ADVISORY PANEL ON CONSUMER PRICES – TECHNICAL

Experimental methodology for regional CPI for Northern Ireland

Expected publication: Alongside minutes.

Purpose

1. The Office for National Statistics (ONS) is working in collaboration with the Department for the Economy (DfE), Consumer Council for Northern Ireland (CCNI) and Northern Ireland Statistics and Research Agency (NISRA) to produce experimental consumer price inflation estimates for Northern Ireland (NI), by boosting the local price collection sample. The aim of this paper is to get the panels input on the methods chosen for creating a regional CPI for Northern Ireland.

Actions

2. Members of the Panel are invited to advise on:
   a) The implementation of the NI price boost
   b) Regional Weighting Methodology
   c) Dealing with sample size issues and volatility
   d) Next steps to take for the project

Background

3. The Office for National Statistics (ONS) is working in collaboration with Department for the Economy (DfE), the Consumer Council for Northern Ireland (CCNI) and Northern Ireland Statistics and Research Agency (NISRA) to produce experimental consumer price inflation estimates for Northern Ireland (NI).
4. In our first publication estimating Northern Ireland prices we highlighted the small sample size for price quotes as a challenge for constructing a Northern Ireland Consumer Prices Index.
5. Locations for local collection of prices are assigned based on the level of consumer spending within that region or country, so the price collection in Northern Ireland is a small fraction of the national sample size. The smaller sample sizes would lead to a much less accurate regional index, both in the context of price movements in the region and the representativeness of spending in the region.
6. To try and overcome the issue of a small sample size we have collaborated with the DfE, CCNI and NISRA to boost the sample size of the price collection in Northern Ireland, mainly the local collection (prices that are collected from local shopping outlets). Every month from January 2022, we have aimed to double the collection in existing outlets, which broadly doubles the local sample in Northern Ireland.
7. For expenditure weights, Living Costs and Food Survey (LCF) data can be reconciled to Consumer Price Inflation spending totals to provide a regional breakdown. However, disaggregating the LCF sample into 12 regions means that regional sample sizes are much smaller than the national sample size. The smaller sample sizes would lead to a much less accurate regional index.

Therefore, we have introduced experimental regional weighting as detailed in out last publication:
https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/boostingthenorthernirelandpricesampleforthecconsumerpricesindexincludingexperimentalregionalweighting

Section 1: Implementation of the NI price boost

8. There are two basic price collection methods: local and central. Local collection is used for roughly half of the CPI basket, by weight; prices are obtained from outlets in 141 locations around the country, with over 100,000 quotations obtained monthly by this method. Normally, price collectors must visit the outlet, but prices for some items may be collected by telephone.
   The local collection is sampled. This is done in the following way:
   i. A sampling frame of locations is defined across the country.
   ii. Locations are sampled with probability proportional to expenditure, stratified by region
   iii. Within each location, a sample frame of outlets is drawn up. This is drawn up by price collectors in the field who start in the centre of a spending hotspot and work outwards up to 1500 outlets.
   iv. Outlets are selected with probability proportional to expenditure (based on floor space as a proxy), and basket items are "placed" in sampled outlets.

   Central collection is used for items where we can collect all prices centrally (within the head office) with no field work. These collections can be further sub-divided into two categories:
   I. Central shops, where the prices come from retailers with national pricing (as in, the retailers assign the same price for a product in each region) and these prices are combined with prices obtained from the local collection to produce the price index
   II. Central items - typically web collections, where we collect the prices centrally and they are used on their own to construct centrally calculated price indices.

   See chapters 5 and 6 of our Technical Manual for more info: Consumer Prices Indices Technical Manual, 2019 - Office for National Statistics).

