Advisory Panel on Consumer Prices Technical Panel

11 April 2025

Office for National Statistics

Agenda

Time	Item	Presenter & Paper	Description
10:00	Introductions, apologies, and actions	Stephen Burgess	Make introductions if necessary. Inform Panel members of any apologies. Update on outstanding actions.
10:05 – 10:30	HPI imputation impact analysis	Aimee North, Malik Khalid	Update on proposed improvements to HPI's monthly imputation
10:30 – 11:00	Retailer type stratification	Mario Spina	Update on transformation work
11:00 – 11:10	**Tea break**		
11:10 – 11:30	Update on Grocery scanner impact analysis	Emily Hopson	Presentation to show the high-level impacts of the introduction grocery scanner data to CPI, CPIH and RPI.
11:30 – 12:15	Business Prices methodology	Andrew Carey, Markus Sova	Chain-linking methods use in PPI
12:15 – 12:35	**Lunch**		
12:35 – 12:55	Panel membership	Stephen Burgess	Standard review of membership
12:55	Publication status of papers & AOB	Stephen Burgess	SA CPI update - Stephen Burgess
13:00	Meeting close		

Actions from previous meeting

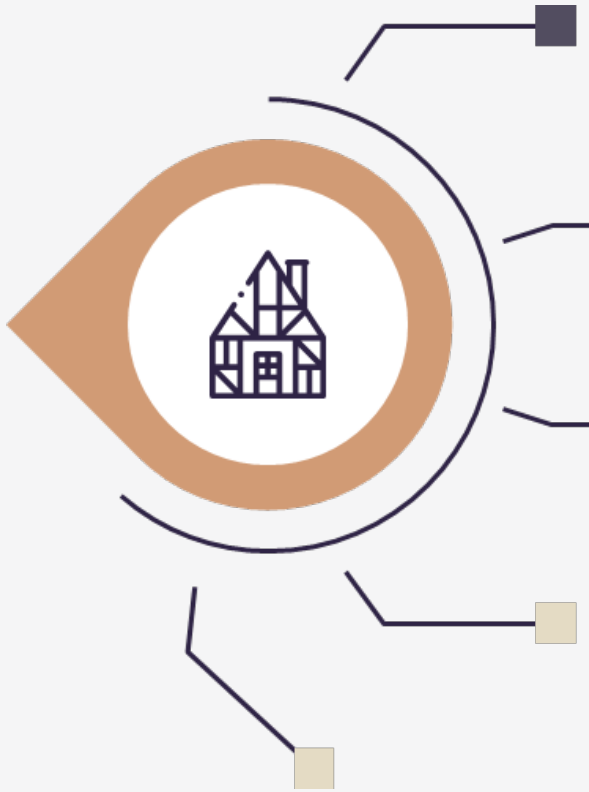
No.	Action	Person responsible	Progress
APCP-T (January 2025)			
1	ONS to set up follow ups with Huw and Monica	Chris Jenkins	Complete
2	ONS to resend GEKS-T video to panel to receive feedback by correspondence	Secretariat and Panel members.	Complete
3	ONS to send details on how to access MoveIT (will be referred to as secure file transfer software on official notes) and instruction files	Secretariat	Complete
4	ONS to update APCP-T on plans for improving the HPI monthly imputation	Aimee North/Secretariat	At today's meeting

APCP-T: Improvements to UK HPI monthly imputation

Aimee North and Malik Khalid
Housing Market Indices team
Prices Division

11 April 2025

UK HPI monthly imputation improvement



OBJECTIVE

What are we trying to achieve?
Why is it important?

SITUATION

Overview of challenges and current mitigation methods
Investigation into attributes data missingness
How improving imputation can help

CHOICES

Current method (missing category, set floor area to zero)
CanCEIS
K-Nearest neighbours (KNN)
Median

