

Using alternative data sources in consumer prices, UK

Tanya Flower, tanya.flower@ons.gov.uk
ONS

Outline

- Plans for alternative data sources
- Pipeline for processing data
- Experimental results
- Further research

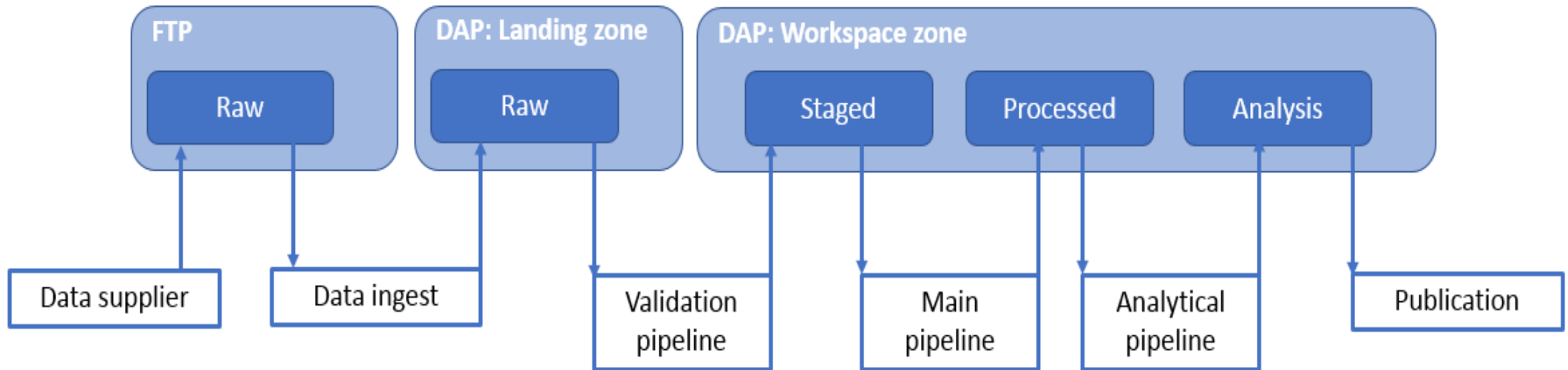
Plans for alternative data sources

- Incorporate into consumer price statistics by Jan 2023
- Develop new pipeline to process prices data including alternative data sources
- Focus on particular product groups

