

Missing in action

testing alternative imputation
methods in price statistics

Patrick Kelly and Mathatsi Mogalanyane
Statistics South Africa

8 May 2019

Aim





Why do
prices go
missing?

And why do
we care?

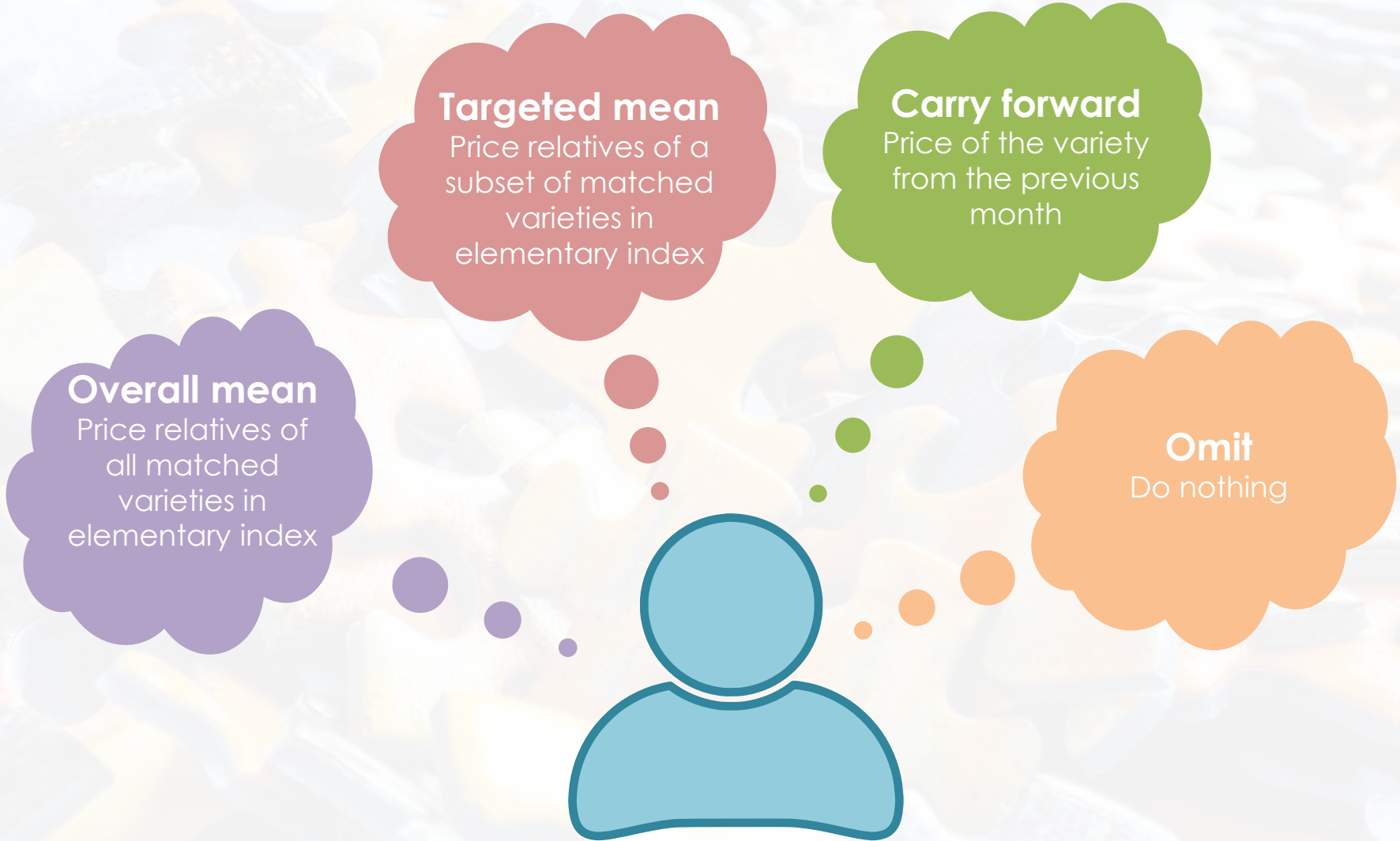
Why do prices go missing?



Aim of imputing

- Missing price creates a gap in sampled items – could create bias
- Objectives of imputing
 - “First do no harm” – Minimise bias
 - Complete matrix of prices for immediate use of next price when available

Imputation methods



What about regression?

- Time product dummy commonly used for quality adjustment when substituting for permanently missing prices
- Uses longer time series than traditional imputation techniques
- Test it as a method for imputing temporarily missing prices (will it help with seasonal products?)