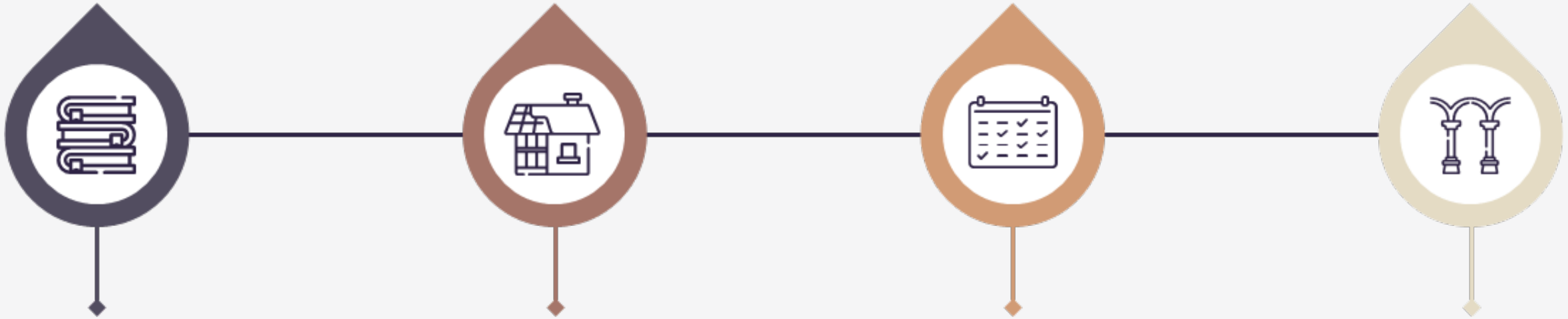
ACTIONS

How did we assess different imputation choices?
How did we evaluate performance?

RESULTS

Which is our selected method and why?
What is the impact on UK HPI revisions?

Objective



Sources of revisions: delays receiving property transactions and attributes data

Effect: over-estimation in early provisional estimates of new build price in Great Britain

Impact: revisions in the new build and headline UK HPI estimates

Solution: improve HPI's monthly imputation for Great Britain, proposed in Jan-25 APCP-T(25)01 paper

Situation

Existing approaches utilised in UK HPI to mitigate against over-estimation of new build prices in early provisional estimates:

- Temporary **pooling** of new builds (England and Wales only, 2nd to 7th estimate only)
- **Dampening factor** (2nd estimate only)
- **Reduced estimation model** (1st estimate only)
- Price **ratio** (1st estimate only)

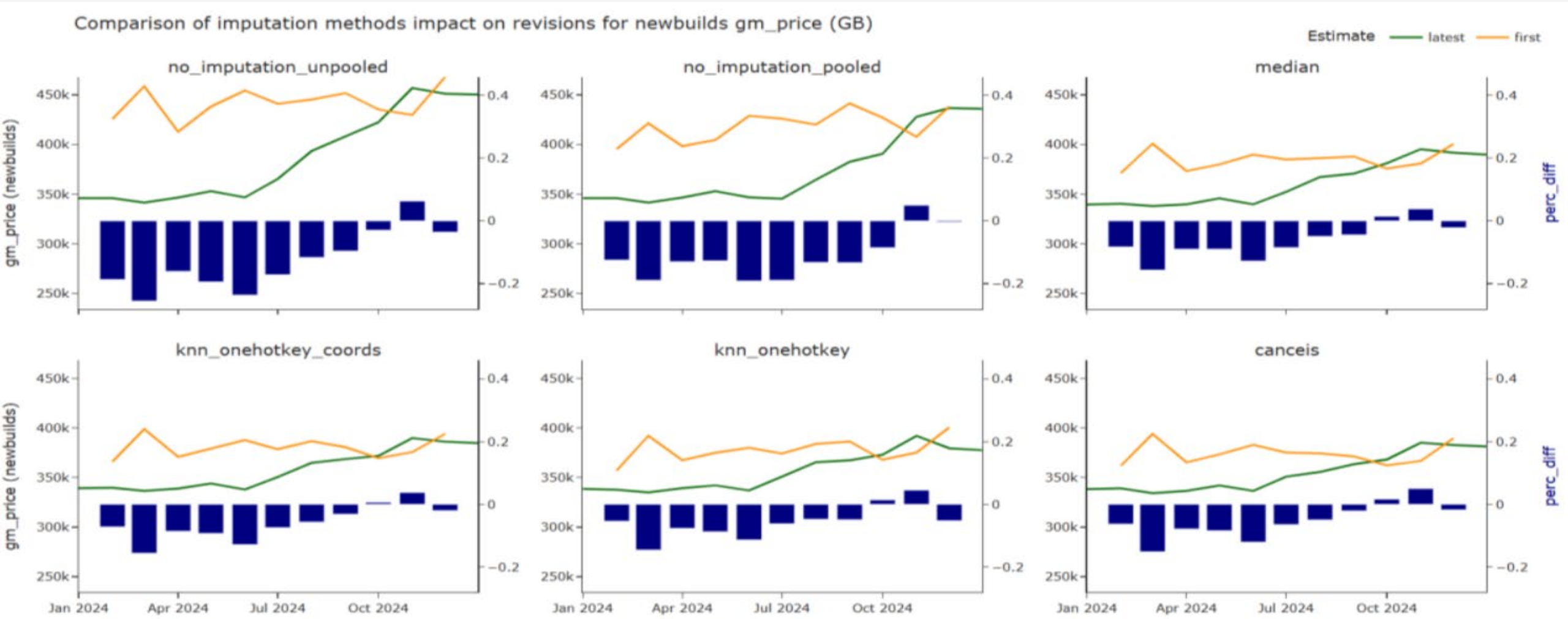
Current handling of missing values:

- Floor area = 0
- Categorical variables = “missing label”

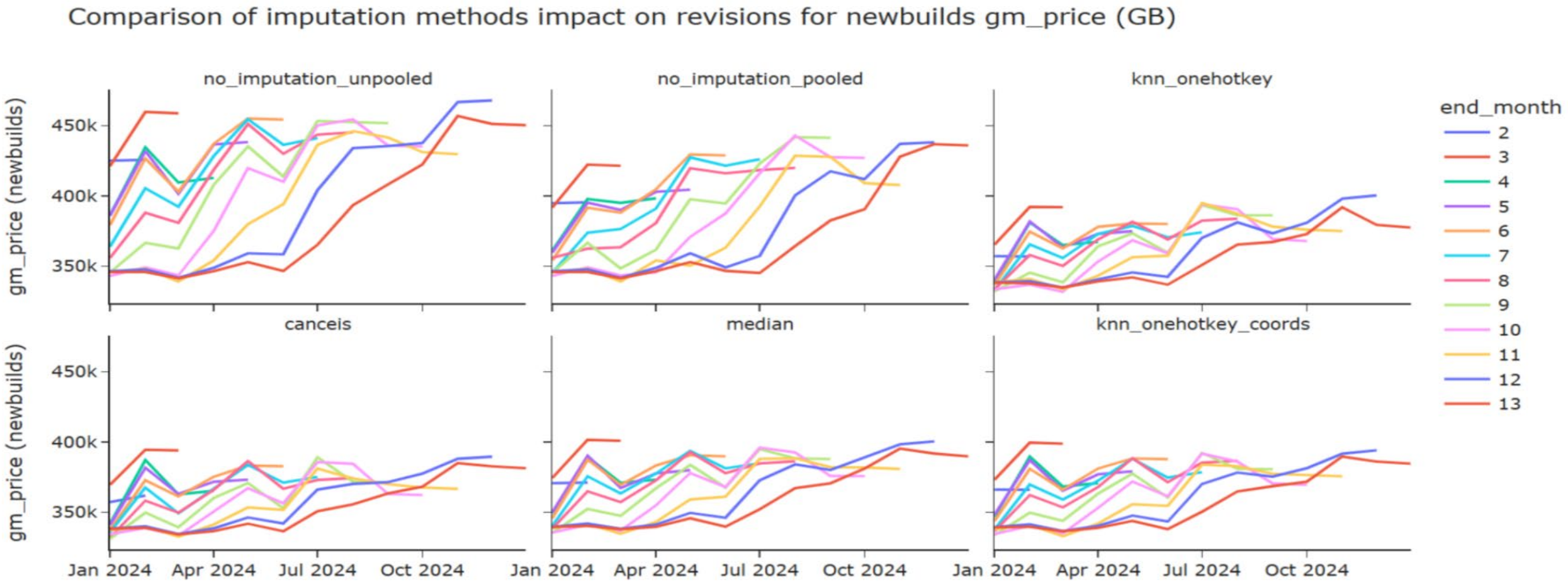
➔ Leads to over-estimation of the new build regression coefficient

➔ Drives over-estimation of new build prices

Analysis: first vs latest new build price, GB



Analysis: impact on successive revisions



Results

- Implementation of any of the tested imputation approaches would, by a similar magnitude:
  - **Improve accuracy** of early provisional estimates of new build price.
  - **Reduce overall revision size** in new build and headline UK HPI estimates.
  - **Be more effective** than the temporary ‘pooling’ approach at reducing revisions in early provisional estimates.

- KNN imputation with one-hot-key encoding and postcode coordinates was the most effective scenario tested. It:
  - Produces more stable and accurate early provisional price estimates, which are closer to the final estimate.
  - Preserves information without assuming any ordinal relationship between categories, thus avoiding biases.
  - Is computationally efficient, scalable and fast, suitable for use in monthly production time pressures.