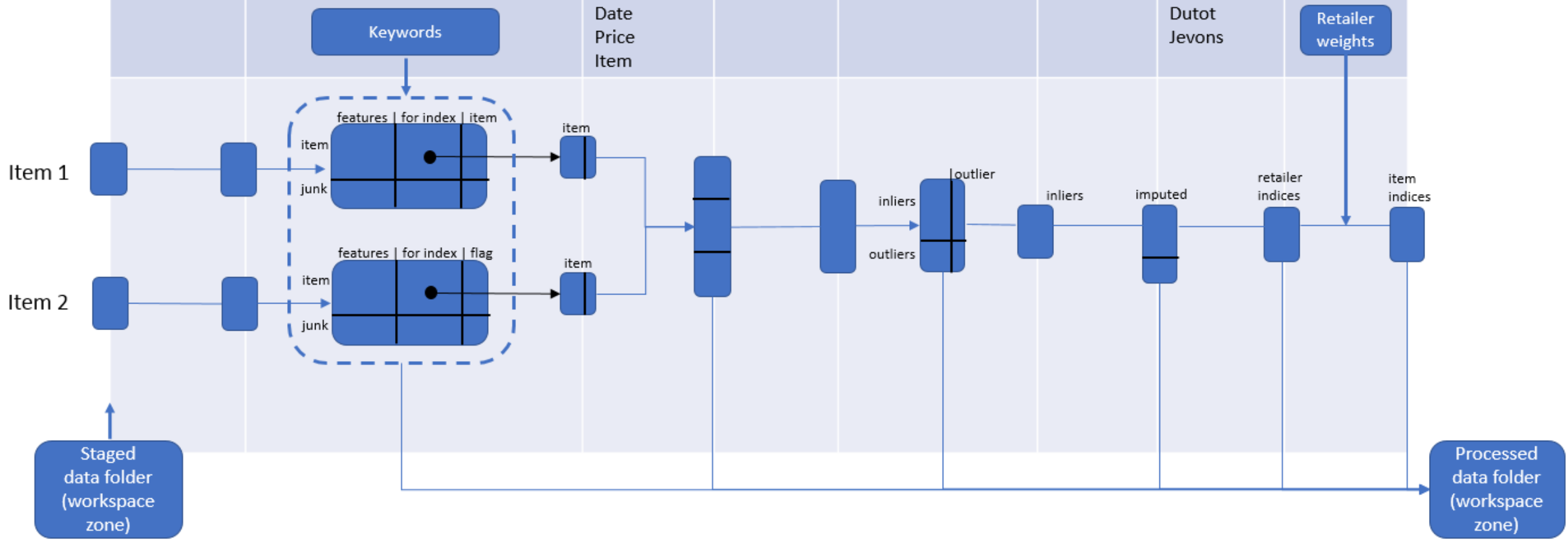
Plans for alternative data sources

- Regular updates and experimental indices scheduled
- Consultation with advisory panel and interest groups
- Intended formal consultation in 2022
- Impact assessment

Pipeline and methods



Pre processing	Classification (optional unless multi-item dataset)	Append items	Averaging	Outlier detection (optional)	Imputation (optional)	Retailer indices	Index aggregation
Each item has its own module and configuration parameters	Classify products based on decision rules applied to product name Create a 'for index' flag column	Columns of interest: Product ID Retailer Date Price Item	Monthly: Arithmetic Geometric	Tukey User defined fences	Fill forward	GEKS-J <i>Fixed-base:</i> Dutot Jevons <i>Chained:</i> Dutot Jevons	To item level



Main pipeline

- Pre-processing
- Classification
- Append items
- Averaging
- Outlier detection
- Imputation
- Retailer indices
- Index aggregation



Create additional variables

Main pipeline



ID	Product name	Retailer	Scrape date	Date	Price
L1	Laptop 1	A	01/04/2019	Apr-19	350
L3	Laptop bag	A	01/04/2019	Apr-19	30
L1	Laptop 1	A	08/04/2019	Apr-19	400
L4	Laptop 3	B	08/04/2019	Apr-19	30000000
L1	Laptop 1	A	01/03/2019	Mar-19	350
L2	Laptop 2	B	01/03/2019	Mar-19	600
L1	Laptop 1	A	08/03/2019	Mar-19	350
L2	Laptop 2	B	08/04/2019	Mar-19	600

Main pipeline

- Pre-processing
- Classification
- Append items
- Averaging
- Outlier detection
- Imputation
- Retailer indices
- Index aggregation



ID	Product name	Retailer	Scrape date	Date	Price	Item
L1	Laptop 1	A	01/04/2019	Apr-19	350	Laptop
L3	Laptop bag	A	01/04/2019	Apr-19	30	Junk
L1	Laptop 1	A	08/04/2019	Apr-19	400	Laptop
L4	Laptop 3	B	08/04/2019	Apr-19	30000000	Laptop
L1	Laptop 1	A	01/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	01/03/2019	Mar-19	600	Laptop
L1	Laptop 1	A	08/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	08/04/2019	Mar-19	600	Laptop

Main pipeline



ID	Product name	Retailer	Scrape date	Date	Price	Item
L1	Laptop 1	A	01/04/2019	Apr-19	350	Laptop
L1	Laptop 1	A	08/04/2019	Apr-19	400	Laptop
L4	Laptop 3	B	08/04/2019	Apr-19	30000000	Laptop
L1	Laptop 1	A	01/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	01/03/2019	Mar-19	600	Laptop
L1	Laptop 1	A	08/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	08/04/2019	Mar-19	600	Laptop

Main pipeline

- Pre-processing
- Classification
- Append items
- Averaging
- Outlier detection
- Imputation
- Retailer indices
- Index aggregation



ID	Product name	Retailer	Scrape date	Date	Price	Item
L1	Laptop 1	A	01/04/2019	Apr-19	350	Laptop
L1	Laptop 1	A	08/04/2019	Apr-19	400	Laptop
L4	Laptop 3	B	08/04/2019	Apr-19	30000000	Laptop
L1	Laptop 1	A	01/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	01/03/2019	Mar-19	600	Laptop
L1	Laptop 1	A	08/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	08/04/2019	Mar-19	600	Laptop

Other items...