Literature

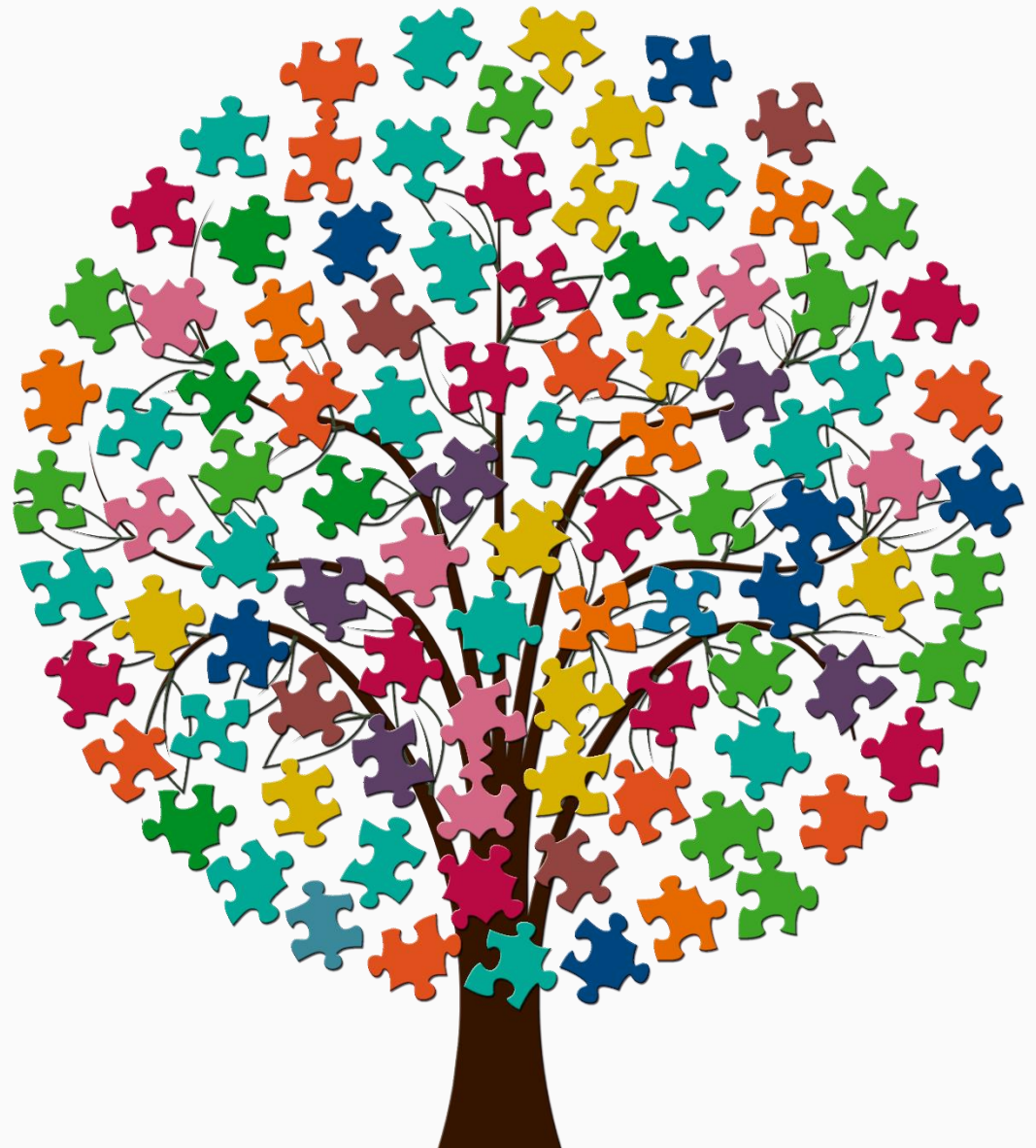
What have
others said?



Literature

- Very little quantitative analytical work
- Exception is Swiss study – test performance of overall and class mean for quality adjustment when substituting permanently missing varieties of clothing
- Class mean performed better than overall mean but needs bigger sample

Methodology



Methodology

- 25 month dataset for 7 products using South African CPI
- Imputed all missing prices to get complete matrix of data
- Randomly deleted 10% of prices in each product group

Description of data

COICOP group	Product	Pricing behaviour	Number of observations
Bread and cereals	Bread	Stable	758
Fruit	Peaches	Strong seasonal	10
Vegetables	Broccoli	Weak seasonal	57
Milk, cheese and eggs	Eggs	Stable	414
Clothing	Men's shorts	Strong seasonal	197
Clothing	Men's jeans	Stable	180
Furniture	Bedroom suites	Sticky	135