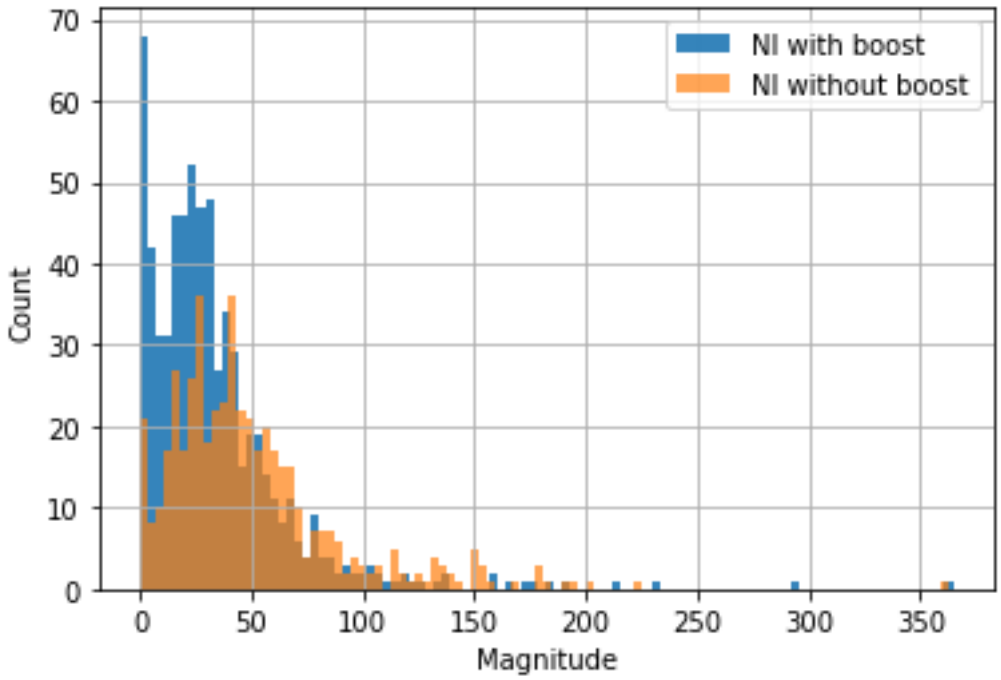
9. For regional indices generally we have considered two options for regional pilot boost collections. One option was to double the number of geographic locations we collect from in the region and recruit shops in those new locations - effectively doubling the sample. However, this wasn't possible in Northern Ireland given that there are far fewer locations on the sampling frame, and most are already in the sample. For Northern Ireland therefore we have doubled the collection in existing outlets. In order to improve the coverage of the boost collection we have also recruited new shops in existing geographic locations to collect price data where it was not possible to collect a second quote in the existing shop. New shops have steadily been added to the boost and coverage has increased to 86% in 2024 up from around 76% in 2022.

10. Only the local collection was boosted. Centrally collected items – that is, items collected by ONS staff from websites, or by email contact or phone – were not included in the boost.

Many centrally collected items are based on the assumption that a national pricing policy applies. However, some central collections have a regional aspect to them, which is tracked in the collection, but we have been unable to boost the sample for a number of items that are centrally collected. In most cases, this is unlikely to result in any significant bias. However, through our engagement with the Consumer Council for Northern Ireland (Consumer Council), we have identified a number of items that may be of greater importance to the Northern Ireland economy. For example, the Northern Ireland Department for the Economy's report highlights that 68% of Northern Ireland households are reliant on Kerosene oil as a source of home heating, compared with 4% of the UK as a whole. Some of these items can be effectively disaggregated using the existing data to give a reasonable price index for Northern Ireland and will therefore be incorporated into the Northern Ireland index as part of ongoing development.

11. We have not been able to investigate how much the boosted sample size is benefiting the results. It is clear that it improves the quality of the Northern Ireland CPI, however, work that we have been doing with Southampton University on producing standard errors will help us to define the true impact of the boost and what sample sizes are needed per region or country.
12. In figure 1 we compare the distribution of the total absolute magnitude in price change of items in NI month to month, with and without the boost, in order to see how much item prices are moving each month. The addition of the boost data reduces the magnitude of price changes month to month, reducing volatility of NI prices. The magnitude on the x axis in fig 1 is the sum of the absolute value of the monthly changes. For example, for a given item if the price index was 100 in January, 90 in February (-10), 105 in march (+15) and then stayed at 105 until November the total absolute magnitude would be 25 (i,e, 10+15+0+..+0).

