



Motivation:

The classic attributes of Big Data—the 4 Vs: volume, variety, velocity and veracity—are insufficient to explore Big Data suitability for official statistics. A first distinctive factor is the scope for which official statistics needs or may use Big Data. Official statistics may use Big Data for decision-oriented analysis, whereas the scope of the private sector is an action-oriented analysis.

Study case:

An application of scanner data in Luxembourg is proposed and examined with the main aim of empirically explaining why it is currently impossible to efficiently explore Big Data suitability for official statistics using the classic attributes of Big Data.

Sources and Methods:

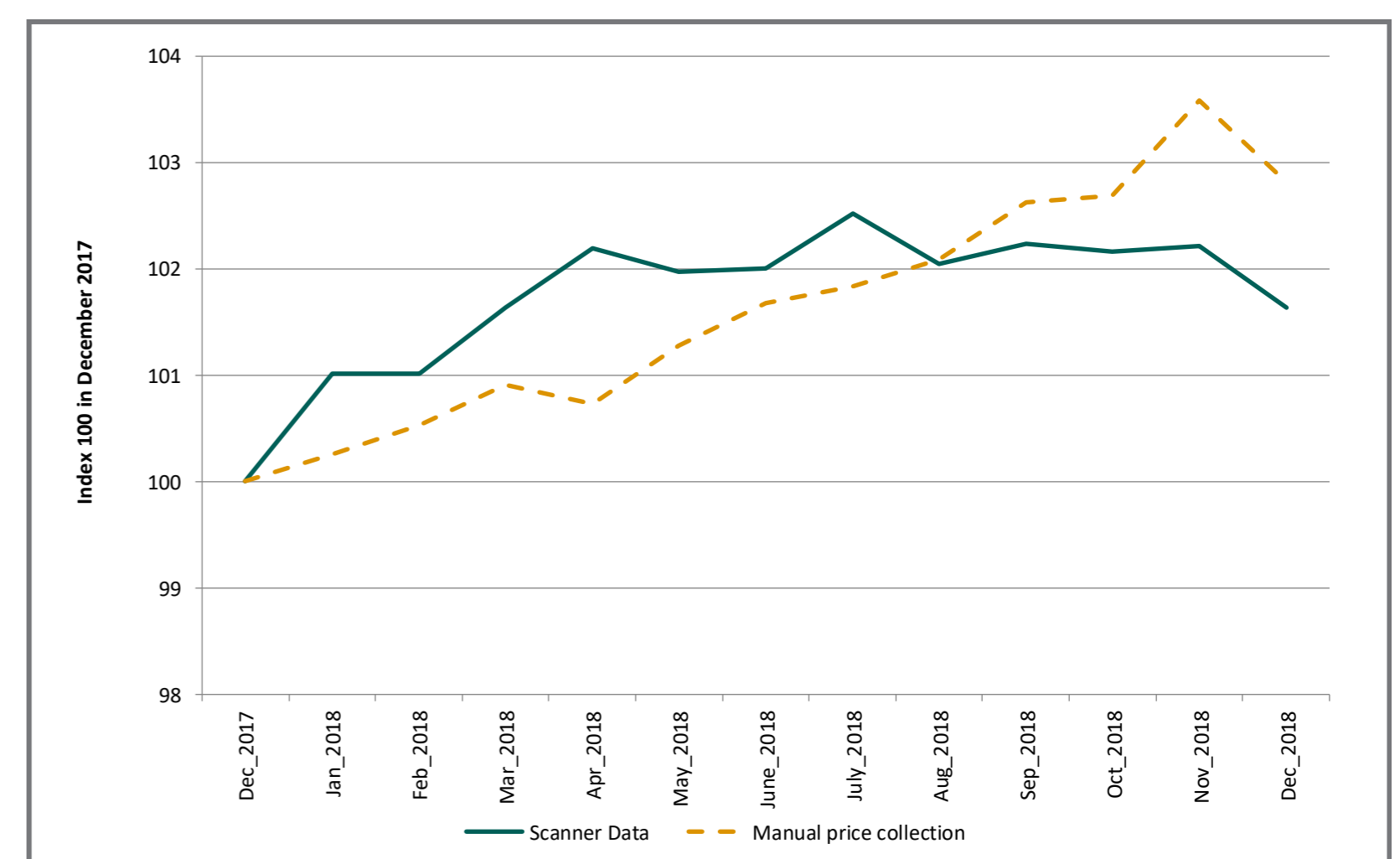
Using the scanner data produced by STATEC under the framework agreement implemented by GOPA Luxembourg, the following steps were implemented:

- Step 1: Automatic item classification procedure through cascading infrastructure which involves tasks such as search for common attribute, usage of dynamic mapping table, text mining;
- Step 2: Treatment of outliers and missing prices through price imputations;
- Step 3: Sampling according to the most sold items;
- Step 4: Index compilation.