Methodology

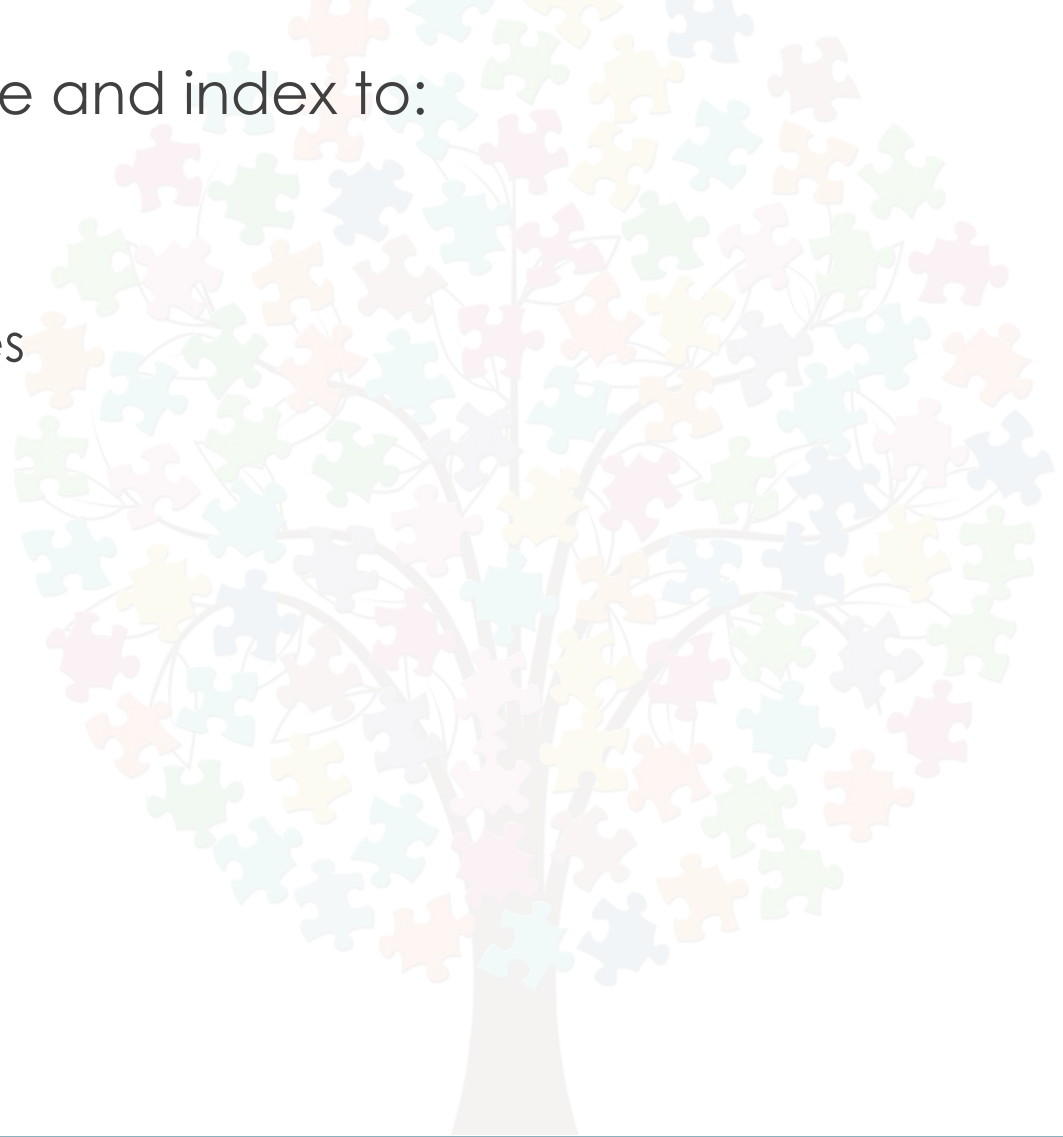
Imputed missing prices using 4 methods

- Overall mean - all available price relatives for product
- Targeted mean - price relatives of a subset of varieties for that product (specific geographic area)
- Carry forward - use price in $t-1$ as the price in t .
- TPD - all data from the product group for the current and previous 12 months.