Main pipeline



ID	Product name	Retailer	Scrape date	Date	Price	Item
L1	Laptop 1	A	01/04/2019	Apr-19	350	Laptop
L1	Laptop 1	A	08/04/2019	Apr-19	400	Laptop
L4	Laptop 3	B	08/04/2019	Apr-19	30000000	Laptop
L1	Laptop 1	A	01/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	01/03/2019	Mar-19	600	Laptop
L1	Laptop 1	A	08/03/2019	Mar-19	350	Laptop
L2	Laptop 2	B	08/04/2019	Mar-19	600	Laptop

Main pipeline



ID	Product name	Retailer	Date	Price	Item
L1	Laptop 1	A	Apr-19	374.17	Laptop
L4	Laptop 3	B	Apr-19	30000000	Laptop
L1	Laptop 1	A	Mar-19	350	Laptop
L2	Laptop 2	B	Mar-19	600	Laptop

Main pipeline



ID	Product name	Retailer	Date	Price	Item
L1	Laptop 1	A	Apr-19	374.17	Laptop
L4	Laptop 3	B	Apr-19	30000000	Laptop
L1	Laptop 1	A	Mar-19	350	Laptop
L2	Laptop 2	B	Mar-19	600	Laptop

Main pipeline



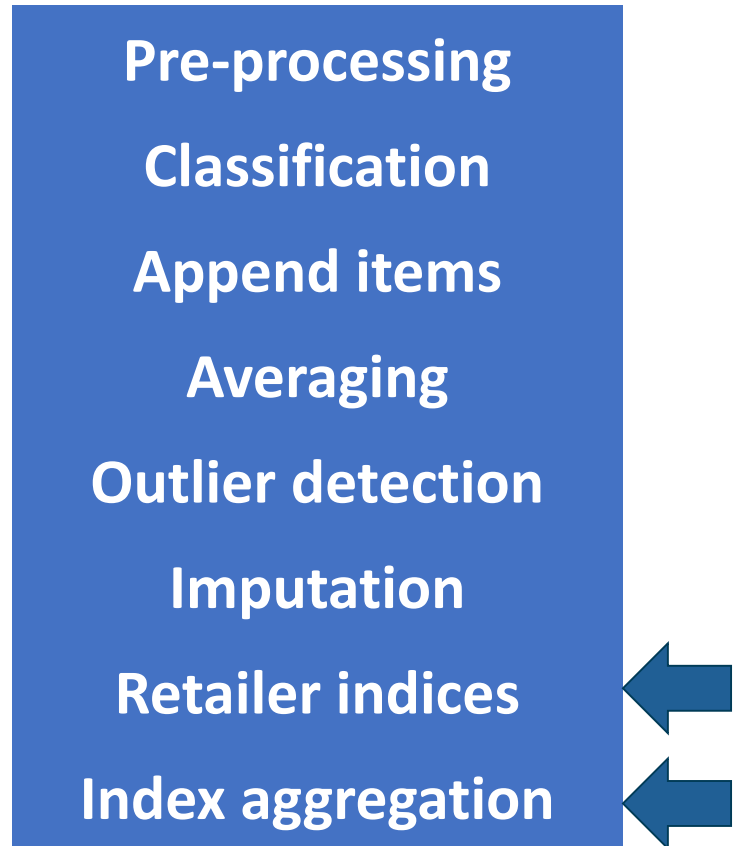
ID	Product name	Retailer	Date	Price	Item
L1	Laptop 1	A	Apr-19	374.17	Laptop
L1	Laptop 1	A	Mar-19	350	Laptop
L2	Laptop 2	B	Mar-19	600	Laptop