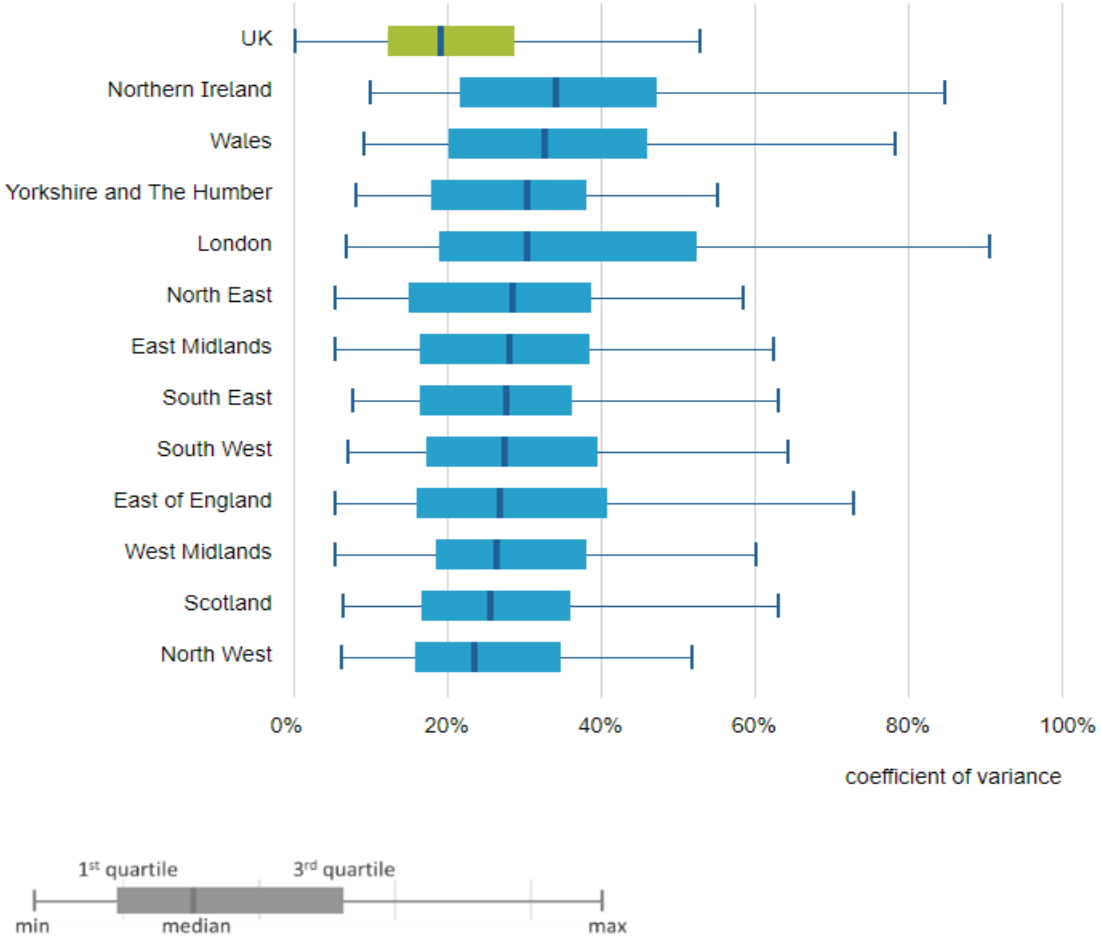
Figure 1. Northern Ireland with and without Boost data: item level price change distribution, January 2022 – November 2022