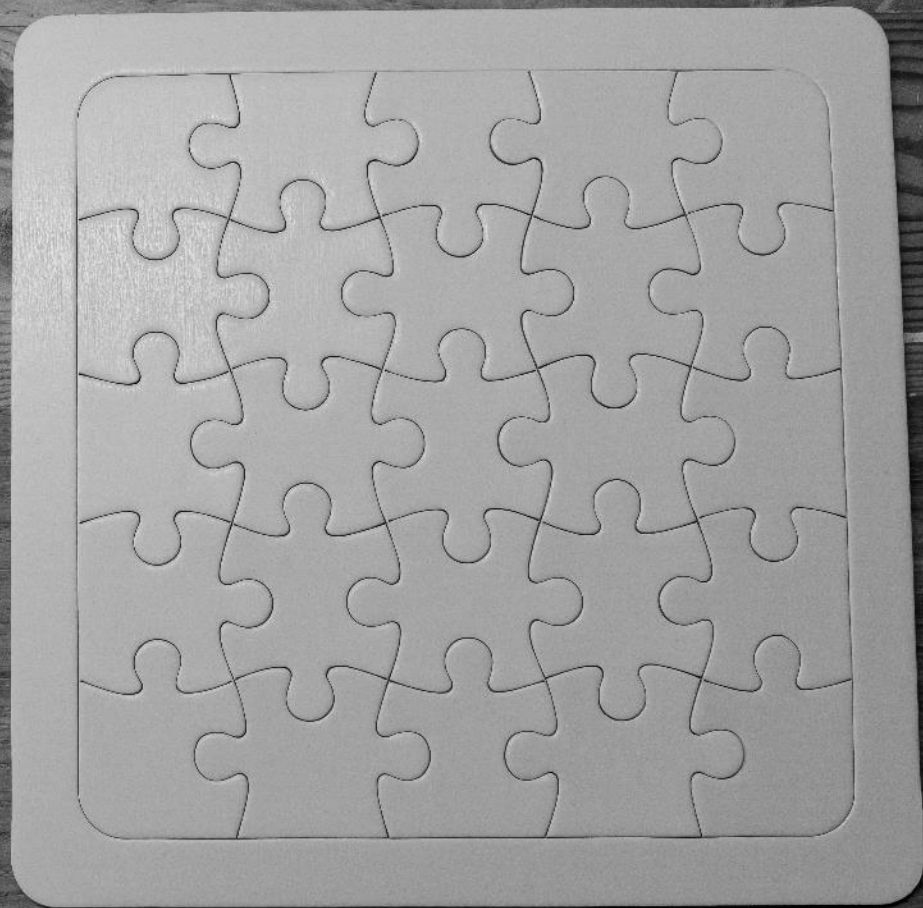
Methodology

Compare imputed price and index to:

- Original data
- Data with missing prices

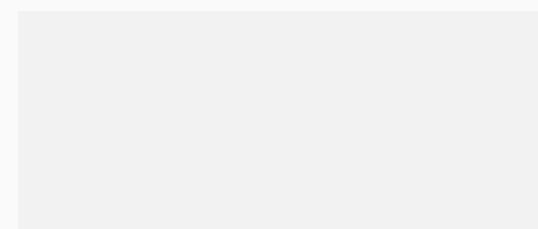
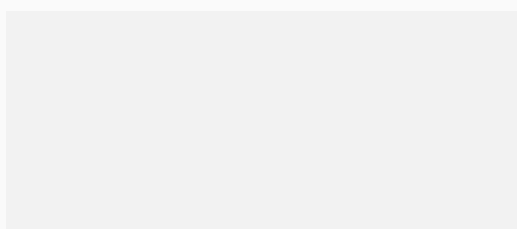
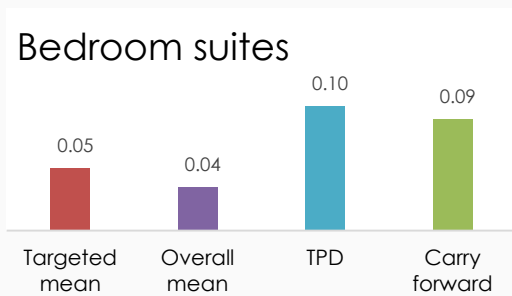
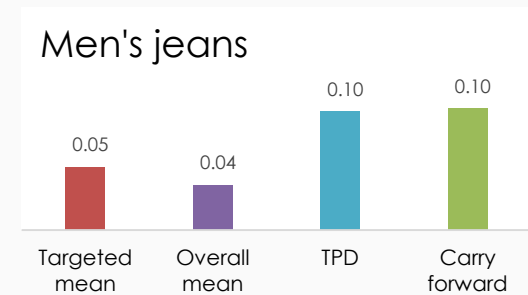
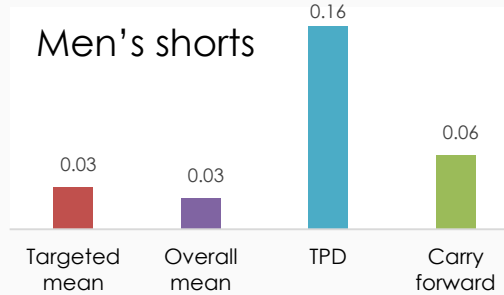
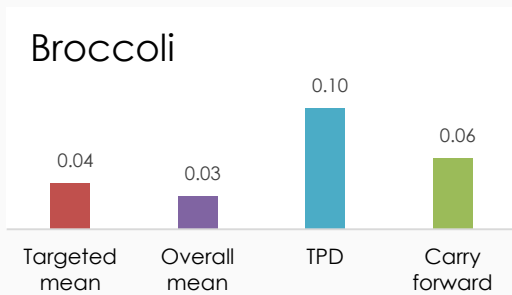
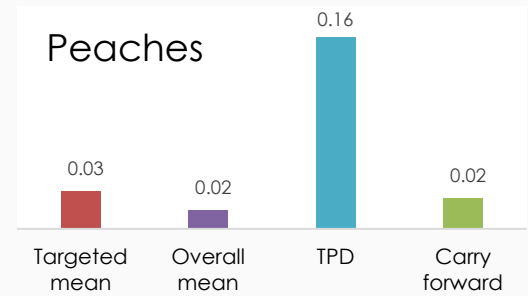
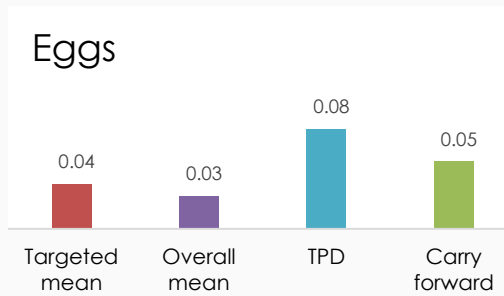
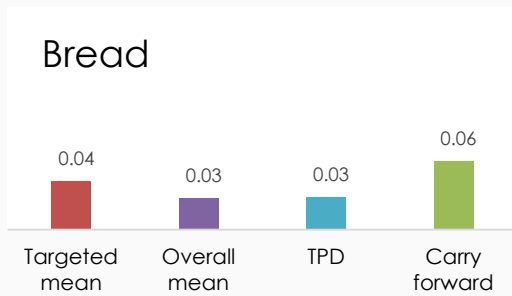