- Proposal: in 2025, to **discontinue the temporary new build ‘pooling’ approach** and simultaneously **implement KNN imputation** as follows…

Proposed improvement to monthly imputation

Variable with missing values	Current	Proposed
 Floor area	England and Wales: Set to zero Scotland: Set to zero	England and Wales: Imputed using 10 nearest neighbours with matching property characteristics. Scotland: Set to zero
 Number of rooms	Assigned “missing label”	Imputed using 10 nearest neighbours with matching property characteristics
 Property type, Acorn	Assigned “missing label”	Assigned “missing label”
 Local authority code, New/Old	Never missing	Never missing

Request of the Panel

- Does the Panel support ONS' proposal to implement improved monthly imputation in the UK HPI, using the selected KNN imputation model?
- Does the Panel support ONS' intention to simultaneously discontinue the temporary 'new build pooling' approach introduced in 2020?
- Feedback ONS' intention to review the annual imputation method used in UK HPI, considering replacing the existing CanCEIS method with KNN or another alternative.

Retailer type stratification

Mario Spina

Office for National Statistics

Aim of the meeting and state of art

- Update on recent work on retailer stratification strategy
- Question: decide whether to continue exploring implicit weighting

- Paper taken to APCP in July 2021 (see pages 5-9): Previous paper
- Final decision required impact analysis. Work was paused to:
  - Prioritise local collection transformation
  - Put in place robust QA

Note – shorter presentation to allow time for discussion

Methods overview

Current approach

- Traditionally, low-level elementary aggregates are stratified on a combination of region, and retailer type.
- Each representative item in the basket is stratified by either:
    1. No stratification
    2. Stratification by region only
    3. Stratification by retailer only
    4. Stratification by both region and retailer
- The different retailer stratification options are relevant for types 3 and 4.

Additional reading: New methods for the calculation of retailer weights

Retailer stratification options

- For items stratified by retailer, we are exploring three options:
  - Multiple/independent (method currently used)
  - Big/small
  - No stratification
    - Assuming implicit weights represent our current retailer type weights

- Retailer type weights are calculated from the ABS, with a three-year gap.
- All items stratified by retailer type represented within the ABS.
- Each option is applied to the item stratified by retailer in the whole basket
- For dual-collection, avoid double weighting by using

$$\mathrm{W}_{\text{other type}} = W_{type} - W_{type,\,ADS}$$

Analysis of implicit weights

Office for National Statistics

Implicit weights analysis

- The no stratification method holds if implicit weights are representative of retailer type weights.
  - We need to verify this assumption

- This is important when including ADS data, as the ADS retailer weight is their market share.
  - Discrepancies in the weights might introduce bias in the index

Implicit weights analysis

- We calculate the ratio between the implicit weight and ABS retailer weight according to

