ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH

Office for National Statistics

Improving the quality of regional economic indicators: Regional consumer prices

Kevin Connolly and Mairi Spowage

ESCoE Discussion Paper 2021-13

October 2021

ISSN 2515-4664

DISCUSSION PAPER

Improving the quality of regional economic indicators: Regional consumer prices
Kevin Connolly and Mairi Spowage
ESCoE Discussion Paper No. 2021-13
October 2021

Abstract

Building on previous work this paper provides updated prices and inflation estimates for the twelve NUTS-1 regions of the UK. One key issue previously noted when moving to regional prices estimates is the reduction in LCFS sample size leading to unstable weights. In this paper, we investigate the data sources to produce regional expenditure estimates, attempting to increase the accuracy of CPIH estimates. The primary focus is on the Regional Household Final Consumption Expenditure measure (HFCE) publication for use as estimates for regional expenditure weights. While we were able to update the UK regional CPIH estimates to 2020 using other data sources we still encounter similar problems to Dawber and Smith (2017), with the small sample size of the price databases being problematic for the calculation of CPIH indices.

Keywords: Regional Prices; CPIH and Inflation

JEL classification: C60, O11, R11

Kevin Connolly, University of Strathclyde, Fraser Allander Institute,
k.connolly@strath.ac.uk

This research has been funded by the ONS as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE)

Published by:
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
United Kingdom
www.escoe.ac.uk

UNIVERSITY of STRATHCLYDE
FRASER OF ALLANDER INSTITUTE

University of Strathclyde Business School

Improving the quality of regional economic indicators

Regional Consumer Prices

Kevin Connolly[1] & Mairi Spowage

University of Strathclyde – Fraser of Allander Institute

JEL codes: C60, O11, R11

ESCOE ECONOMIC STATISTICS CENTRE OF EXCELLENCE

[1] Corresponding author email: k.connolly@strath.ac.uk

Abstract

Building on previous work this paper provides updated prices and inflation estimates for the twelve NUTS-1 regions of the UK. One key issue previously noted when moving to regional prices estimates is the reduction in LCFS sample size leading to unstable weights. In this paper, we investigate the data sources to produce regional expenditure estimates, attempting to increase the accuracy of CPIH estimates. The primary focus is on the Regional Household Final Consumption Expenditure measure (HFCE) publication for use as estimates for regional expenditure weights. While we were able to update the UK regional CPIH estimates to 2020 using other data sources we still encounter similar problems to Dawber and Smith (2017), with the small sample size of the price databases being problematic for the calculation of CPIH indices.

1 Introduction

Since 2017, the Consumer Prices Index including owner occupiers' housing costs (CPIH), has been the lead indicator of inflation in the UK. First introduced in 2013, CPIH is a comprehensive measurement of inflation as occupiers housing costs (OOH) - costs associated with the owning, maintaining and living in a home (Dawber and Smith, 2017) - are included. This is important to inflation measures as OOH costs are a large proportion of overall household spending.

Similarly to the Consumer Prices Index (CPI) and Retail Prices Index (RPI), CPIH has generally only been published at the national level (all UK), however it would seem advantageous to be produce and publish estimates at a regional level. The objective of this paper is to extend the work of Dawber and Smith (2017) in developing regional CPIH estimates for the 12 NUTS-1 regions of the UK (Scotland, Wales, Northern Ireland and the nine regions of England).

The idea of using regional price/inflation indices in UK in addition to national ones is not new, with researchers suggesting that regional estimates are important to determine if there are differences across regions and to understand why. However, even though they would be seen as advantageous, under the current data collection system regional indices are seen as challenging with both the Living Cost and Food survey (LCFS) and the price quote database being too small on a regional scale. This could be overcome, at a cost, by increasing sample size of the price quote and LCFS. There is some data available on changes in regional prices from Relative Regional Consumer Price Levels publications (ONS, 2018), with the data coming from additional prices collected made for the Purchasing Power Parity database every six years. While these publications do give some indication of price changes over a six-year period they cannot be used for inflation measurements due to the differences in methodologies and weights.

The ONS does not currently have standard regular regional measures, there has been work in other countries. Nagayasu (2011) investigates the difference in region inflation in Japan, finding significant differences, which are in contrast with the conventional view that regions within a monetary union (like the UK) see similar inflation rates. In Duran (2016) the authors examine the regional inflation rates in Turkey over an 11-year period from 2004 to 2015. Like Nagayasu (2011) there are significant differences in regional inflations rates with the disparities declining over time, particularly after the financial crash of 2008.

The objective of this work is to provide updated regional CPIH estimates in a similar manner used for national values – following closely the methodology outlined in Dawber and Smith (2017). This work is extension of Dawber and Smith (2017) as we report CPIH rates up to the end of 2020 and investigate different data sources for regional expenditure weight calculations.

Following the national method for regional CPIH estimates, Dawber and Smith (2017) note several challenges including a small regional LCFS sample size, leading to unstable weights. The previous work used small area estimates to attempt to resolve the issue of unstable weights, but the results were inconclusive with the method (currently) not considered viable for regional CPIH. In this paper, we instead investigate the data sources to produce regional expenditure estimates, which may allow for the calculation of regional CPIH weights to increase accuracy of measurements. The primary focus is on the Regional Household Final Consumption Expenditure measure (HFCE) publication for use as estimates for regional expenditure weights.

The report is structured as follows. The next section details the method used for CPIH calculations where we first give an overview of the national method and then how this can be adapted for regional indices along with outlining the LCFS and HFCE databases. Section 3 describes the adjusted and unadjusted weights as well as the results for the regional CPIH and inflation simulations. In Section 4 we note other data sources which could be useful in developing regional household expenditure weights with Section 5 concluding.

While we are able to update the UK regional CPIH estimates to 2020 using other data sources we still encounter similar problems to Dawber and Smith (2017), with the small sample size of the price databases being problematic for the calculation of CPIH indices.

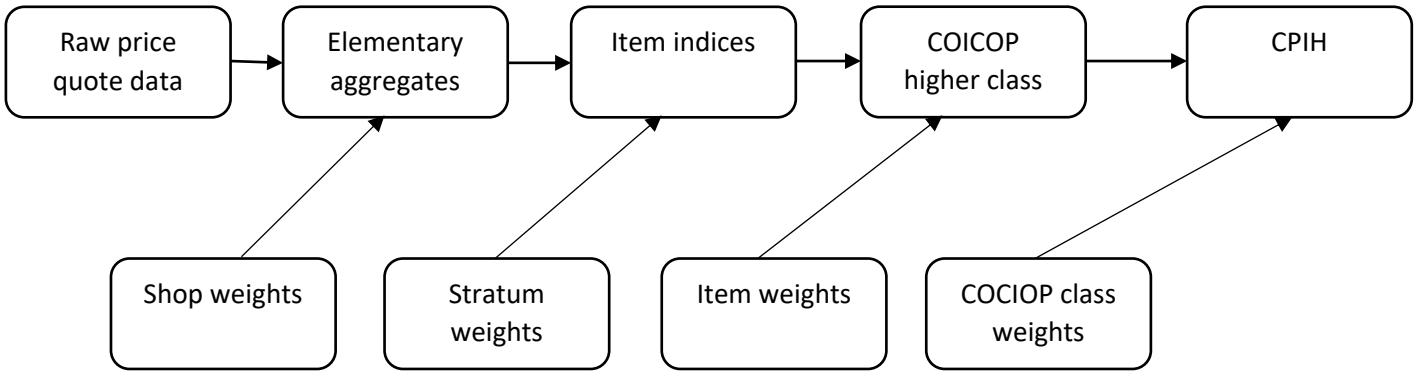
2 Methodology

Outlined in the introduction, the methodology for calculating regional CPIH follows closely, with some adaption) the national calculation, which we explore in detail first.

2.1 National CPIH

The consumer price index manual (ONS, 2019b) thoroughly describes the rationale and methodology of the CPIH calculation, while Dawber and Smith (2017) outline the practical steps – with both summarised in this section.

Figure 1: Schematic of CPIH national methodology.