Results



Measure deviation of price and index
from the original data set

Price level: Difference between actual and imputed price

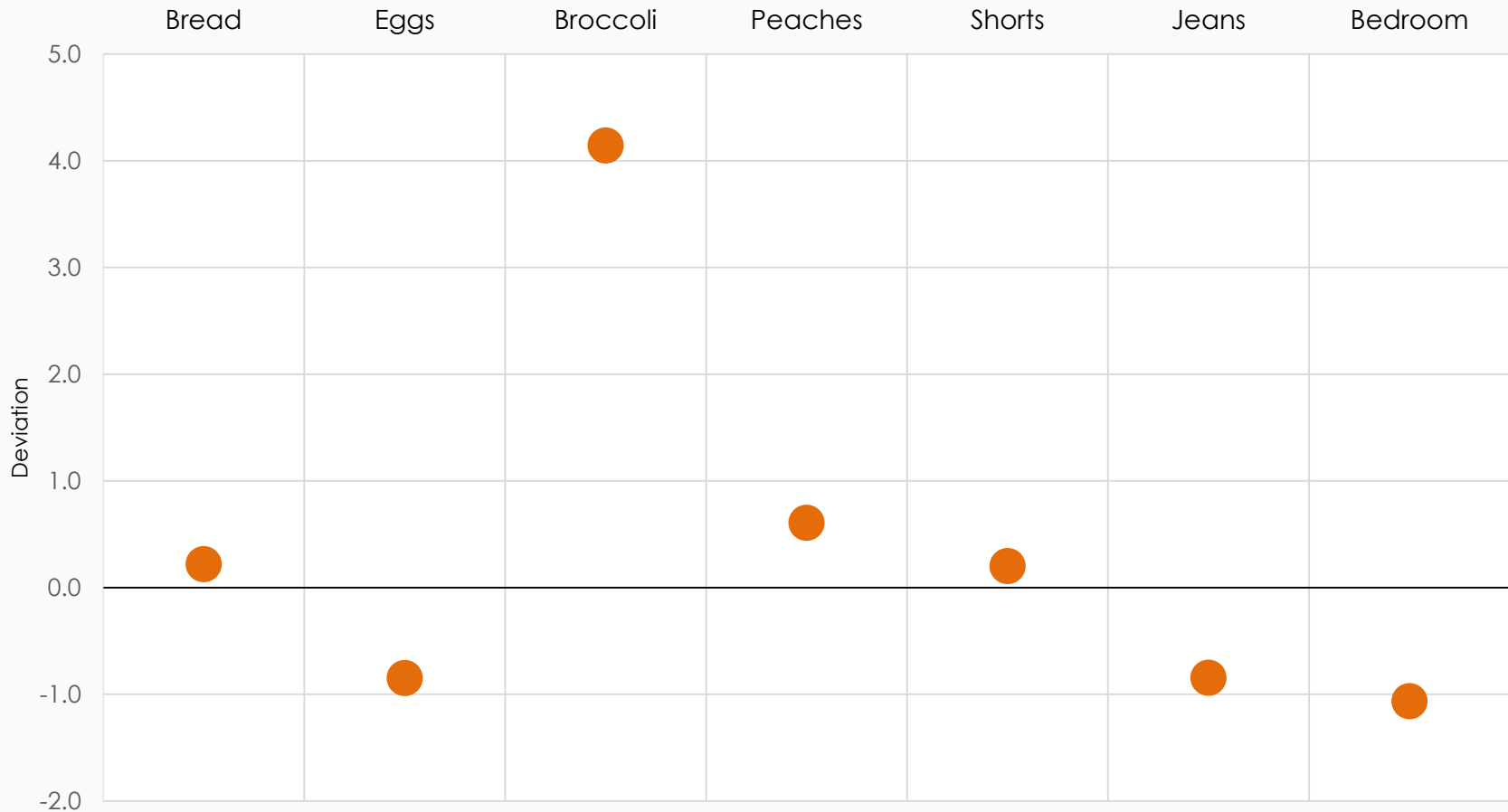


