price Collection in CPI

Before COVID-19

- 100 pricers around Ireland
- Pricing was carried out using specially developed mobile app
- No alternative sources of data

During COVID-19

- Forced withdrawal of pricers from field
- New sources of data for CPI
- Webscraping – Automated and Manual
- Price files from Supermarkets
- 1 supermarket provided scanner data





An
Phríomh-Oifig
Staidrimh

Central
Statistics
Office

Post Covid and the introduction of Scanner Data



Scanner Data

- Only 30 pricers returned to stores.
- Pricing continued in office
- Statutory instrument updated to allow for CSO to request scanner data from retailers
- Scanner data now being collected from 3 retailers and one more coming in 2024. This is 4/5 major supermarkets in Republic of Ireland.
- New sources of scanner data had to be assessed for quality – no assessment structure in place



Voorburg Group and Assessment tool

- Voorburg Group is an UN group on Service Price Statistics
- Developed a questionnaire based on GSBPM
- Suggests measures that can be taken to assess and mitigate risks to quality
- This was determined to be the best method of assessing an alternative data source.
- <https://voorburggroup.org/Documents/2022%20Ottawa/Papers/1006.pdf>



Questionnaire Example

- 4. Collect
 - E. Is there sufficient consistency across records in the file to meet your needs?
 - Report on births and deaths of objects in the alternative dataset
 - Report on stability of variables and metadata
 - Total and Partial Response Rate, item response, unit response
 - Does the volatility of the variables in the data source warrant tracking

CSO Answer for Supermarket X:

Scanner Data has a constant flow of products entering and leaving the data. There is no record of new items or replacement in the data set, this must be tracked by the CSO instead. It is worth tracking the number of new items in a dataset from week to week and querying with Supermarket X if there are a large number of new products or items replaced.



Conclusions from GSBPM Analysis Tool

- Scanner data required two stages of assessment
 - Initial assessment
 - After a number of weeks of consistent data provision
- Initial evaluation determined that the proposed data was a possible data fit for use in CPI
- Analysis tool made for longitudinal analysis and final evaluation confirmed that data was suitable for use.
- Recommendation from questionnaire that analysis be continued on a every transmission to track new items, missing item and other metrics.
- This has allowed for the quick flagging and resolution of potential data errors.



Ongoing Assessment

- Tracking ongoing quality of scanner data is performed using an R shiny app
- App tracks general data quality metrics such as missing/negative entries and duplicates
- App also assesses quality of scanner data such as reliability on in store codes, item presence and significant changes in variables such as price or quantity
- This ongoing assessment flags issues before the data enters production



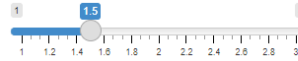
Scanner Data

Select Shop

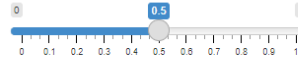
Select Year

Select Month

Significant Upper Price Change (% change)



Significant Lower Price Change (% change)



Select

Currently Selected: July 2023

Select an option

Total Entries: 5942

Total Unique GTINs: 2051

Weeks Present:

21 (1983)

22 (1958)

23 (2001)

Unique Groceries: 2052

Unique Groceries GTINs: 2052

Presence Length (in Months):

3: 1821

2: 117

1: 114

Duplicate GTINs: None

Entries with Values Missing: None

Entries with Negative Values: None

Previous Month Comparison:

Present Both Months: 1938

Added: 114

Removed: 103

Significant Price Changes: 3

Significant Price Changes Upper: 2

Significant Price Changes Lower: 1





Questions?