Source: Dawber and Smith (2017)

Fundamentally, the calculation of CPIH involves aggregating the prices of items within a shopping basket (of goods and services), using a series of weights, to estimate inflation. Weights are used as not every item within a basket is consumed equally. As we spend more on some items than others, these items should have a greater influence on inflation rates. For example, you would expect a 10% increase in petrol to have a higher impact than that for the same relative price change in tea.

The first stage of development of national CPIH, outlined in Figure 1, is to produce elementary aggregates using the raw price quote data and stratum weights. These are the lowest aggregation of prices which cover all prices collected for one item in one stratum. An item within a stratum is identified by shop type – either multiple (10 or more outlets) or independent (fewer than 10 outlets) and by region. Most elementary aggregates are calculated as the weighted mean of relative prices within each stratum using equation (1).

$$I_{i,j} = \frac{\sum W_s * P_{i,j}}{\sum W_s} * 100 \quad (1)$$

Where $I_{i,j}$ is stratum index (elementary aggregate) for item i, stratum j. W_s is the shop weight with $P_{i,j}$ the relative log price (log of price divided by base price). These stratum indices are combined with stratum weights, using equation 2, to produce item indices.

$$I_j^k = \frac{\sum I_{i,j} * W_{i,j}}{\sum W_j} \quad (2)$$

With I_t^k the index of item k and W_{ij} being the stratum weights. Using item weights (w_k), class indices (I_t^c)are calculated.

$$I_t^c = \frac{\sum I_t^k * W_k}{\sum W_k} \quad (3)$$

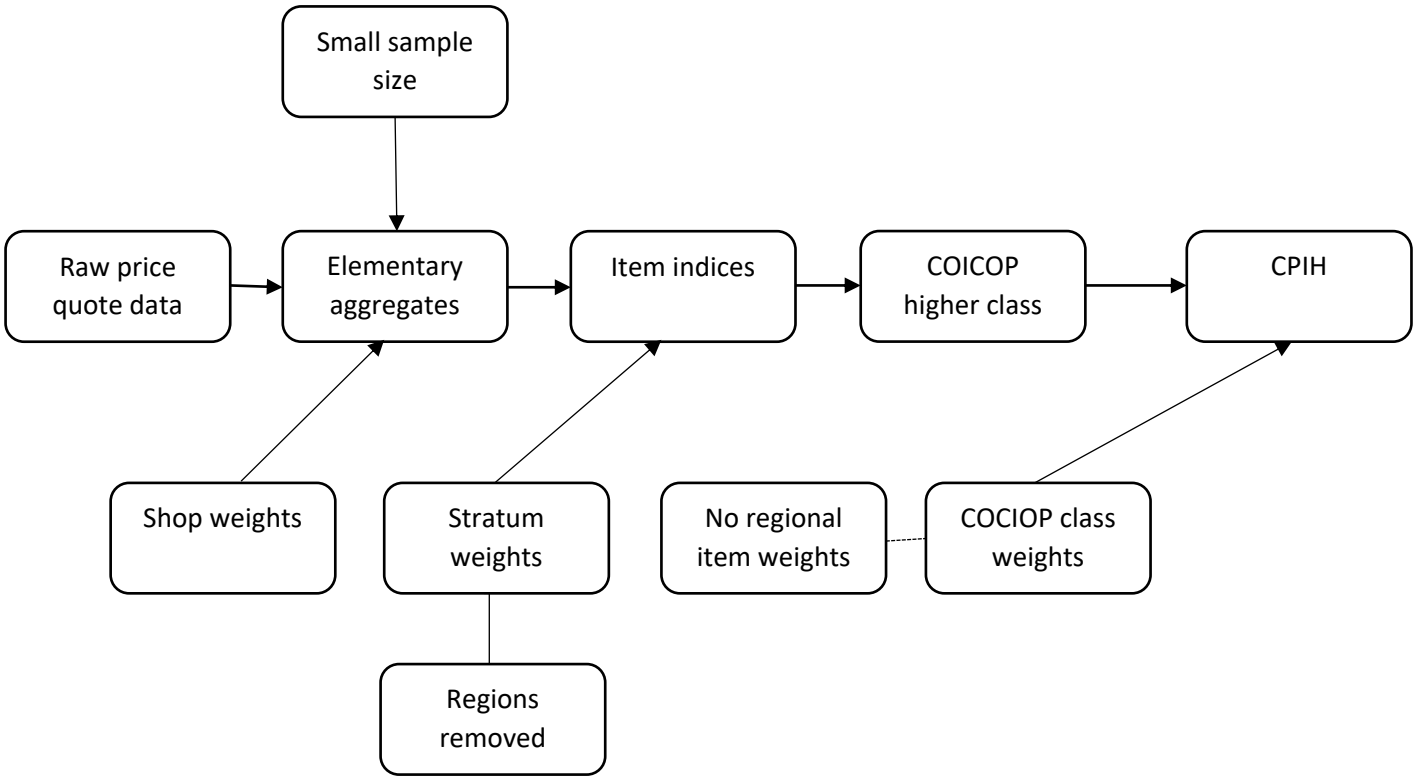
Equation 3 is applied for all COICOP classifications levels from COIOCP level 3 to COICOP level 1 creating the headline CPIH value. Above item aggregation requires COICOP classes to be matched with item identifications.

Several data sources are used for the national CPIH calculation. For elementary and item level indices, ONS provides the price quote and item indices databases available from the SRS. The price quote database, generated monthly, contains the prices of a fixed basket of goods and services throughout the UK. Collection of prices involves both physically attended samples as well as internet collection. Prices are presented in both pounds and relative log price (needed for equation 1) and include a range of identification categories such as region, shop type and stratum. The item indices database, again generated monthly, contains the required weights for equations 1 and 2. Above item index aggregation (equation 3) requires weights derived from national accounts expenditure, the Living Cost and Foods Survey, market research as well as other data sources.

2.2 Regional CPIH

Detailed in the previous section is the method for producing national CPIH, which we adapt for regional estimations. Dawber and Smith (2017) outline the changes in the methodology shown in Figure 2.

Figure 2: Schematic of CPIH methodology for regions



Source: Dawber and Smith (2017)

The previous work (Dawber and Smith, 2017) identified challenges/limitations when applying the national CPIH methodology for regions. Firstly, the authors note that there is the problem with small sample sizes in the price quote dataset. For CPIH at the national level the sample size is suitably large with gives an accurate estimate of the price changes for a range of items. However, at a regional level this may not be the case as the price-quote data for each region will be approximately a 12^{th} the size of the national database. Reducing the database scale means under representation of some items, potentially leading to inaccurate estimates of price changes at the regional level. To increase regional accuracy the sample size could be increased, but this will come at significant cost.

Secondly, ONS (Dawber and Smith, 2017) note the use of improper Stratum weights. Stratum weights within the national CPIH calculation account for the different pattern of spending across regions and shop types. However, within a regional CPIH context there is no adjustment for regional differences needed thus the stratum weights reduce to shop-type weights only.

Finally, the authors identify the COICOP (and higher level) weights are also problematic for the production of regional indices as they rely on national accounts household final consumption expenditure (HFCE) data, and are therefore subject to revision over time. These

weights are the focus of this paper as a combination of LCFS and regional HFCE data in their calculation. An extension of Dawber and Smith (2017) was provided in Dawber et al (2019) in which the authors use small area estimations to try to stabilise the household expenditure weights. In short, this involved using the relationship between key descriptor variables and expenditure on items (the most commonly used being household income, for example) to predict expenditure for some items within some regions. Ultimately, there are limitations to the approach given the limited number of regions.

Early in the process of this project we had discussions with the prices team in ONS to get their views on the next steps for the research. Their view was that much of the volatility in the weights that was found by Dawber and Smith (2017) could be explained by the source data used. Dawber and Smith (2017) had used the LCFS only, and the volatility in the data at NUTS 1 level for some COICOP classes contributed to the volatility in the weights.

In this paper we therefore take a different approach by investigating the base LCFS data available from the SRS further, looking at unweighted estimates and using the regional HFCE for weighted estimates.

2.3 The Living Costs and Food Survey (LCFS)

The fundamental source of data used in the calculation of the expenditure weights for regional CPIH was the Living Cost and Food survey (LCFS), with data obtained for 2008-2018. Previous regional CPIH work used LCFS data available from the UK data services that includes households and individual data level files. For this paper, we were given access to all underlying weights data through the ONS secure research service[2].