Index level: Deviation of index by imputation type



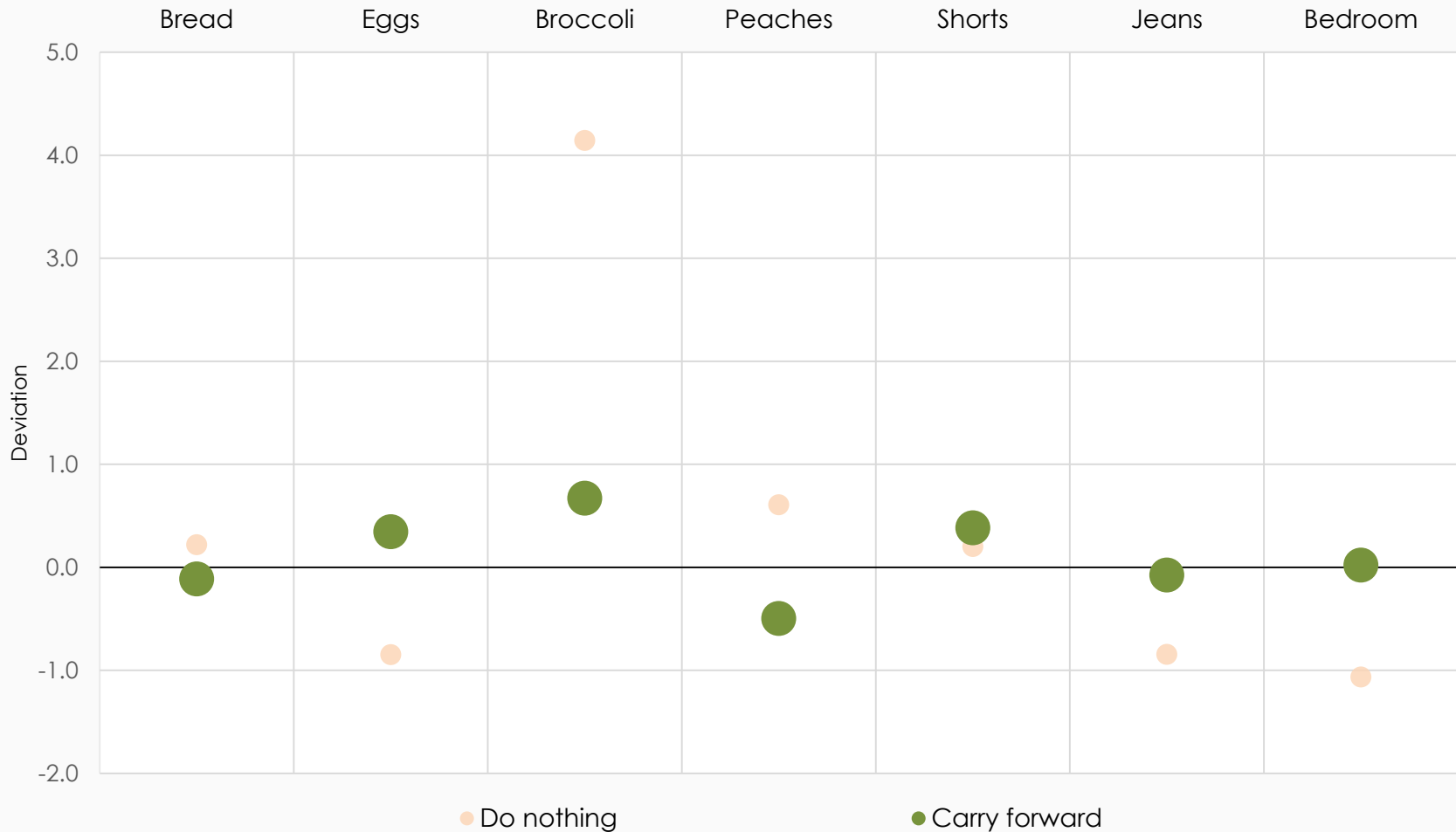
Do nothing

Index level: Deviation of index by imputation type – Do nothing