Section 2: Regional Weighting Methodology

13. Expenditure data for the UK-level weights at Classification of Individual Consumption by Purpose Level 5 (COICOP5) and above are drawn from Household Financial Consumption Expenditure in the National Accounts. For subnational expenditure weights, Living Costs and Food Survey (LCF) data can be reconciled to CPI spending totals to provide a regional breakdown. Similar work was done to calculate CPIH-consistent inflation rates for UK household groups and we aligned with this methodology. However, disaggregating the LCF sample into country and regions means that the sample sizes are much smaller than the national sample size, particularly in Northern Ireland.
14. Previous research by Southampton University looked at using a rolling three-year average as a smoothing technique but found the weight series were still quite volatile. This presents us with an inherent trade-off between having region-specific weights that are excessively volatile or, alternatively, using UK-specific weights that are less volatile but do not accurately represent Northern Ireland spending.
15. As it was not feasible to increase the sample size of each regional breakdown in previous years of LCF data, we have developed an experimental method of producing more robust regional weights.
16. The LCF sample collection varies across regions and countries of the UK and the sample size is based on the size and the population of the region. When expenditure data are broken down by region, the sample size is small leading to greater volatility in any regional weights produced from it.
17. Figure 2 shows the volatility across class-level weights (COICOP4) across UK regions. The UK has a much shorter box plotting meaning the volatility is considerably lower than individual regions, which have a much longer box plot. This is expected as the UK is the combination of all countries and regions, and therefore has the largest sample size. Northern Ireland has a much higher level of volatility across class-level weights than most other UK countries and regions. Given that Northern Ireland is the smallest region of the UK, this is to be expected. This also means that the data challenges from a NI CPI perspective are particularly acute.

Figure 2. The coefficient of variance of the weights timeseries at class level (COICOP 4) across all UK countries and English regions compared with the UK, January 2005 to February 2022.



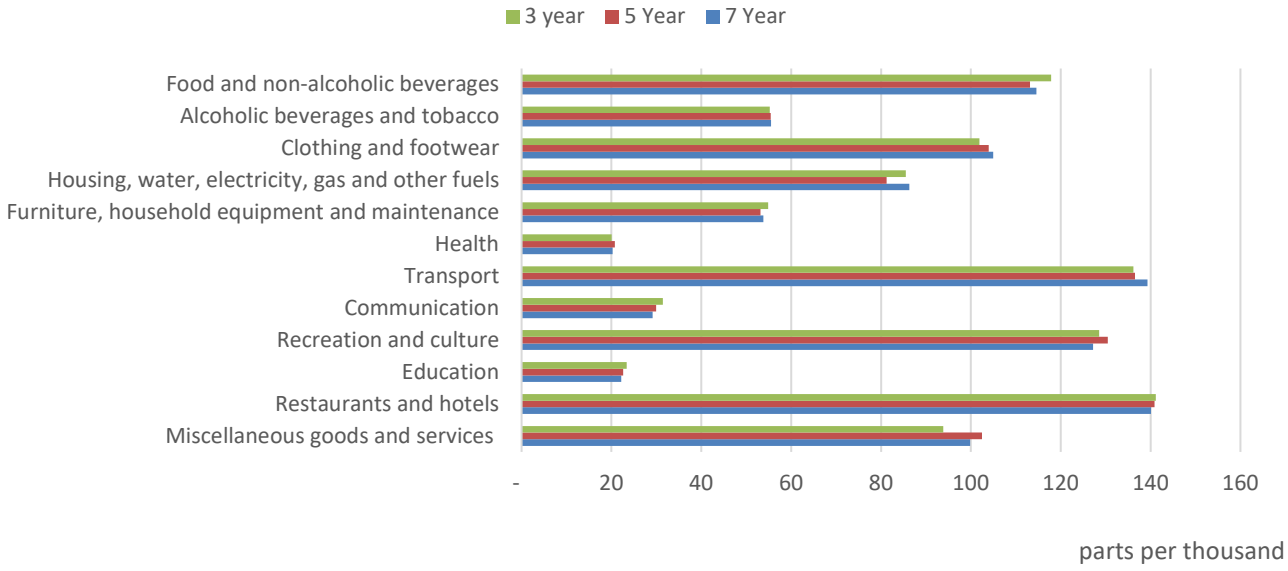
Reducing volatility of weights by using an average starting point

18. A small expenditure data sample for Northern Ireland leads to volatile Northern Ireland-specific weights. The volatility means that year-on-year movements are unlikely to represent true changes in spending habits in Northern Ireland, it can also make it difficult to show price changes over the year, since annual inflation rates are confounded by weights effects. We have therefore looked at using an experimental methodology to reduce the volatility of the Northern Ireland weights.
19. Under the assumption that the general distribution of spending in Northern Ireland differs to the UK, but changes over time are broadly similar, we looked to create a stable estimate of

the spending distribution in Northern Ireland at an appropriate point in time and then adjust it temporally using UK-wide annual changes. We therefore looked for a reasonable "starting point" of Northern Ireland expenditure to then adjust by UK weight changes.