Started in 1957, and conducted every year, the LCFS provides detailed expenditure information for each of 6000 surveyed households randomly selected from across the 12 NUTS-1 regions of the UK. For each household an initial survey interview is carried out, with data collection occurring over a two-week period. During this two-week period each individual over the age of 16 is given a diary and asked to record all their purchases. The diary consists of 10 sections with the first six related to everyday purchases (ranging from food and drink to lottery tickets) and the other four covering the entire two-week period (holiday expenditure, special occasion purchase)[3]. Further information on the LCFS can be found in ONS (2019). These purchases are then classified using the COICOP hierarchy, matching with the information found in the item and price quote data sets.

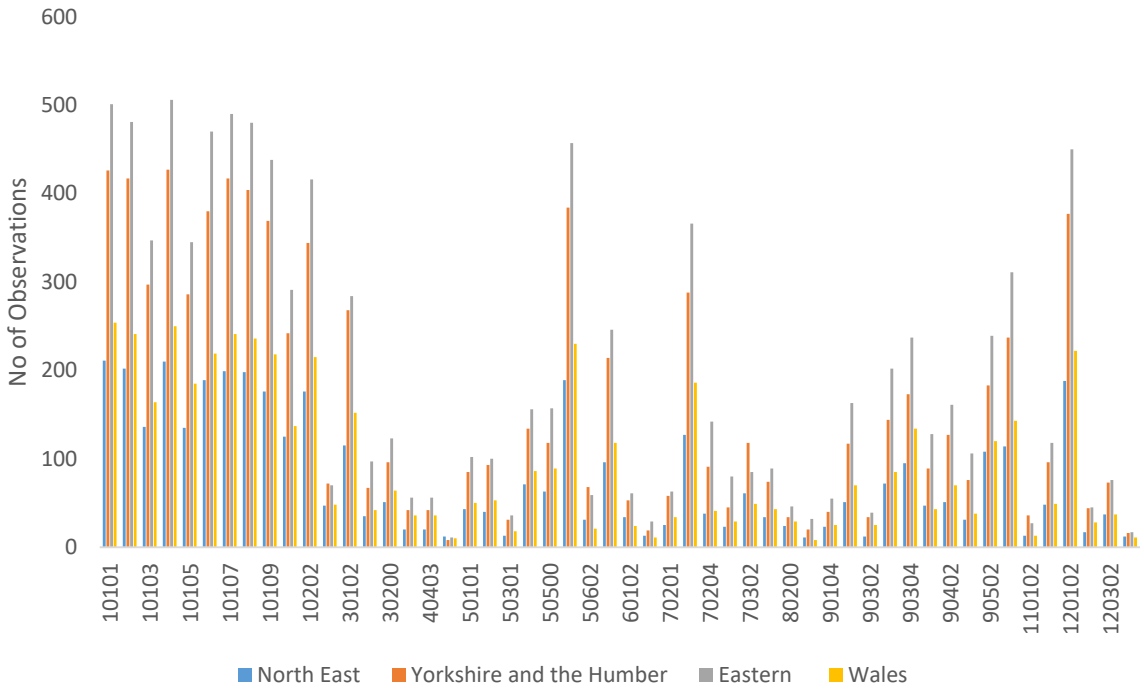
Ultimately the information from the LCFS is used to estimate a mean normalised expenditure profile by COICOP class for each of the 12 NUTS-1 regions, used as the initial weights within the CPIH calculation (input as parts per thousand (ppt)). A variety of methods can be used to

[2] See https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme
[3] Information on some larger purchases, like vehicles, is obtained retrospectively through the interviews

calculate these weights, for this work both unweighted and weighted profiles were estimated – which we detail in the following sections.

In the LCFS, households are given both a case number and area identifier, which enables separation into the 12 NUTS-1 region. Included in the LCFS data is the purchased item (by COICOP plus level) and cost for each purchase within the household. This means several entries are input for some items, particularly everyday items like food and drink. From Figure 2 and 3 we find that the number of observations varies greatly by COICOP class and regions.

Figure 3: Observations by region, 2017[4]



Source: LCFS database

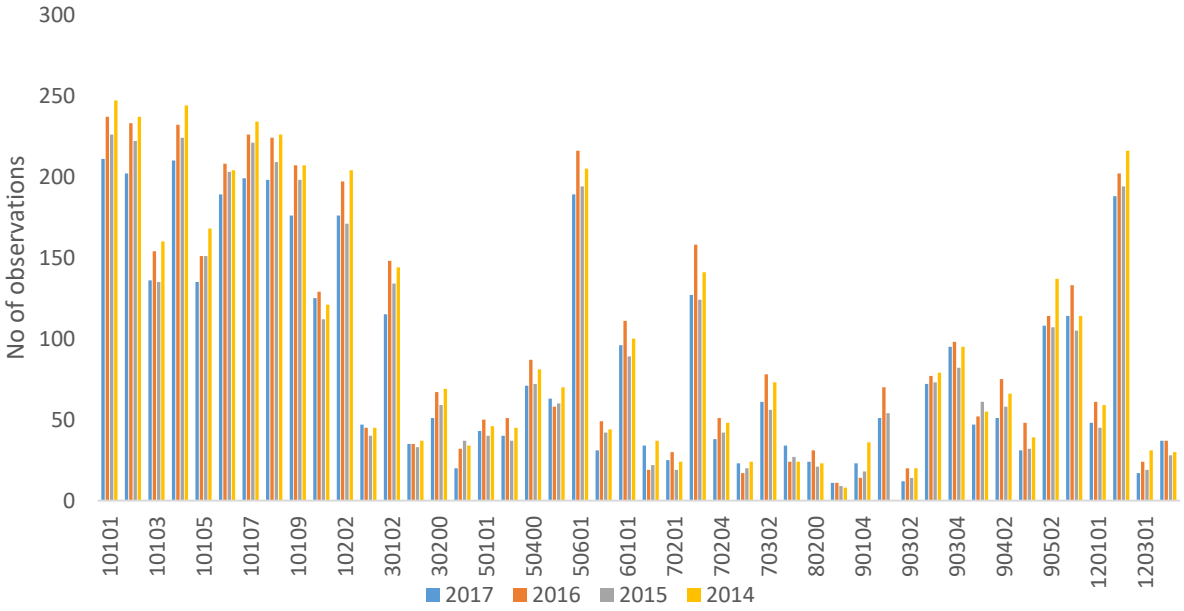
Each observation identifies if a member of the household purchases a good or service within the COICOP class but not the number of purchases within each COICOP – these are accounted for in in the CPIH calculation. The above figure demonstrates the large variation between region and COICOP classes. Classes 10101 (Cereals) to 10202 (Coffee and substitutes) represent everyday purchases (food and drink) thus these have the largest number of observations for each of the four identified regions. In general, for the higher COICOP classifications the number of observations drop significantly as these are not everyday goods and services. There are however some exceptions which still have large observation numbers such as 50601 (non-durable household goods), 70202 (fuel) and 120102 (insurance).

Similar to COIOCP classifications, the numbers of observations in the LCFS data also vary depending on the region. In the first COICOP (10101) classification we find that the North

[4] Due to data sensitivity some COICOP levels have been omitted.

East only has 211 observations compared with the Eastern region which has 501. Ideally, we would have the same number of observations for each region and COICOP but different spending patterns and response rates make this near impossible. The LCFS has been designed in such a way as to maximise the observations. While the difference in observation numbers may not be significant for COICOP levels with large numbers of observations, there will be significant impact with lower number COICOP classifications like COCIOP 50601 (Non-durable household goods) where the North East has 189 observations compared with the 457 for the Eastern region.

Figure 4: Observations by year, North East region[5]



Source: LCFS database

Figure 4 shows the number of observations by COICOP for the North East region for the four-year period between 2014 and 2017. Overall, as would be expected, the pattern is similar to that seen in Figure 3 i.e more observations in the everyday COICOP classification levels (for example, food & drink etc.) compared with others. We also find that the variation in observation numbers between years is much more stable when compared with the differences between regions (Figure 3). While these may be more stable there are still yearly differences in observations. For the everyday items there is, on average, a 20% difference between the maximum and minimum observations whereas this difference increases with the higher level COICOP classifications due to the smaller number of observations.