Main pipeline



ID	Product name	Retailer	Date	Price	Item
L1	Laptop 1	A	Apr-19	374.17	Laptop
L2	Laptop 2	B	Apr-19	600	Laptop
L1	Laptop 1	A	Mar-19	350	Laptop
L2	Laptop 2	B	Mar-19	600	Laptop

Main pipeline

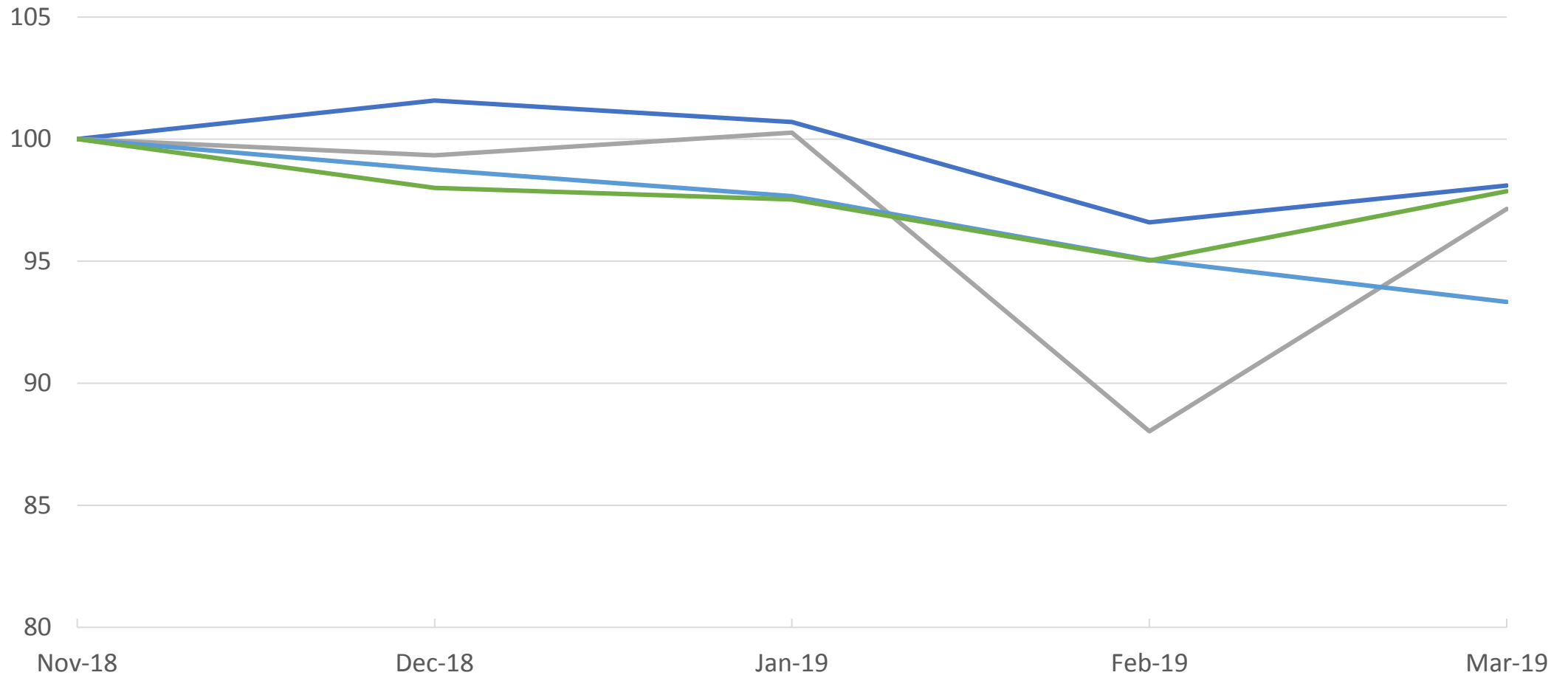