20. The assumption that changes over time are broadly similar between the UK and Northern Ireland is a key one. We expect that over the short-term changes in spending patterns may be small but that differences may accumulate over the longer-term. If this approach were used on an ongoing basis, it would be necessary to have a periodic benchmarking exercise to realign the weights with the Northern Ireland expenditure distribution. This may create large structural breaks in the weights series leading to issues with annual rates in the benchmarking year. The revision policy may therefore be an important consideration. Nevertheless, the method was chosen as a way of managing the trade-off between volatility and representativeness.
21. Because of sudden changes in spending habits during the pandemic period (2020 to 2021), we only considered pre-coronavirus-pandemic starting points. We also did not want to pick a year that was too far back and be unrepresentative of current spending patterns in Northern Ireland. We chose 2015 as the least volatile year and most suitable "starting point".
22. To ensure the starting point was a good representative of each class weight, we applied a smoothing method to pool survey samples over several years. We investigated smoothing Northern Ireland weights around 2015 across three-, five- or seven-year periods. A five-year average was chosen as it gives a reasonable trade-off between robustness and relevancy. The series was then adjusted by annual changes in UK weights since the relevant period around 2015, and rescaled to sum to 1000, to create a set of Northern Ireland-specific weights.
23. There is a trade-off between averaging over more years but over-smoothing, and averaging over fewer years but not stripping out much volatility.

Figure 3. Three, five and seven-year averages of Northern Ireland weights centred around January 2015 by divisions. I.e. 2014-2016, 2013-2017 and 2012-2018 averages.



24. Based on this methodology, we have produced experimental Northern Ireland weights at division level. Table 1 shows these new weights for CPI aggregated to division, uprated to 2024, and how they compare to the UK.

Table 1: Experimental Northern Ireland expenditure weights compared with the overall UK weights used in the headline UK CPI inflation (parts per thousand), February 2024.

Division		Northern Ireland	UK	Difference
01	**Food and non-alcoholic beverages**	116.5	113.1	3.4
02	**Alcoholic beverages and tobacco**	49.6	39.1	10.5
03	**Clothing and footwear**	87.8	58.6	29.2
04	**Housing, water, electricity, gas and other fuels**	93.6	131.3	-37.7
05	**Furniture, household equipment and maintenance**	56.2	62.5	-6.2
06	**Health**	23.4	26.2	-2.8
07	**Transport**	126.4	135.5	-9.1
08	**Communication**	24.1	23.1	1.0
09	**Recreation and culture**	125.1	142.6	-17.5
10	**Education**	25.9	29.7	-3.8
11	**Restaurants and hotels**	165.3	145.1	20.2
12	**Miscellaneous goods and services**	105.9	93.1	12.8

25. Over time we would plan to rebase this 5 year average. In order to avoid impact of the covid pandemic we would need to take weight from 2024 onwards (which would be based on 2022 expenditure) so with a 5 year average this would not be able to be done until 2029. It would be interesting to see how our uprated 2015 weights would compare to this 2024-29 average.
26. For future work, we plan to look at a 3 year average 2022-2024 as this is the most recent data we have while still avoiding much of the covid pandemic years and seeing how this compares to our 2015 uprating. With this, the question remains on which weight is more

reliable. The 2022-2024 3-year average which is the most recent and relevant but potentially volatile; or the 2013-2017 5-year average which should be a less volatile, longer time series but not as relevant. The 2015 weights are also uprated using UK changes which may have differed from NI movements.

27. Table 2 shows the February 2022 and February 2023 Northern Ireland weights calculated directly (without smoothing) and compares the average of these against the 5 year 2015 based uprated weight.

Table 2: Experimental Northern Ireland expenditure weights compared with the Northern Ireland single year calculated directly (parts per thousand).