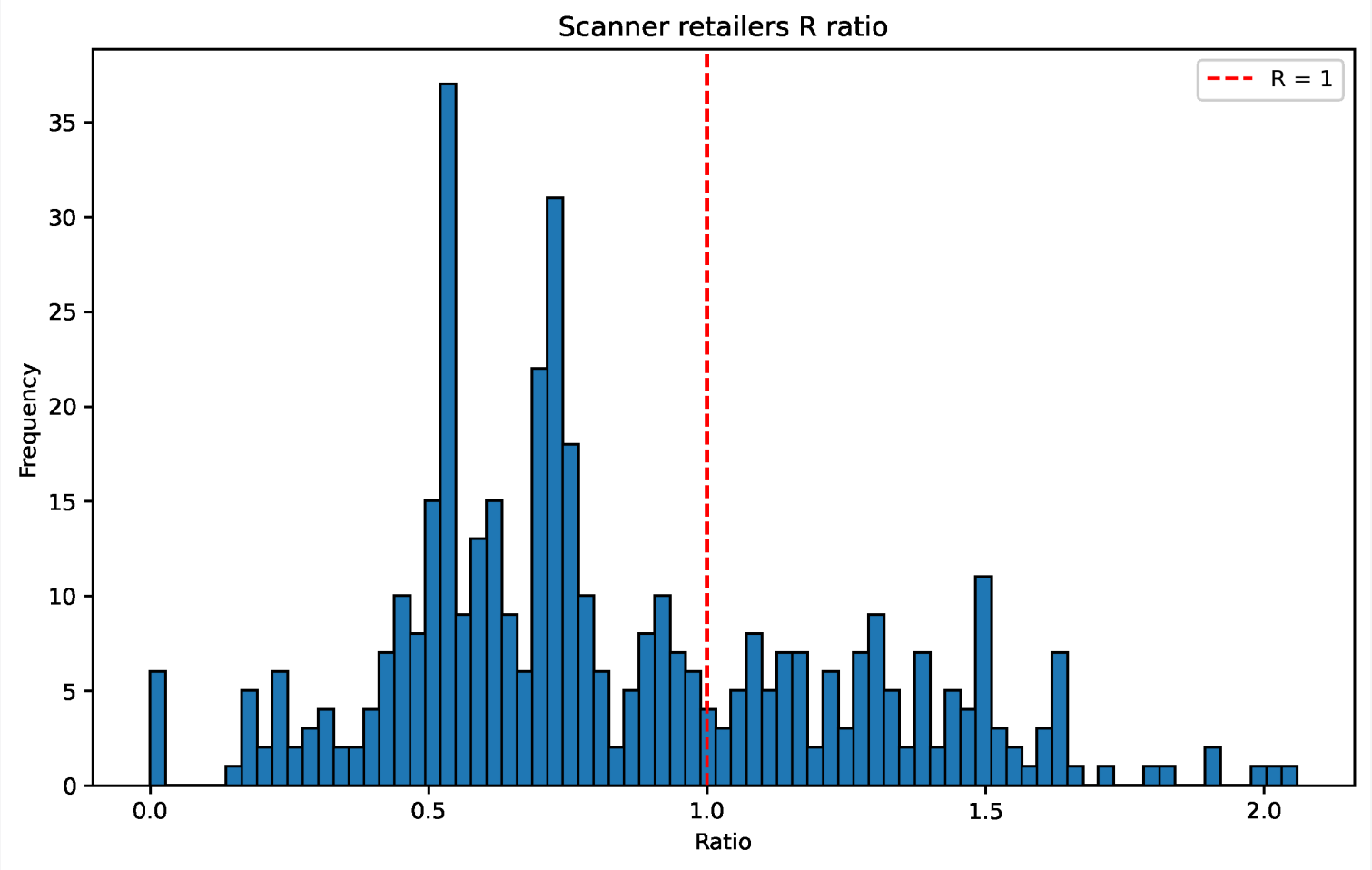
$$R = \frac{w_{type}}{w_{implicit}}$$

  - For distributions of R incompatible with 1, implicit weights do not represent the retailer type weight

- Focus on two case studies:
  - Ratio with scanner retailer for grocery items
  - Ratio of independent weights for clothing and footwear items

Implicit weights analysis

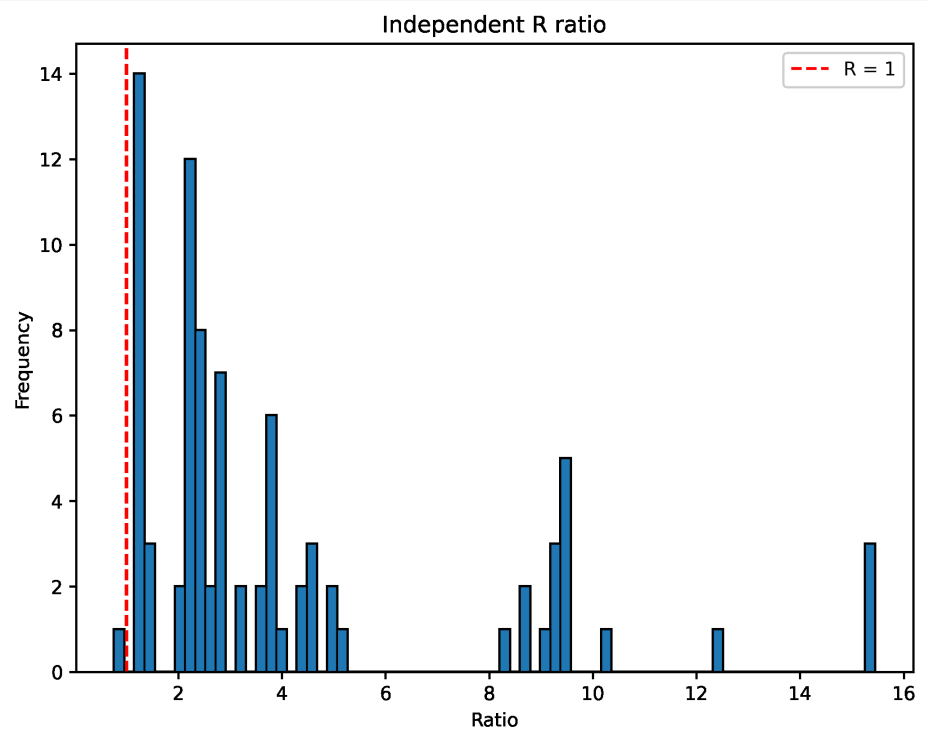
- Case study: R ratio for grocery scanner retailers