Elementary indices created at retailer-level

Expenditure weights used to aggregate to item-level

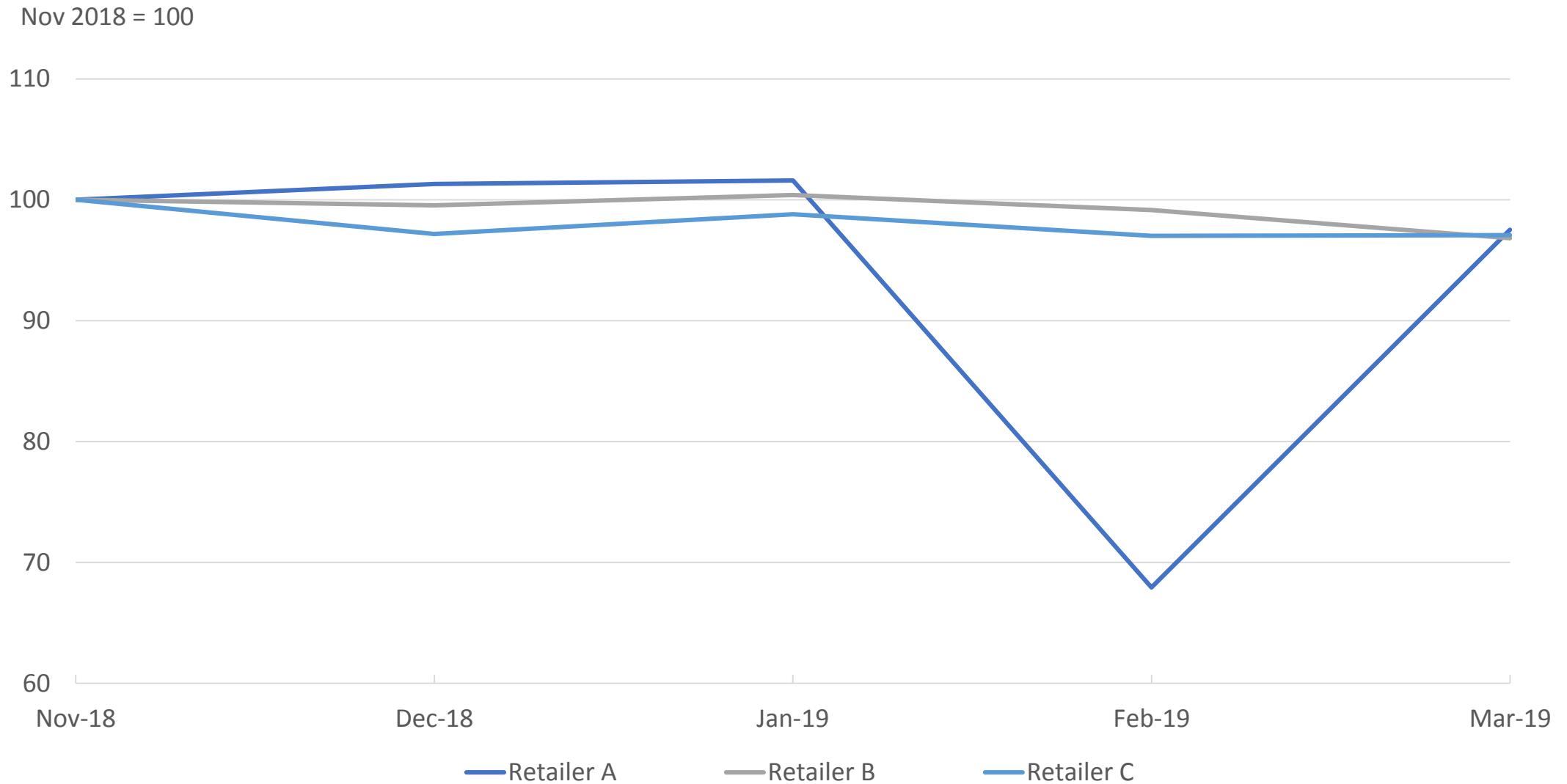
Fixed base Jevons indices for desktops, laptops, tablets and smartphones