Both Figure 3 and 4 show that the number of observations vary depending on the COICOP classification, region and year. The difference in COICOP classification observations is unavoidable as some goods and services are purchased more frequently than others, indeed

[5] Due to data sensitivity some COICOP levels have been omitted.

this is why we are carrying out this work to generate base weights. The yearly differences are also unavoidable as these are mainly caused by differences in response rates.

2.4. Household Final Consumption Expenditure (HFCE)

Published for the first in 2018, the ONS Regional Household Final Consumption Expenditure measure (HFCE) publication (ONS, 2018) gives experimental regional household consumption estimates across the 12 NUTS-1 regions of the UK at COICOP group level. The increased devolution of powers to local and combined authorities within the UK has led to the need for statistics to monitor and inform policy at regional level – the key driver for developing the HFCE estimates. Two HFCE estimates are made, national and domestic, with latter being the key estimated for rCPIH.

Previously the only household information at the regional level was the gross disposable household income (GDHI), which measures the total amount of money households are able to spend or saving by measuring income (wages, property) and outgoings. The HFCE extends this by estimating the goods and services that households are purchasing. These HFCE estimates were developed to meet a range of user needs including:

- Planning facilities and infrastructure by local government;
- Aid in planning of goods and services investments;
- Expanding household accounts at regional level; and
- Making the UK compliant with the European Union's European System of Accounts 2010: ESA 2010 regional transmission tables.

These estimates also have the potential to inform fuller regional accounting approaches, with HFCE being one of the many challenges in the estimation of regional Supply and Use and Input-Output tables.

It is worth saying that these estimates are currently experimental statistics and are at an early stage in their development, and are not necessarily at the moment consistent with HFCE estimates produced by parts of the UK (e.g. estimates for Scotland which are produced by the Scottish Government), some of which are established National Statistics products. It is likely that these issues will be explored in the production of the next edition of these statistics.

The main data sources used for the HFCE estimates are the LCFS and the Annual Business Survey (ABS). Estimates were produced for both the domestic (spending in a region, no matter by whom) and national (spending by the residents of a region, no matter where) concepts, in order to estimate net flows in and out of regions of particular goods and services. In this way, the estimates can also shed an interesting light on the flows of goods and services between regions, which can again be informative for regional accounting.

Other data sources are also utilised for other goods and services, where available:

- Department of Business, Energy and Industrial Strategy (BEIS) data on the consumption of fuel;
- Data from local authorities on parking charges; and
- Bank of England regional measures of fees and Financial Intermediary Services Indirectly Measured (FISIM);

However, despite the use of these other sources, the data for measuring the national concept is still overwhelmingly dominated by the LCFS. In order to deal with the volatility present, ONS statisticians have using expert judgement, as happens for many statistics, adjusted the data to smooth it out.

Helpfully, the team in ONS have provided detailed information on the adjustments that are made to the raw data. These adjustments are manual and by eye, based on the judgement of statisticians. The table overleaf shows the largest absolute adjustments that are made to the raw LCFS data.

These demonstrate the items that require the largest adjustment, which include mostly infrequently bought goods (such as holidays, furniture and motor cars) or those which cover expenditure by a limited amount of households (such as payments for Education). As these are absolute values, London also dominates.

It is worth flagging that the purpose of the adjustments is to smooth out the figures over time. They therefore vary in direction and size: some examples as shown overleaf: the blue line being the raw data, and the orange the adjusted figure used.

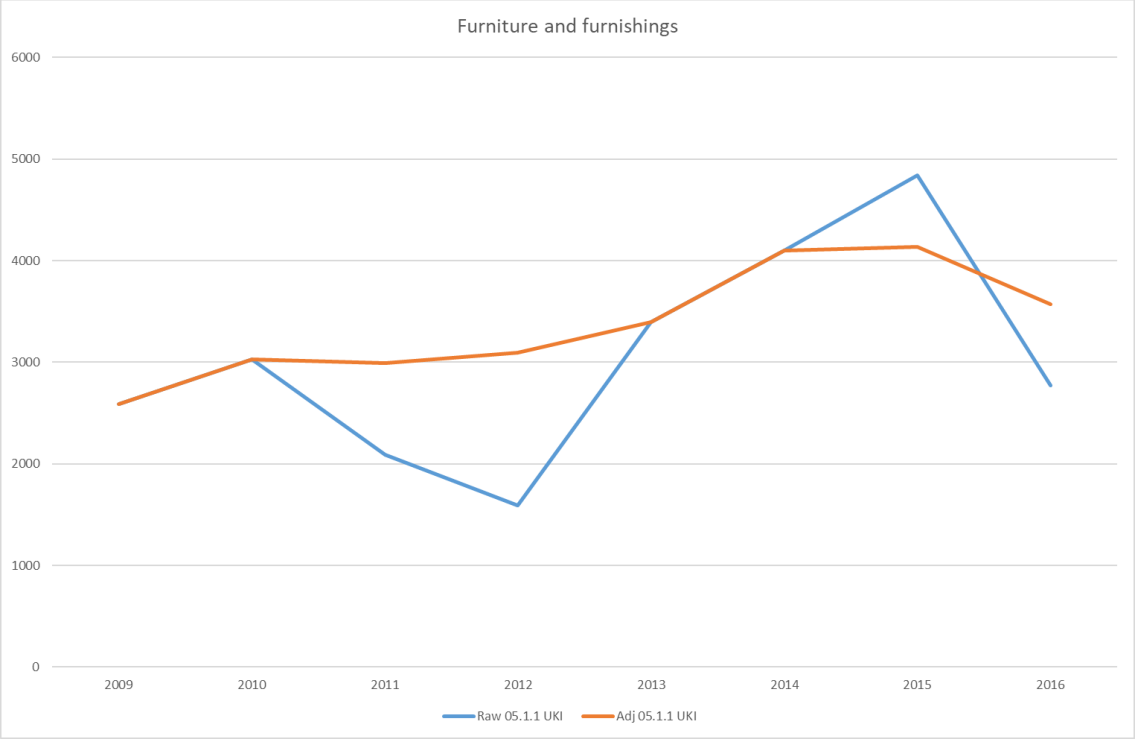
Table 1: Total Absolute adjustments, 2009 to 2016

COICOP	COICOP label	Region	Absolute Adjustment	Change over 5 years as % of total spending
10	Education	East Midlands	4350	98.6%
05.1.1	Furniture and furnishings	London	3900	14.5%
07.1.1	Motor cars	North West	3400	14.4%
10	Education	Wales	3400	161.6%
04.1.1	Actual rents paid by tenants for housing	London	3200	3.6%
09.6	Package holidays (UK and abroad)	London	2900	11.2%
07.3.3	Air	London	2680	22.2%
07.1.1	Motor cars	Yorkshire and the Humber	2650	15.7%
10	Education	South East	2650	21.2%
09.6	Package holidays (UK and abroad)	East of England	2600	11.9%
09.4.1	Recreational and sporting services	London	2450	19.6%
07.1.1	Motor cars	East Midlands	2350	12.3%
09.6	Package holidays (UK and abroad)	North West	2350	9.8%
10	Education	London	2300	11.4%
09.4.3	Games of chance	South East	2110	59.0%
10	Education	East of England	2100	33.2%
09.6	Package holidays (UK and abroad)	South West	1950	10.8%
09.6	Package holidays (UK and abroad)	Scotland	1800	10.0%
11.1.1	Restaurants, cafes etc.	East of England	1800	5.7%
11.2	Accommodation services (UK and abroad)	London	1800	12.8%
12.4	Social protection	London	1760	20.7%
03.1.2	Garments	South East	1750	7.0%
09.4.1	Recreational and sporting services	South East	1750	15.1%
10	Education	Scotland	1750	37.1%
07.1.1	Motor cars	Scotland	1600	7.2%
09.2.1	Major durables for outdoor recreation	London	1600	95.9%
04.1.1	Actual rents paid by tenants for housing	Yorkshire and the Humber	1550	7.9%
05.1.1	Furniture and furnishings	Wales	1550	24.5%
10	Education	South West	1500	20.1%
11.1.1	Restaurants, cafes etc.	London	1500	2.6%