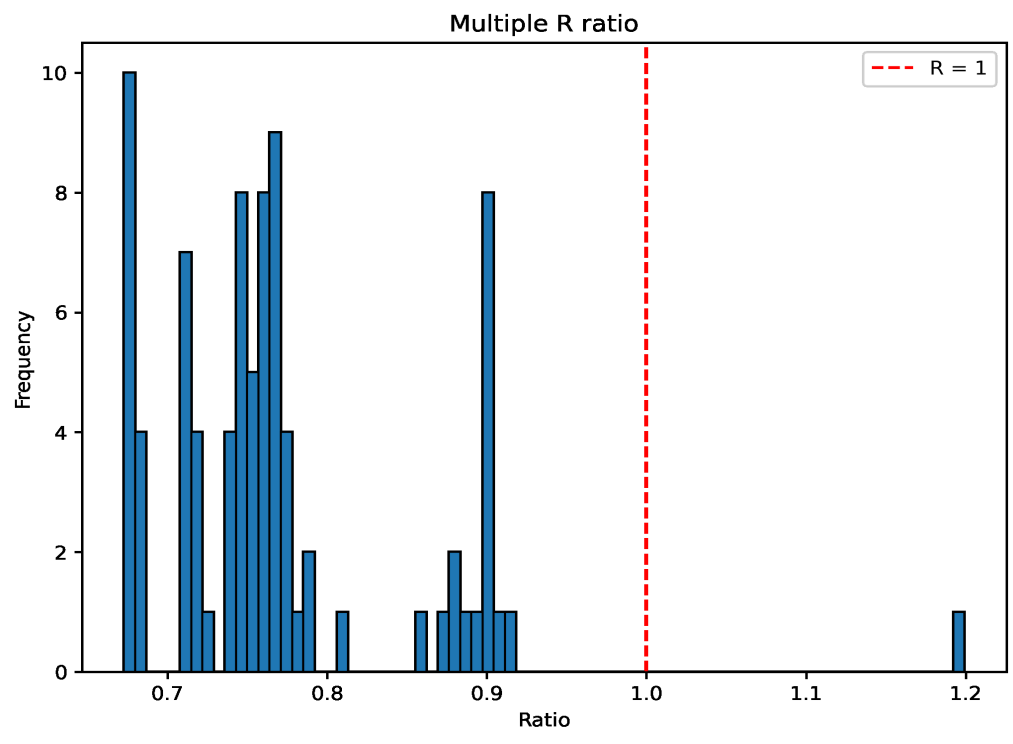


- R calculated on weights for the scanner retailers.
- Broad distribution away from 1
- Implicit weights don't represent ADS weights.

Implicit weights analysis

- Case study: R ratio for clothing and footwear





- R calculated on independent weights for traditional collection
- Broad and skewed distribution away from 1
- Implicit weights don't represent retailer type weights.

Limitations of implicit weights

- Sampling frequency within a location based on square footage as a proxy for expenditure, to maximise coverage while minimising outlet visits.
- Prices data is collected to capture accurately price movements, and not retailers' market share.
- The three-years gap with ABS data causes comparing weights calculated in different years for ADS retailers.
- Maintenance in the long-term is complex when introducing new scanner retailers or transforming new categories.
  - Implicit weight could be less representative of new data providers or categories.
- Distribution of ratio of weights consistently different from 1.

Question for panel

Do we keep working on implicit weight?

Indicative timelines

- July – update panel with full methods paper with impact analysis, with and without scanner data
- December – publish updated scanner data impact analysis including new retailer stratification method

Summary

- Updates on retailer stratification work
  - Presented options to our current strategy
- Initial analysis on implicit weight
  - Question for panel on whether keep working on implicit weight
- Next steps timeline

[REDACTED]

Break

11:00 – 11:10

APCP-T: Grocery scanner data impact analysis

Emily Hopson

Business Change and Implementation Lead

Prices Division

11 April 2025

Office for National Statistics

Background

- From February 2025 (published March 2025), we introduced:
  1. improved imputation methods
  2. consumption segments
  3. Northern Ireland private rental price statistics