Nov 2018 = 100



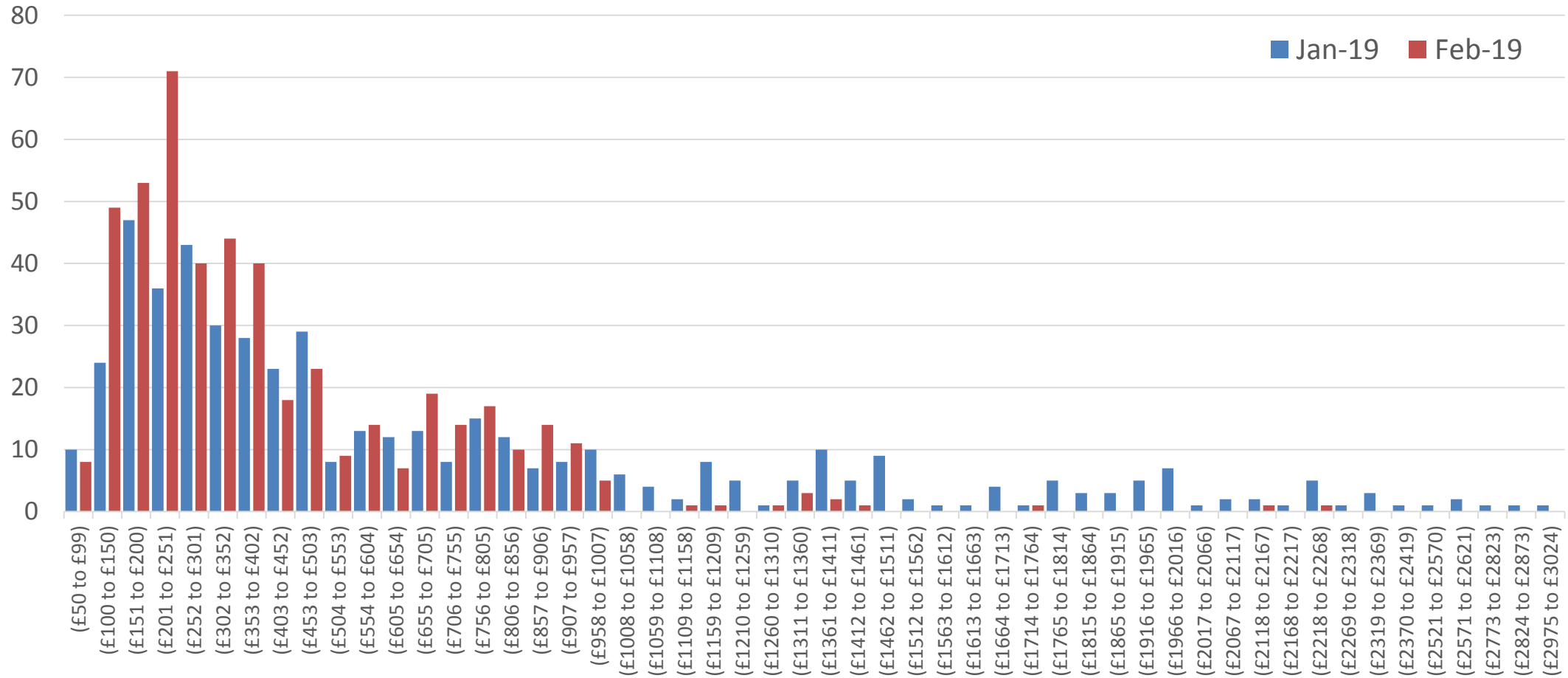
— desktop — laptop — smartphone — tablet