Source: ONS

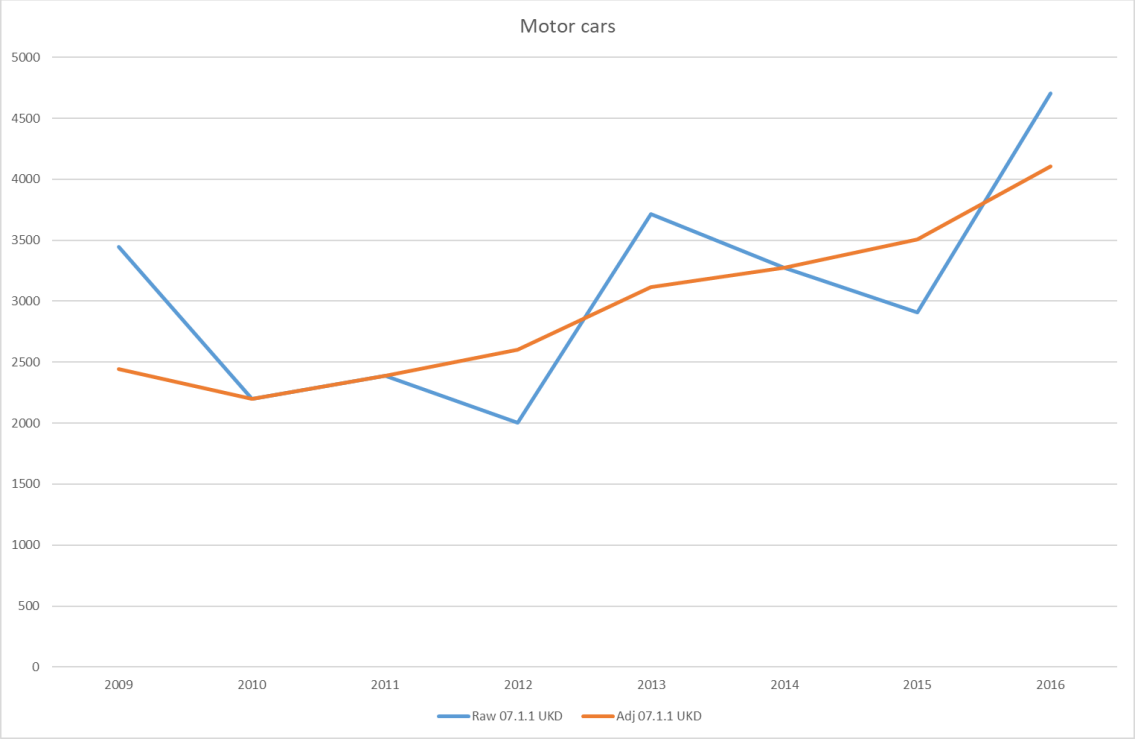
The updated publication was published in July 2020 which we use to develop adjusted expenditure weights for regional CPIH estimation. Section 3.2 gives detail on how the HFCE can be used in the calculation of expenditure weights.

Figure 5: Furniture & furnishings expenditure in London



Source: ONS

Figure 6: Motor Cars expenditure in North West



Source: ONS

3 Calculation of base weights

3.1 Unadjusted weight

The main objective of this paper is to use the fundamental LCFS data to calculate the base expenditure household weights used in the calculation of regional CPIH. We focus on two types of weights, adjusted and unadjusted, for which we use the direct estimate method outlined in Wurz (2017).

Unadjusted weights are the simplest as we are only interested in the amount paid by each household. Within the LCFS database each purchase by each individual in the household is identified as a separate data point thus the first stage is to aggregate across the household to give the total spend for a specific good or service. Secondly, as the fundamental data is in COICOP plus level we aggregate the data from COICOP level 5 to level 3 which match with the information in the price quote datasets.

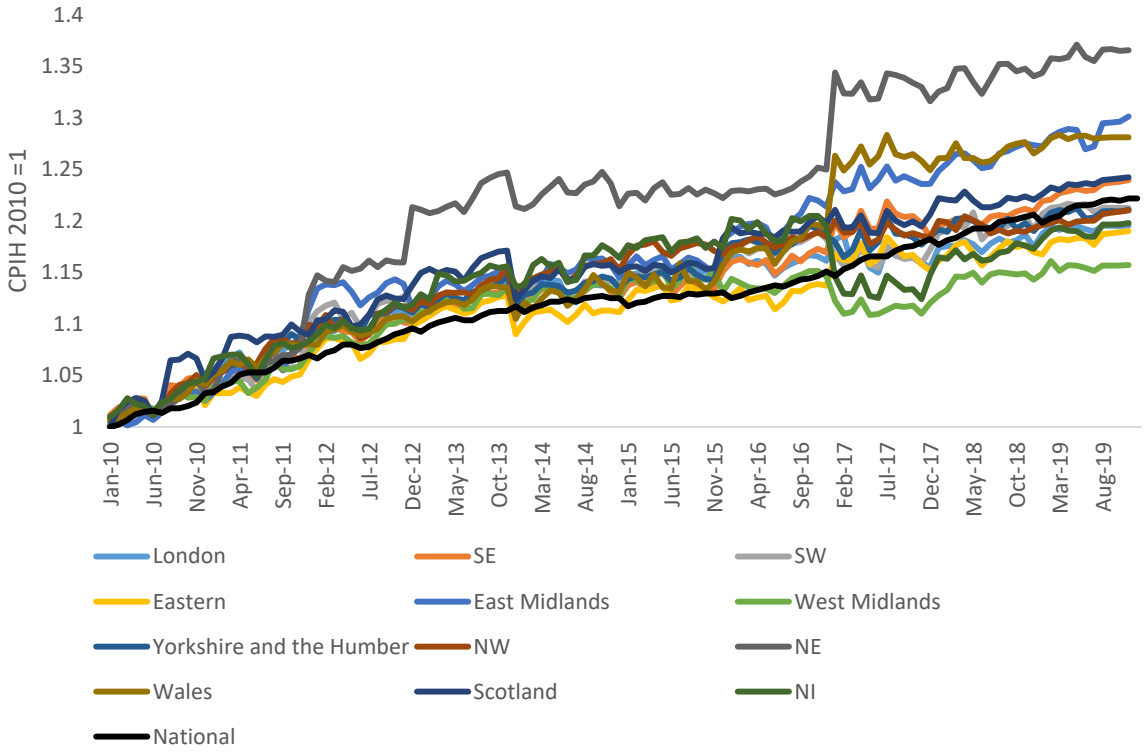
This aggregation gives the total spend by COICOP level 3, by all households within each of the 12 NUTS-1 regions in the UK. The input weights to the CPIH calculated are based on average regional baskets thus we must normalise, to 1000, the aggregated totals previously calculated using equation (4):

$$N(x_i) = \frac{x_i}{\sum x_i} * 1000 \quad (4)$$

N(x_i) is the normalised value[6], in parts per thousand (ppt), of x_i compared with the total spend within the region. Higher values give more weight to those COICOP classifications in the regional CPIH estimates. In the calculation it must be noted that the CPIH weights are based on LCFS data from two years previous (i.e the 2019 weights are based on the 2017 LCFS). Similar to the observations, CPIH weights are dependent on year, classification and region.