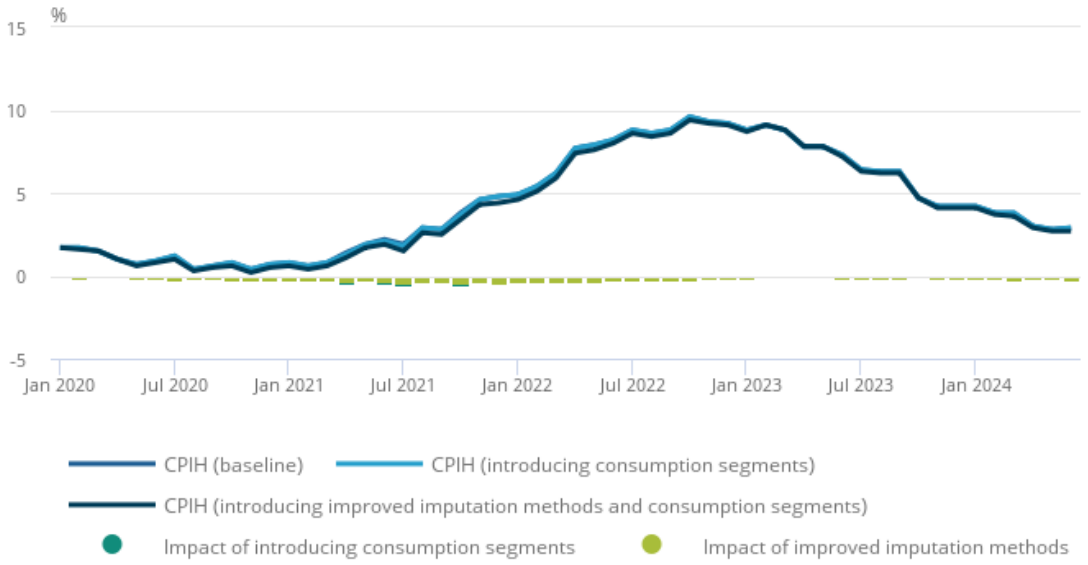
  into our UK consumer price statistics.
- Due to the complexity of data and methods and the importance of these statistics, we took the decision to **delay the introduction of grocery scanner data**.
- Further quality assurance is required to ensure they are of the **highest quality**.
- We plan to parallel run groceries scanner data for a year before incorporating into live production in **February 2026**.
- Impact analysis including grocery scanner data publication on **29th April.**
  - **This presentation is for your information.**

How we are measuring impact of scanner data

- We present the differences between our series incorporating groceries scanner data and our series using **the improved methods** and **consumption segments** we introduced this year.
- Demonstrate the impact of introducing groceries scanner data **in isolation** from the other changes we introduced in February 2025.

Figure 4: The average impact to the annual rate of change of CPIH was negative 0.17 percentage points

Indicative impact of transformed methodology on Consumer Prices Index including owner occupiers' housing costs (CPIH) 12-month rate, UK, January 2020 to June 2024



Source: Impact analysis of transforming UK consumer price statistics from the Office for National Statistics

Reasons for differences

Different product coverage

Different treatment of promotions

Different data frequency

Different index methodologies

Headline impact CPI and RPI

Office for National Statistics

[REDACTED]

Average impacts on the annual and monthly rate of change of CPI and RPI

CPI

Year	Average annual growth difference (percentage points)	Average monthly growth difference (percentage points)
2019	-	0.00
2020	0.03	0.01
2021	0.04	0.00
2022	-0.05	0.00
2023	-0.12	-0.01
2024 (up to June)	-0.14	0.01

RPI

Year	Average annual growth difference (percentage points)	Average monthly growth difference (percentage points)
2019	-	0.00
2020	0.03	0.01
2021	0.14	0.00
2022	-0.11	-0.01
2023	-0.11	-0.01
2024 (up to June)	-0.13	0.01

[REDACTED]