Fixed base Jevons indices for laptops by retailer



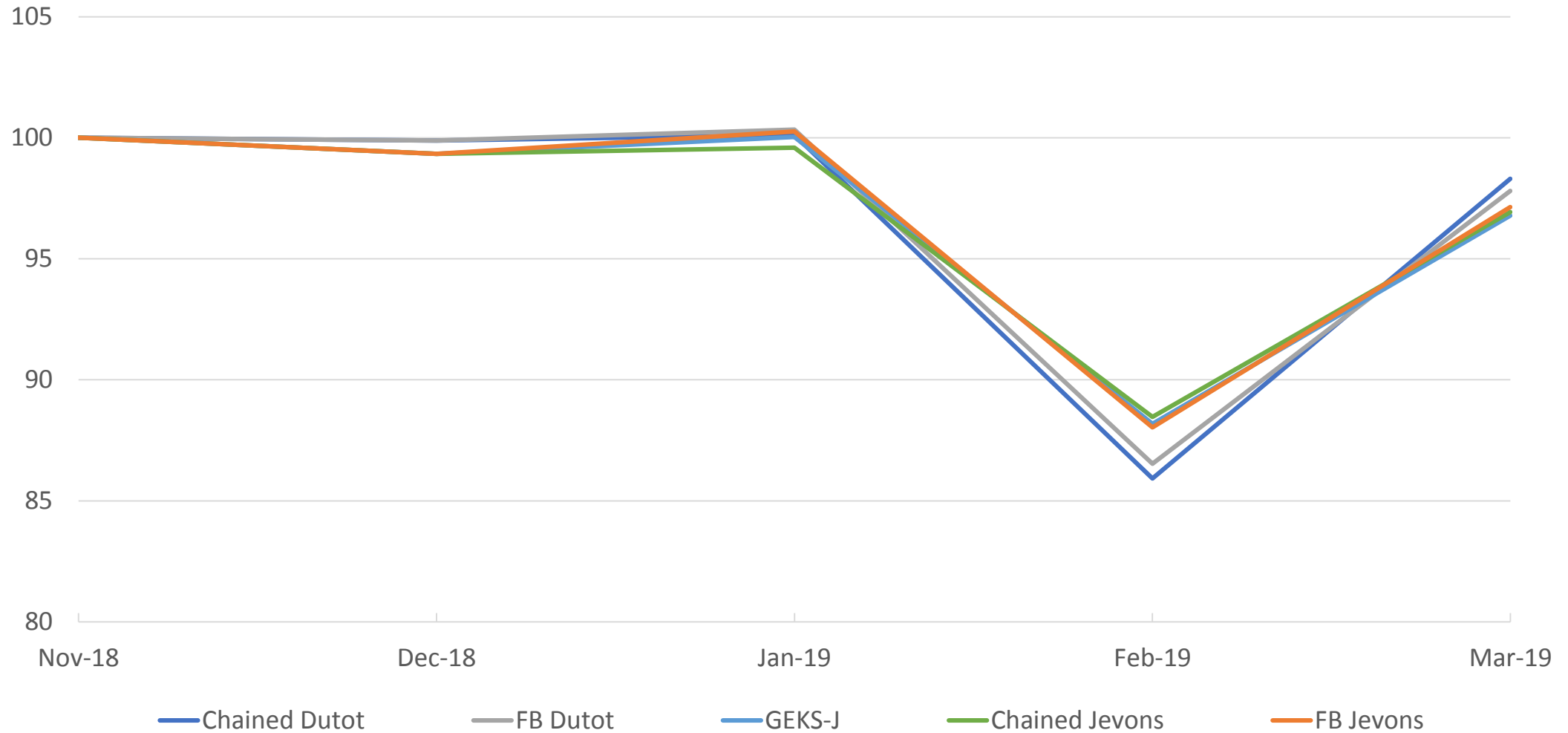
Price distribution for laptops, Retailer A, January and February 2019