[6] i are the different COICOP classifications spend

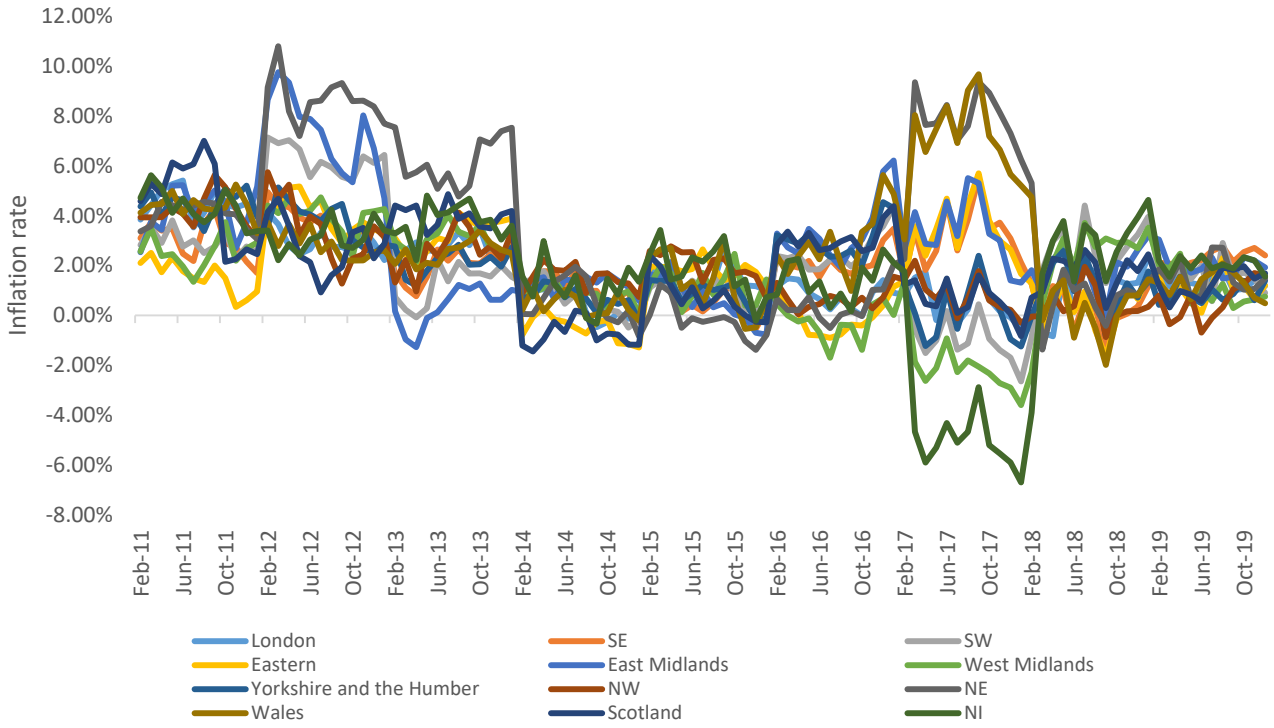
Figure 7. Regional CPIH estimates using adjusted expenditure weights, 2010 -2019



Source: author's calculation

From the figure above, prices across the 12 UK NUTS-1 regions follow a similar pattern with there being an increase from 2010 to 2019 and the North East CPIH being the highest and Eastern the lowest difference. In fact, the North East CPIH is significantly greater than all other regions throughout the time series. Figure 7 does also demonstrate that using unadjusted weights can cause unstable regional CPIH estimates. Over the 10-year period there is overall an increase in CPIH across all regions however this is not constant, with there being several noticeable changes particularly around 2017. By comparing all regions to the UK estimate at all times most regions have a higher value than the national average again suggesting problems with using unadjusted weights.

Figure 8. Regional inflation estimates using adjusted expenditure weights, 2010 -2019



Source: author's calculation

Figure eight illustrate that the use of unadjusted weights are problematic for regional CPIH calculations as they produce unstable estimates with unrealistic inflation rates. To try and improve these estimates we use other information within the LCFS to develop adjusted weights.

3.2 Adjusted weight

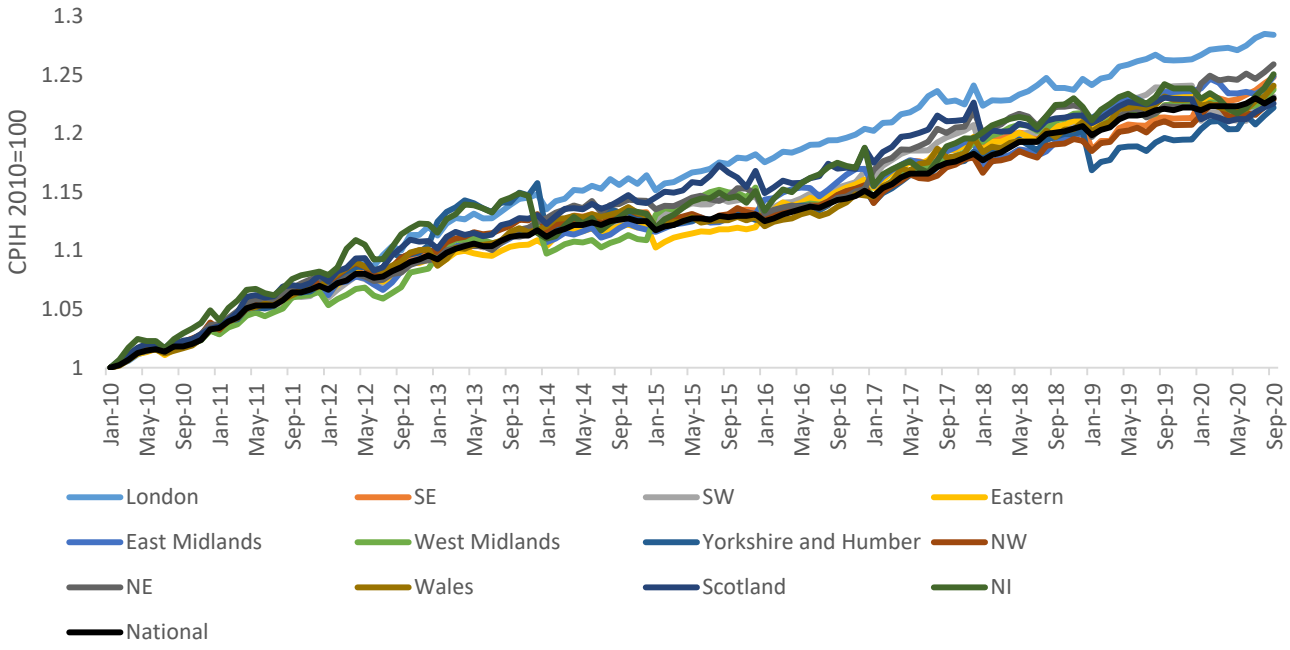
In the calculation of unadjusted weights, the focus was solely on the net amount paid by household by COICOP class, not accounting for non-response bias and non-private households (which are not in the LCFS). To overcome these problems adjusted weights can be used. The adjusted weights for this report utilize the HFCE publication as the base input, with some adjustments from the primary LCFS data.

The first step in the process is calculation of the proportion of total expenditure by COICOP classification from the HFCE for each of the 12 regions. There are several COICOP classifications- such as 2.1 - which have been published at level two for the HFCE but are identifiable at COICOP level three using the LCFS data. In these cases, the HFCE is supplemented by the LCFS to separate the level two classification into two (or more) level three classification. This is achieved through the following equation.

$$HFCE_3 = HFCE_2 * \frac{LCFS_3}{LCFS_2} (5)$$

Where $HFCE_3$ is the level three classification expenditure from the HFCE, $HFCE_2$ the level two classifications and $LCFS_2$ and $LCFS_3$ being the level 2 and 3 expenditure from the LFCS. Separating the levels in this way means there is the same number of expenditure weights as in the unweighted analysis (but the values will be different). Using these adjusted weights we calculated the CPIH in using the previously outlined method, the results are found in Figure 9.

Figure 9. Regional CPIH estimates using adjusted expenditure weights, 2010 -2020



Above shows that the regional CPIH estimates for all 12 regions between 2010-2020 follow a similar pattern over the years but there are some fundamental differences. For comparison, the national CPIH index has also been included in the series.

Through the first 3 years the indices follow a similar pattern but by 2014 the difference overall becomes apparent. From 2014 the regional CPIH for London become noticeably larger than the other 11 regions, with the CPIH by the end 2019 being around 23% larger than 2010. Also around 2014 the regional CPIH estimates for the North West and Scotland become significantly lower than the national average.

One noticeable issue with the regional CPIH, particular from 2014 onward is the sizable difference in some indices between December and January (Scotland in in 2020) for example[7].