● Do nothing

Index level: Deviation of index by imputation type – Carry forward



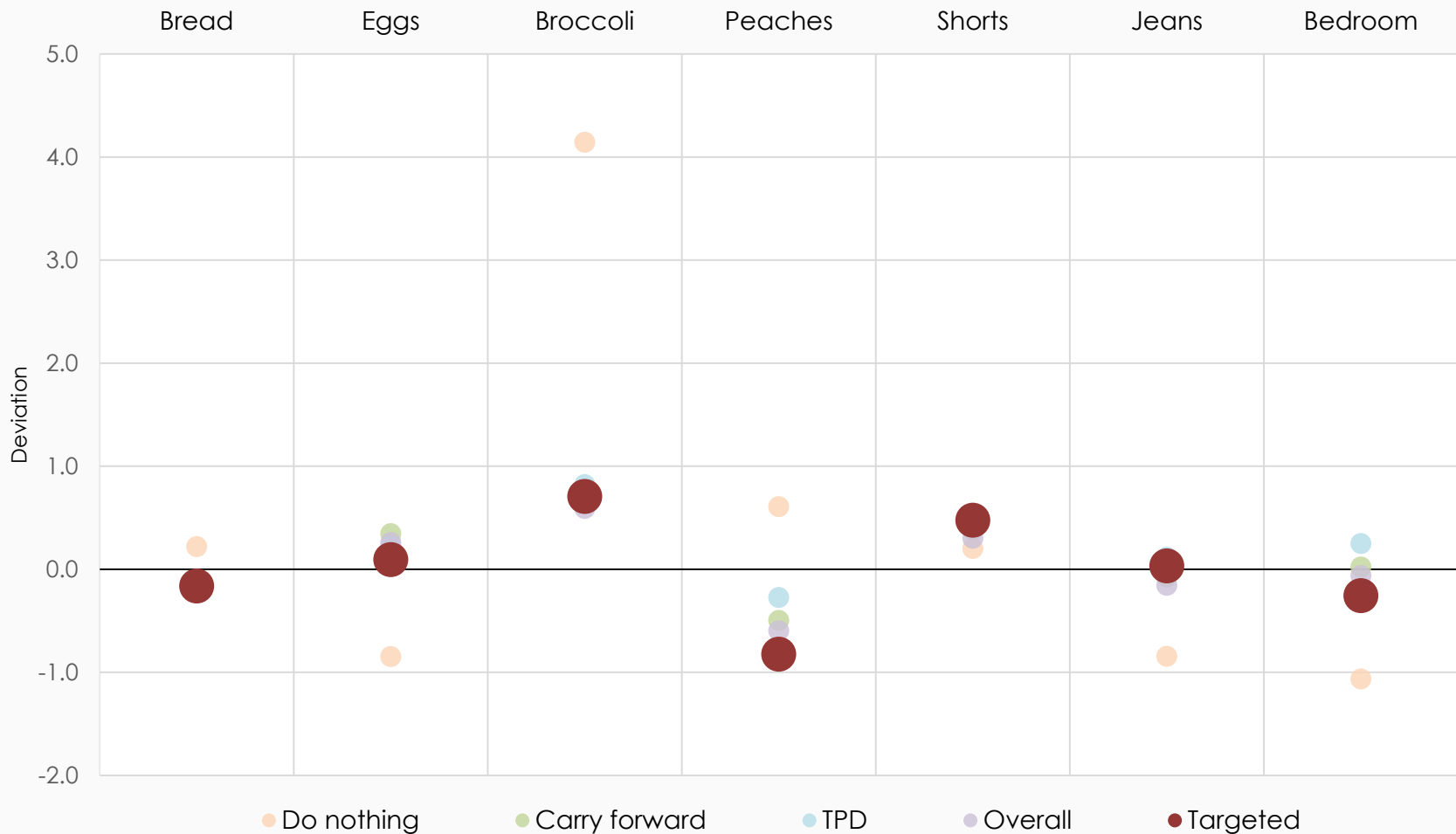
Index level: Deviation of index by imputation type – TPD