Results:

- ⇒ Scanner data represents roughly 5 % of the 2018 HICP Luxembourg (food and non-alcoholic beverages)
- ⇒ The usage of Big Data for the public sector is, indeed, different from the usage by the private sector; i.e. decision-oriented analysis requires the continuity of Big Data (the availability of scanner data should be continuously ensured to produce these statistics)
- ⇒ The classic attributes of Big Data—the 4 Vs—are insufficient to explore Big Data suitability for official statistics; i.e.:
 - scanner data should be somewhat structured, always including certain variables, e.g. month of purchase, item label, item price, etc.
 - the speed by which scanner data is created and stored is fast enough, but it is certainly slower when scanner data needs to be analysed and visualised.

Figure 1: Comparison between the CPI manual price collection and scanner data for COICOP division 01

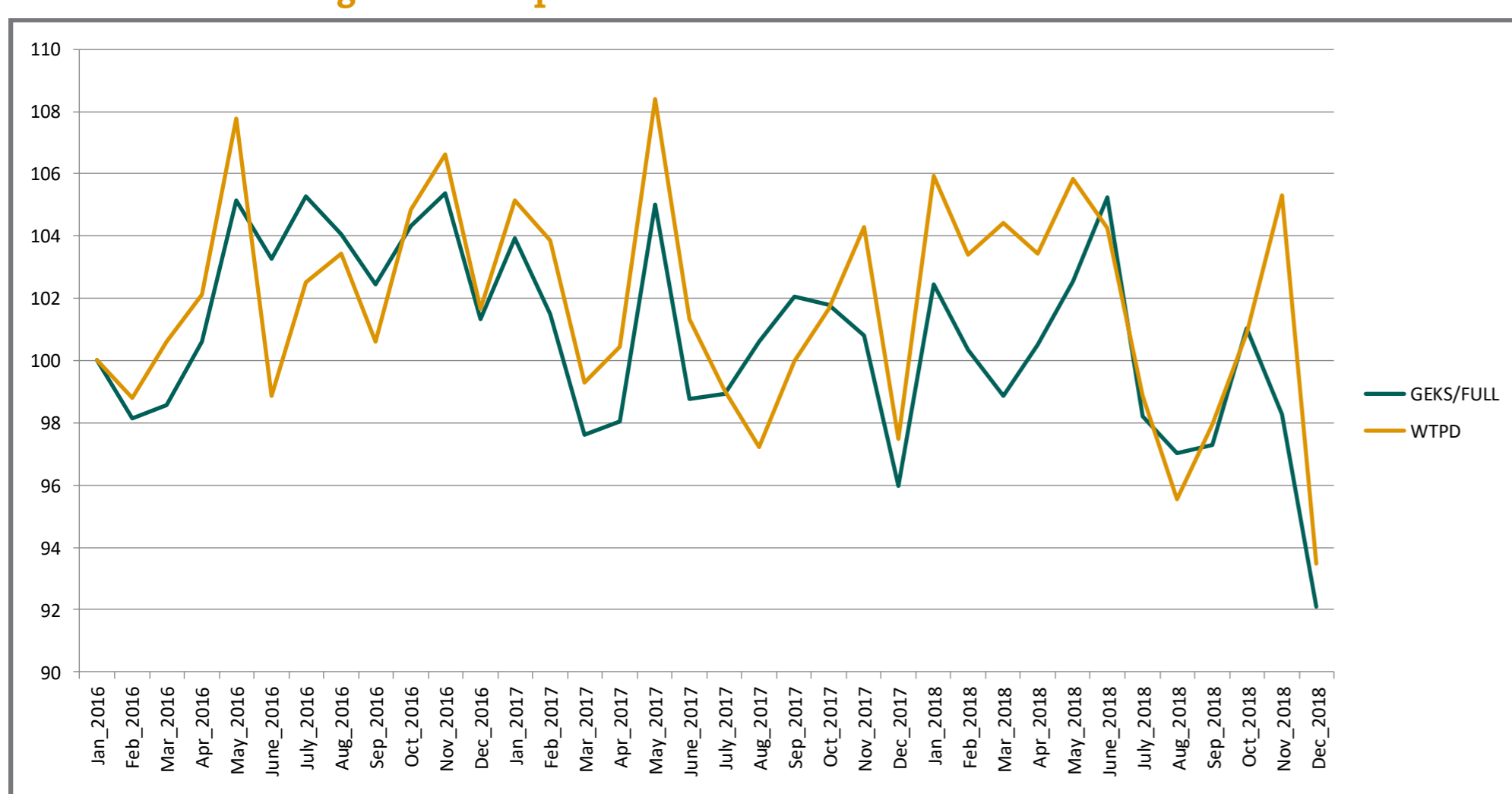


Source: STATEC

Conclusions and future plans:

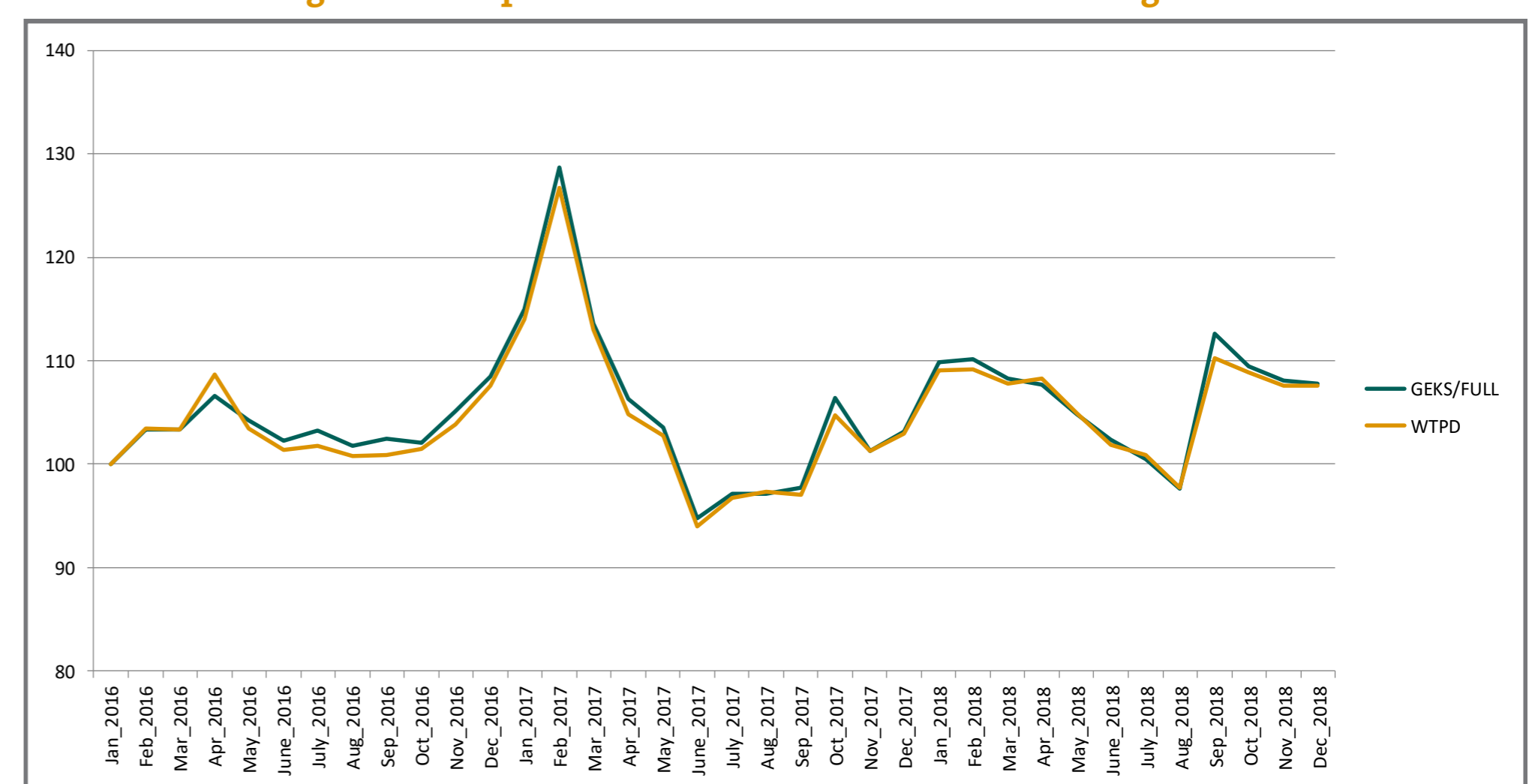
- ⇒ Scanner data - introduction of more retailers and more COICOP groups into STATEC production system as well as testing and development of multilateral methods

Figure 2: Comparison of GEKS and WTPD for fresh fruits



Source: STATEC

Figure 3: Comparison of GEKS and WTPD for fresh vegetables



Source: STATEC

- ⇒ Refining of the list of attributes in the context of Big Data for official statistics is needed.