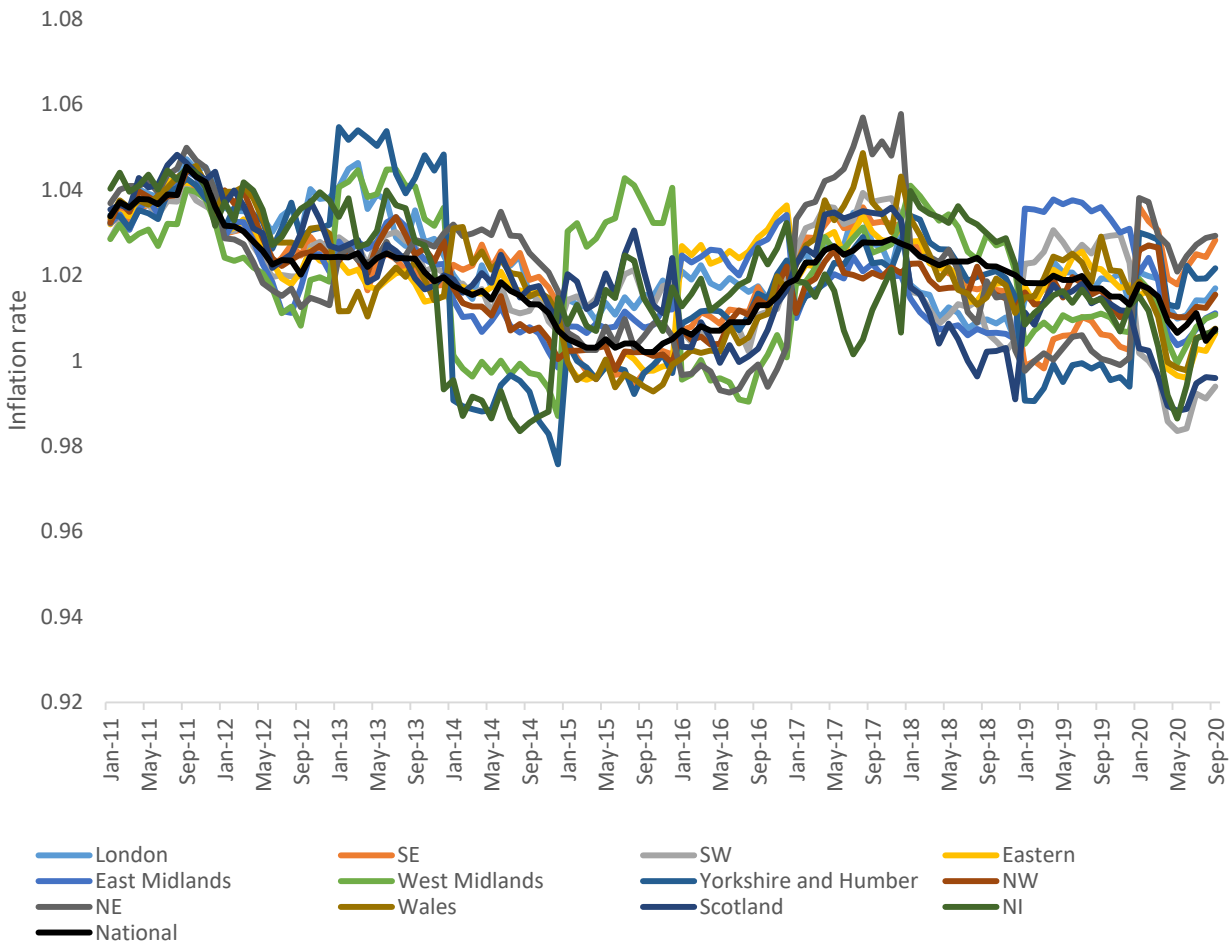
This is assumed to be caused by the chaining of indices. These 'jumps' occur in the national CPIH index but are much more subdued than the regional ones indicating that the issue may

[7] This also occurs for unadjusted weight

be in the price/item quote sampling size. With the regionalisation on national data on average, the sample for each item is a 12th that used in the national calculation.

Over the 10 year period there were some large reductions in sample size for certain price quotes, for example the sample size for 90104 (recreational services) reduced by 45%, which is more problematic when moving to regional analysis. Figure 10 gives the yearly inflation rates using the HFCE adjusted weights.

Figure 10: Regional inflation estimates using adjusted expenditure weights, 2010 -2020



From the above figure we find that similar to the index, the fluctuations in inflation rate are much more prominent for regions than the national estimations. Similar to the CPIH indices we find that the fluctuations are enlarged between December and January again suggesting that the chaining of indices is much more problematic for regions due to the small sample size in parts of the price quote database.

4 Improving the regional CPIH estimates

This paper has demonstrated that while the HFCE data can be used (in conjunction with the LCFS) to produce timely rCPIH estimates, there are still some issues with the estimates.

4.1 The Impact of Price Quote Sample Size

One of the main problems identified with rCPIH estimates is the sample size of the price quote database when moving towards regional estimates. The current design of the price quote dataset is sufficient for national indices as many observations are taken for each COICOP classification, which are used in the chain-liking of indices. However, as there are 12 regions of the UK, when using this dataset for rCPIH there are fluctuations in the number of observations, which feeds into the linking of indices between years.

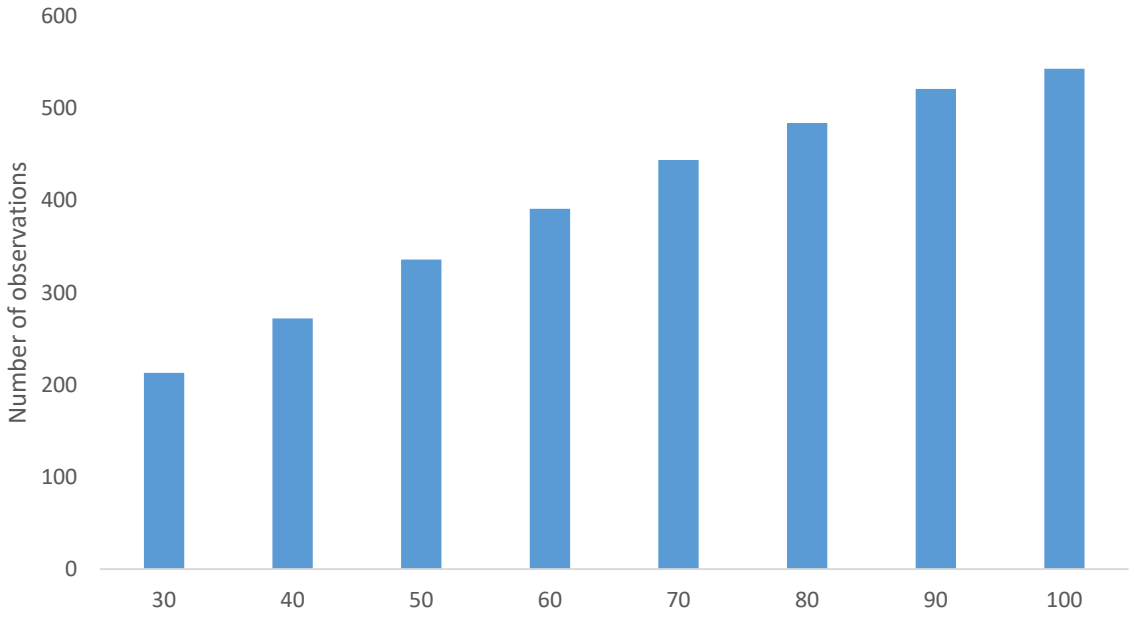
Overall, in the past 10 years, the number of observations within the price quote database has increased by 9% with classification 90105 (cultural goods) having the lowest number of average national monthly observations (92) in 2020. When moving towards price quote data for regional estimates, the average number of monthly observations significantly reduces, illustrated in Table 2.

Table 2. Average number of monthly observations per COICOP classification by region, 2020

Region	Average
NE	184
NW	209
Yorkshire and the Humber	134
East mid	139
West mid	114
Eastern	123
London	132
SE	167
SW	77
Wales	78
Scotland	133
NI	61

The North West has the largest number of average observations at 209 per month, more than three times than of Northern Ireland, which has an average of 61 observations by COICOP classification. In general, for many classifications, moving towards a regional price quota database is unproblematic as there are still a significant number of observations. However, for some regional classifications, there is a considerable reduction in the number of observations, impacting the chain-linking process. Figure 11 illustrates the number of regional COICOP classifications under a certain observation level.