	Feb-22			Feb-23		
Division	Single Year NI CPI weight	Experimental weight	Difference	Single Year NI CPI weight	Experimental weight	Difference
1	131.3	121.3	10.1	144.8	123.5	21.4
2	49.0	64.4	-15.4	41.6	54.5	-13.0
3	67.6	91.6	-23.9	79.2	87.9	-8.8
4	110.4	99.0	11.4	86.0	102.9	-16.9
5	81.7	69.8	11.9	79.2	61.8	17.4
6	24.9	18.3	6.6	27.1	21.6	5.4
7	140.6	116.3	24.3	130.1	122.4	7.6
8	25.3	26.3	-1.0	25.8	24.3	1.5
9	113.3	120.8	-7.6	118.4	122.1	-3.7
10	32.1	29.2	2.9	28.5	25.5	3.0
11	141.8	132.9	8.9	161.6	158.9	2.7
12	81.9	110.0	-28.1	77.9	94.5	-16.6

Limitations

28. A larger sample size in expenditure data would support a more accurate weight for Northern Ireland, but this was not possible to obtain. The method we use here overcomes a small sample size in expenditure data using a methodology that is a trade-off between having Northern Ireland-specific weighting and having volatility in the data. However, there are still limitations to this experimental method:

    a) Northern Ireland trends are ignored completely once imputation/uprating begins in 2015 – if there are any differences between the UK and Northern Ireland economies reacting to events such as the withdrawal of the UK from the European Union or the covid pandemic, these will not be captured in the weights. This could be particularly important for items that are significantly different in NI than the UK, for example oil vs gas heating.

    b) There is a trade-off between having region-specific weights, with volatility over time, suggesting error, or using UK-specific weight changes that are reliable over time, but not relevant to Northern Ireland changes over time.

    c) Manual flagging of outliers is based on human judgment which is susceptible to errors – however human judgement is necessary.

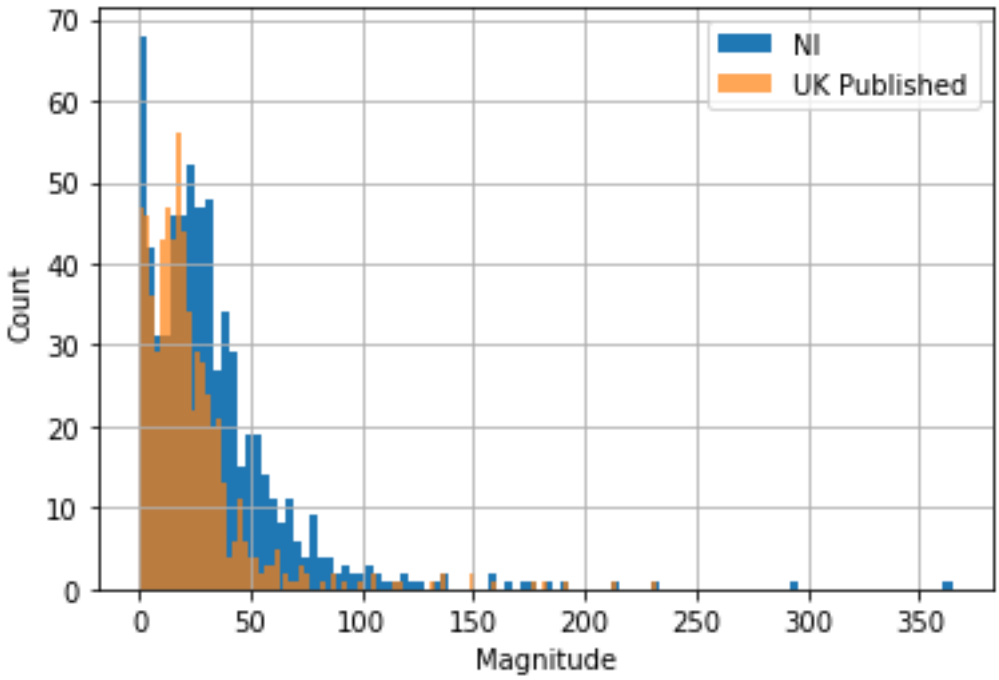
d) It is necessary to combine price indices for basket items to estimate an overall aggregate inflation rate. Items are aggregated according to an internationally defined structure: the Classification of Individual Consumption by Purpose (COICOP). However, below COICOP, at the basket level, the data are too granular to support reconciliation against the LCF data. As a result of these limitations, we used the national basket weights at the lowest level, rather than Northern Ireland specific regional weights. Volatility at item level is not inspected, however this aligns with the method used in CPIH-consistent inflation rate estimates for UK household groups publication.