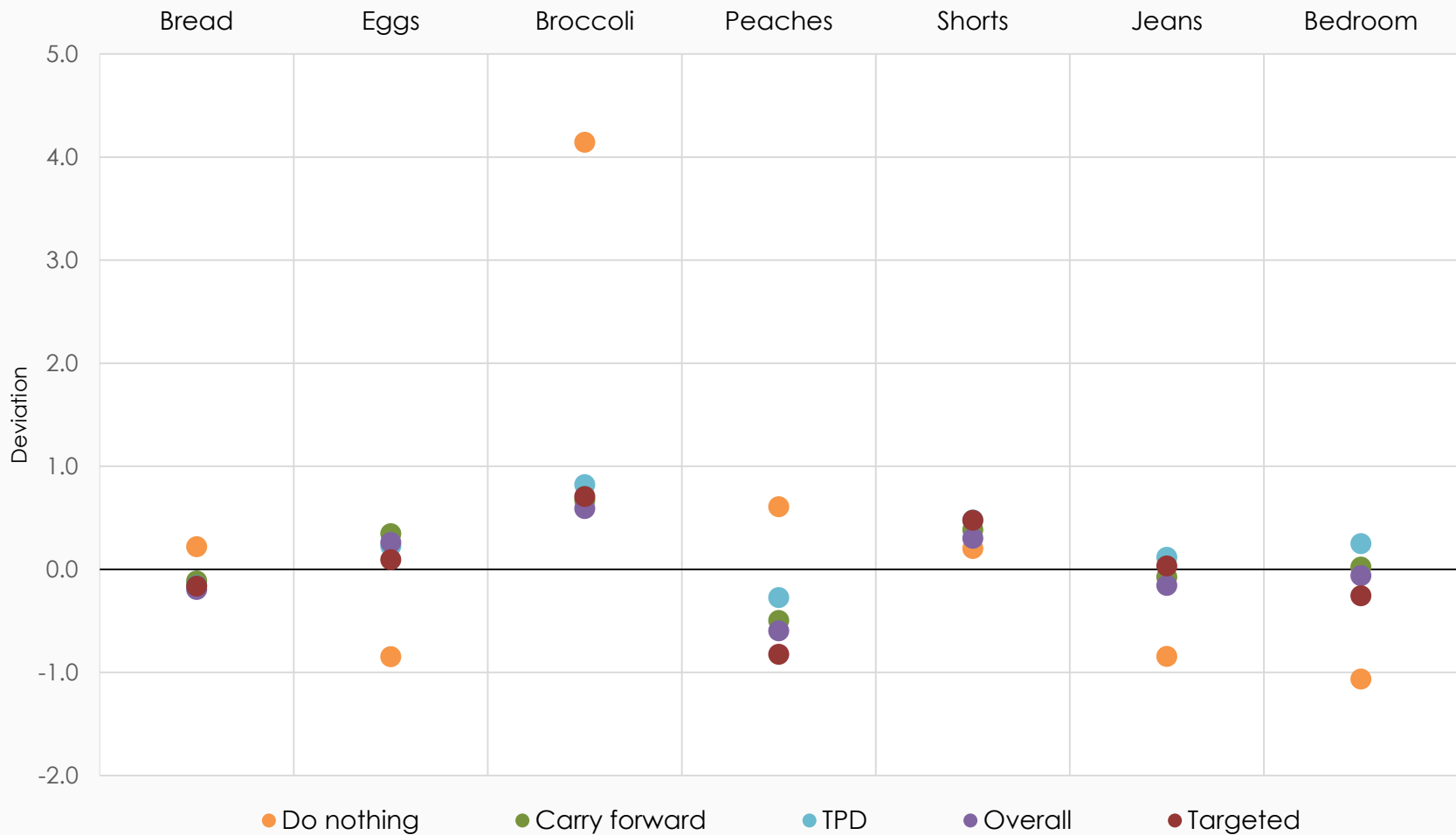
Index level: Deviation of index by imputation type – Overall



Index level: Deviation of index by imputation type – Targeted



Index level: Deviation of index by imputation type



Conclusions

Conclusions

- Confirm overall and targeted mean as most reliable methods
- Overall mean performed better with seasonal items
- Overall mean has larger sample – ensure targeted mean has adequate observations
- Performance of TPD was mixed – needs further investigation
- Carry forward performed well for sticky prices – but will miss change when it does happen
- Do nothing creates biggest bias – Not meeting criteria of ‘do no harm’

The end



Thank you