Count of unique products



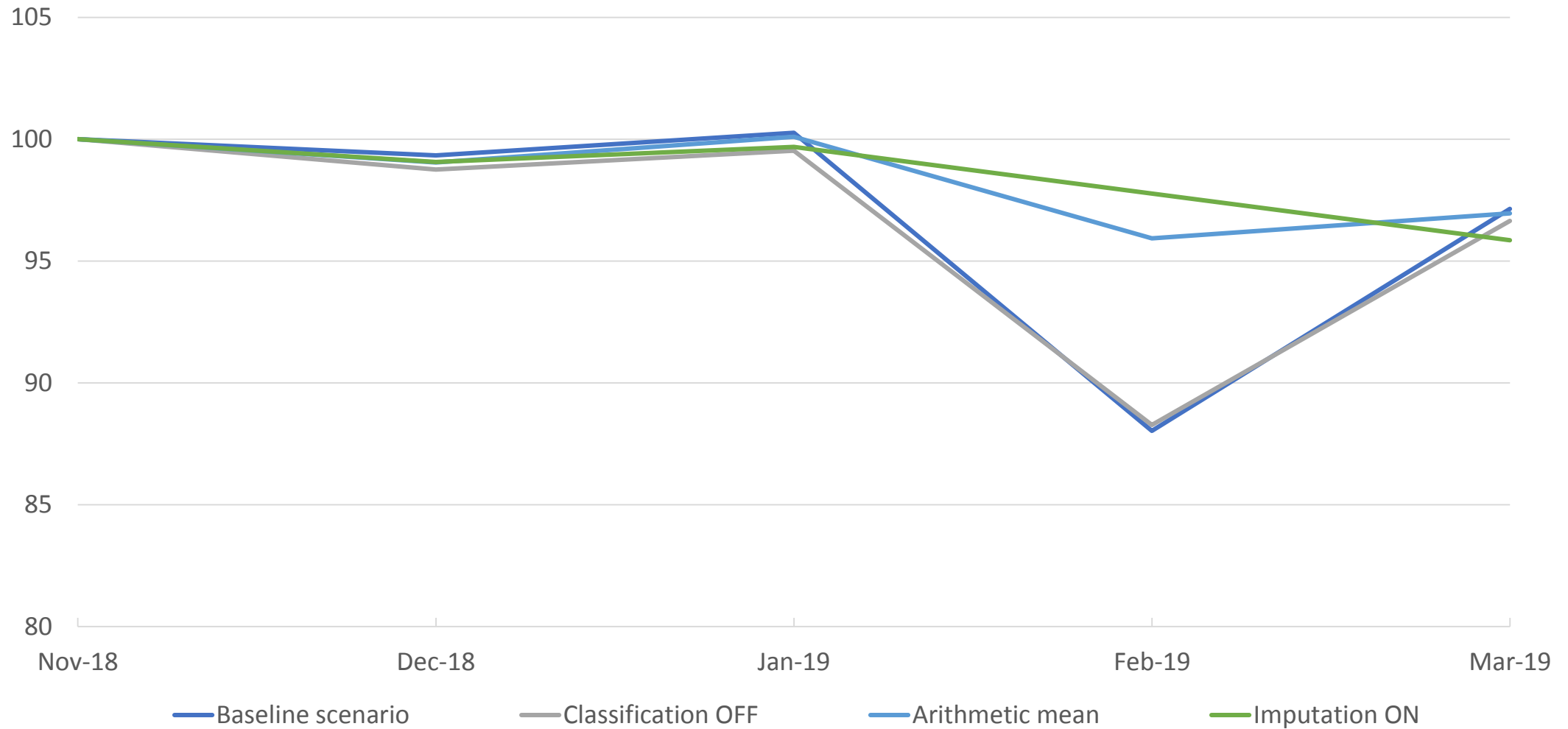
Price indices for laptops, different index methods

Nov 2018 = 100



Fixed base Jevons indices for laptops, alternative scenarios

Nov 2018 = 100



Index Methods

- Framework for assessing quality of consumer price indices produced using alternative data sources
- Product level expenditure weights for web scraped data
- Product definition

Further research

- Classification
- Further development of pipeline
- Aggregation of data sources
- Impact of returns and discounts

Quick wins...

- Can we replace existing data sources whilst using existing systems and methods?
- Can we use data elsewhere in the basket?

Questions?

...thank you for listening!