CPI deep dive

Consumption segment case study

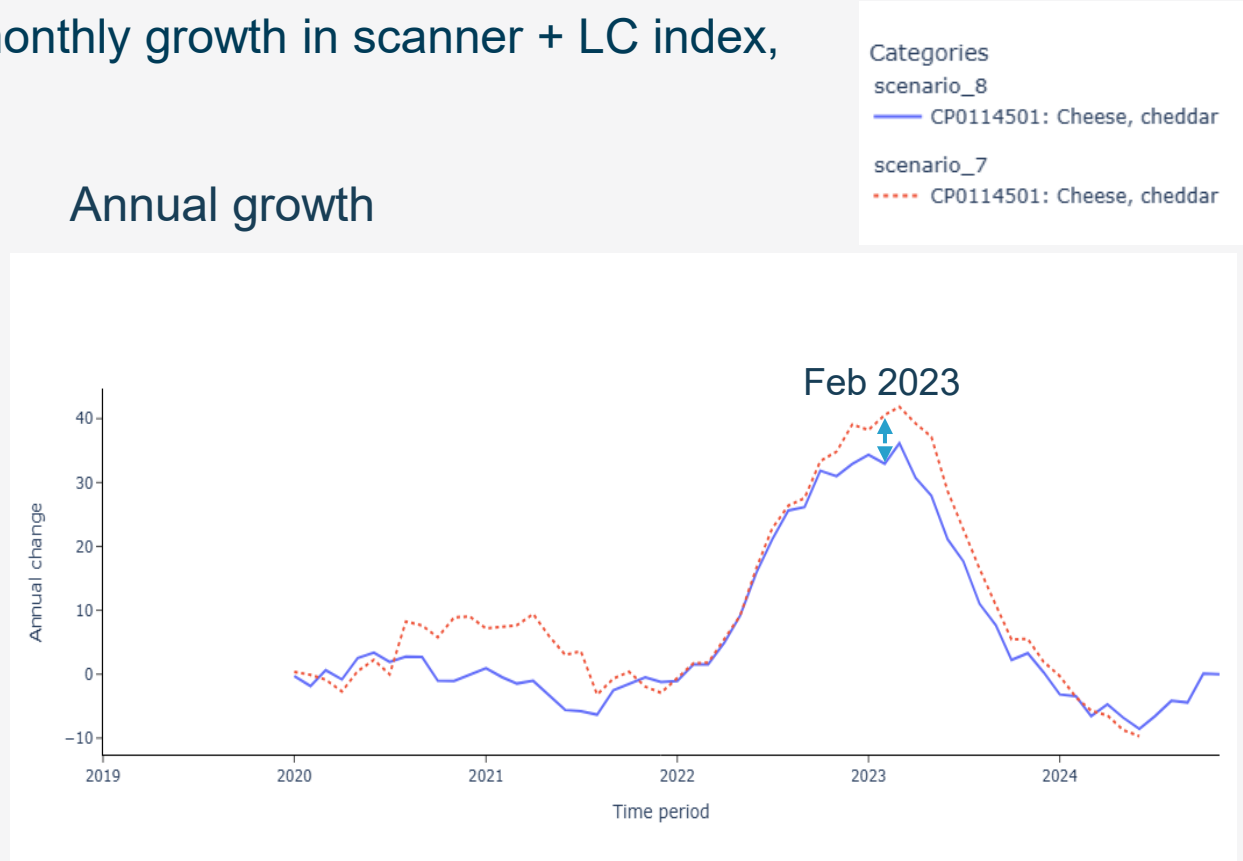
Cheddar cheese

Office for National Statistics

Scanner data and LC had different movements in Feb 2023

In Feb 2023, Cheddar cheese index had negative monthly growth in scanner + LC index, positive monthly growth in LC index

[REDACTED]



Promotional impact causing differences

Difference caused by promotion on cheddar cheese in that month

Questions

Business Prices methodology

Andrew Carey, Markus Sova

Office for National Statistics

Objective

i. Present proposed methodology correction.

ii. Receive feedback/validation regarding methodology.

iii. Provide sign off on proposed changes.

Summary

Summary

- Issue has been present in PPI data since the introduction of chain-linking in 2020.
- The error relates to the failure to introduce a new base period when sales are updated annually.
- The base period should align with the year of the sales data.
- There is also an error in how the sales data is updated to reflect the weights year.
- Publication of PPI data has paused until correction applied.

Methodology

Chain-linking and price updating

Office for National Statistics

Methodology

International best practice is to calculate PPI or SPPI as an annually chain-linked Laspeyres price index. The Laspeyres formula can be expressed:

$$\sum_i R_{bti} w_{bi}$$

where R_{bti} is the price relative for item *i*, comparing current period *t* to base period *b,*

w_{bi} is the weight for item *i* in base period *b* (the weight is the proportion of base period turnover attributable to the item),

and the summation is over all items being aggregated.

To chain-link, we start with a set of Laspeyres price index series, each calculated using a different base period:

$$\sum_i R_{b_j t i} w_{b_j i}$$

where $\{b_j\}$ is the set of base periods, for PPI these should be the years 2015, 2016, 2017…

Finally, each series is linked to its predecessor by a simple scaling to give a single chain-linked series . Thus chain-linking introduces a new base period every year into the series.

The error identified in the PPI calculations is in the formula used for the set of price index series, which is:

$$\sum_{i} R_{bti} w_{a_j i}$$

where base period *b* has been fixed at 2010, but $\{a_j\}$ is the set of years {2015, 2016, 2017…} for which weights are calculated.

Price updating methodology

- Sales used to calculate weights should correspond with each base year.
- Since suitable sales are not available, data from the previous year is used to estimate the correct sales.
- The sales estimate used should be calculated using the growth rate between the annual average of the source year with the annual average of the base year.
- Currently the link month is used instead of the average of the base year.

Impact

[REDACTED]

Summary

Summary of methods changes

- Apply correction to chain-linking methodology to introduce a new base period when sales are updated annually.
- Ensure that the base period aligns with the year of the sales data.
- Correct the error in calculating the growth rate applied to estimate sales in the base year.

Lunch

12:15 – 12:35

[REDACTED]

Panel Membership

Stephen Burgess

Office for National Statistics

AOB

- Seasonally adjusted CPI update – Stephen Burgess