e) The sample size for Northern Ireland in LCF is small and though averaging across five years reduces volatility, the combined sample size should still be treated with caution

Section 3: Sample Size issues and volatility

29. As there are only a few locations in the NI sample, doubling the number of quotes, won't necessarily result in a robust sample size, even if there is no missingness. For some items there are issues with low sample size and/or high volatility. How is this best dealt with? We have looked at item level prices and assessed the volatility over time, once items are identified as volatile, they can then be dealt with accordingly, for example, if a Northern Ireland item index is particularly volatile, the UK item index may be a better proxy.
30. A contributor to this volatility for some items is issues with sample sizes. In our first publication estimating Northern Ireland prices (see Figure 3), hotels and restaurants was the main driver of contributions to the difference between the NI CPI and the UK CPI. This was entirely driven by the index for hotels which only had one unusually large price relative for Northern Ireland. For cases like these we may prefer to replace the NI index with the UK index.
31. Following from the same method from figure 1 we investigated individual items price indices month on month and compared the total absolute magnitude of the index price change between the UK and NI. This gave us a sense of how much item prices are moving each month. This suggests that Northern Ireland has a higher number of volatile items. However, on it's own this indicator is not enough to indicate that the sample size is impacted on the variance of the index, as it may be the case that price movements are larger in Northern Ireland than in the UK as a whole.

Figure 4. UK vs NI item level price change distribution, January 2022 – November 2022



32. We intend to develop criteria based on item volatility, sample size and UK comparison in order to create a framework for dealing with potential outliers in the Northern Ireland data, looking at when to use NI prices and when they would be better replaced with the UK index. It is important that if NI does experience greater price changes, that we are not stripping out genuine inflationary effects. As part of looking at item volatility it would be important to look at the magnitude and frequency of change in direction of prices. If there are large price changes but these vary in direction, and the sample size is small, then it may be sampling volatility rather than a continuous large change in the same direction.

Section 4: Future Work

33. There is a range of areas we would like to explore for further work. This includes:
    - Creating control and logic for when to use NI price vs UK price indices for items.
    - Create a mechanism to include Northern Ireland specific central price collections. For example, using https://www.consumercouncil.org.uk/home-heating/price-checker rather than UK price collection or its regional element with a low sample size.
    - Investigate the assumption that changes over time are broadly similar between the UK and Northern Ireland
    - Investigate Republic of Ireland CPI changes and whether this would be a better proxy for Northern Ireland uprating than UK changes. Likewise, if it would be more relevant for specific areas of the Northern Ireland economy.

- Investigate how much the boosted sample size is benefiting the results in collaboration with Southampton University's work using standard errors to optimise sample sizes per region.
- Compare methods we have used to what other countries are using for regional indices. Some countries, for example Germany start at a regional level and build up to an overall country index.

James McGregor
Prices Division, ONS
July, 2024

List of Annexes

Annex A	Raw Northern Ireland weights and experimental weights

Annex A: Raw Northern Ireland weights and experimental weights

The attached excel sheet shows the raw Northern Ireland weights in tab "NI raw weights".

Tab "Ave_wts_3" is the 3 year average of the weights centred around the date in column A. January and February weights are averaged separately. I.e. January 2015 average will be the average of January 2014, January 2015 and January 2016. The ave_wts_5 and ave_wts_7 similarly are the 5 and 7 year averages respectively.

In tab "imp_wts_3" the experimental weights are calculated. These are based on the January 2015 centred 3 year average and uprated using changes in the UK CPI since that date. Tabs "imp_wts_5" and "imp_wts_7" are similarly using the January 2015 centred 5 and 7 year averages and uprated using UK CPI changes since January 2015.

imputed_weights_cpi_201501.xlsx