Figure 11: Number of average monthly observations under a defined limit



Out of a possible 792 classifications (66 COICOPs for each of the 12 regions) 213 (27%) have, on average, less than 30 observations per month with nearly 70% (543) having less than 100 observations. Again, as with national estimates, the lower number of observations are for COICOP 90105 (cultural goods) classification.

There are a number of options here to deal with issue of small classification sample size when regionalising the price quote database. The first is to investigate options for increasing the sample size of these classes across the regions in the price quote collection and the second, investigate options for alternative data sources for price quotes. Section 4.2 details the latter option. If we increased the sample size to contain the same level data as the national CPIH estimate the collection would increase by a factor of 12, which is unfeasible due to the associated cost, instead the increase could be varied across regions.

Statistical power calculations can be used to identify the levels below which variability is determined to be too great: although of course it is a subjective decision to determine what is “too great”. In practice, many statistical products suppress values where the cell count is below an absolute threshold: thresholds of 5, 10, 25, or 100 observations are used in different official statistics products. The minimum of 5 or 10 observations in cell counts are common minimums used in statistical disclosure control.

Initially, a minimum number of observations of 10 per COICOP classification per region is identified. To achieve this goal, using the most recent year of data, there must be an increase in sample size across 11 of the 12 regions. To achieve this 10 minimum standard the aggregate price quote database size would need to increase by 62%. If the minimum target was set at five observations the database would only increase by 10% whereas a 215% increase is required for a 20 observation minimum. In addition, as outline previously, this is not a universal increase across all regions, rather a targeted approach. Table 3 gives

estimates on the increase in price collections needed across each of the regions individually to meet the 10 observation minimum across all COICOP classifications. This increase in sample size will also reduce the variance throughout the sample. For example, for Northern Ireland COICOP 90504 (Hire, maintenance and repair of major durables for recreation) the increase in sample size reduces the variance by 20%.

Table 3. Estimated increase in price quote collection required to achieve 10 observations per COICOP classification per region

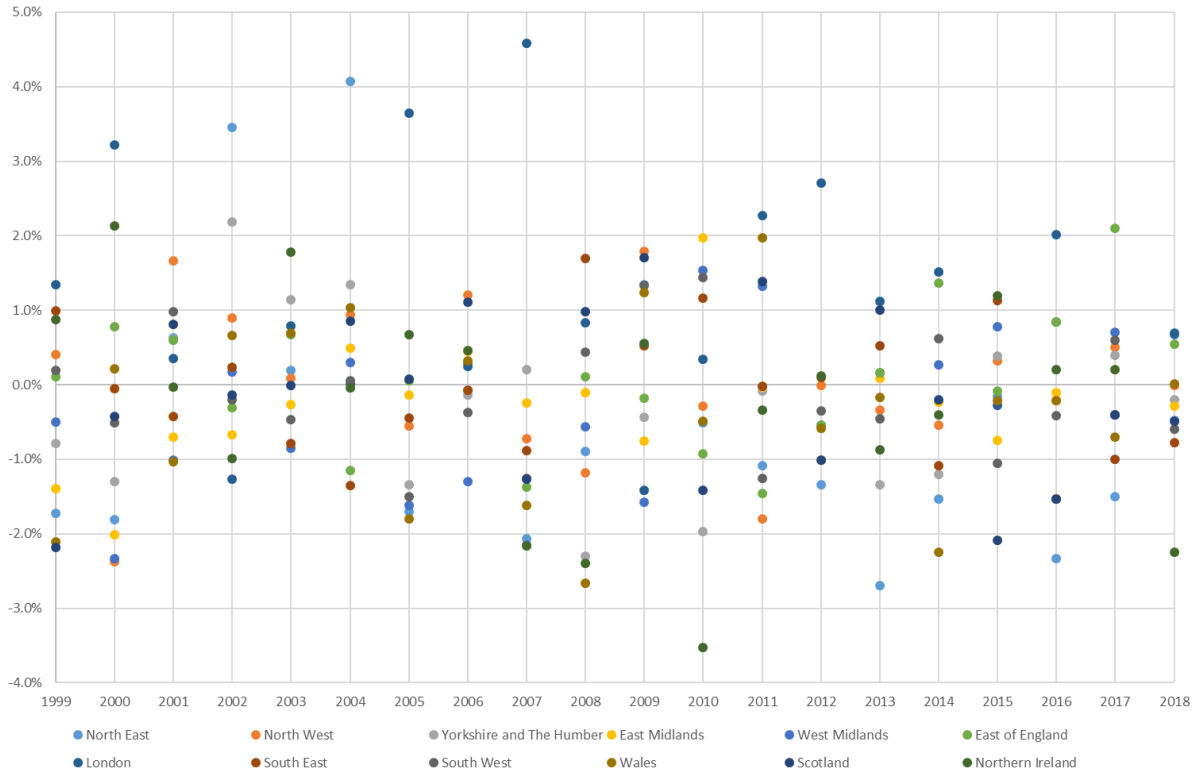
Region	Average
NE	2.56%
NW	0.00%
Yorkshire and the Humber	46.34%
East mid	44.58%
West mid	71.43%
Eastern	34.83%
London	44.58%
SE	12.15%
SW	233.33%
Wales	155.32%
Scotland	44.58%
NI	421.74%

4.2 Considering Quality Estimates

In producing these estimates, the statistical team involved will have to consider when estimates look like they are too volatile to publish, and whether interventions are required to smooth out these estimates. We have flagged above the issues with the December to January transition: and discussed the adjustments that are made to the weights information in the production of the regional household expenditure estimates.

Further adjustments may be required to produce a set of estimates that look sensible. However, this is a subjective judgement. An example is the regional GDP balanced figures (chain volume estimates) that are produced annually by the ONS. The chart below shows the differential between growth in each region and the national estimates in each year.

Figure 12: Growth differential for each region vs the national estimates, 1999 - 2016



We can see that there is a huge variation in growth estimates across regions, with some of the smaller areas of the country, such as Northern Ireland, having particularly volatile growth figures.

We have recommended that additional adjustments may be required if there are particularly influential price quote changes between years. However, once the adjustments to the weights through the HFCE data are made, alongside these additional adjustments, it is likely that the regional CPIH will not be more variable across regions than the data published above.

We do not recommend that formal thresholds should be adopted above which the data would be considered to be not good enough to publish. This would not be consistent with other practice on producing regional estimates. It is for the judgement of the professional statistical staff if the estimates are of sufficient quality to publish.

4.3 Additional Data Sources- VAT Assignment

As part of this project, we have considered if there are alternative data sources that can be used to improve the estimates, with a focus on data to improve the weights. Obviously we have used the HFCE data extensively in the production of this new set of estimates.

One other possibility was the VAT Assignment Model which has been developed jointly by HMRC and the Scottish Government since 2016.

In 2015 the Smith Commission convened to examine which further powers could be devolved to the Scottish Parliament, recommended that "the receipts raised in Scotland by the first 10 percentage points of the standard rate of Value Added Tax (VAT), and the first 2.5 percentage points of the reduced rate of VAT, will be assigned to the Scottish Government's budget." Following the Smith Commission's recommendations, the UK and Scottish Governments subsequently agreed in the Scottish Government's Fiscal Framework that a VAT assignment methodology would be jointly developed by UK and Scottish Government officials, to calculate the Scottish share of UK VAT receipts

This model has now been developed, an initial description and two sets of estimates published (HM Treasury, 2018, 2019, 2020). Given that the purpose of this model is essentially to regionalise VAT-able expenditure, many of the data sources that are used in the production of Consumer Trends at the UK level are used here. As we would expect, there is a large reliance on the Living Costs and Food Survey.

Like the ONS publication, the volatility for infrequently bought goods is dealt with by smoothing: in this case, by the use of 3 year averages. A particular example given is expenditure on Jewellery and Watches. A full list of the categories that have been smoothed out in this way is not currently published.

HMRC have chosen to employ a much wider range of data sources, some of which are HMRC information that may not be readily accessible to ONS. They have shared the full list of sources with us in confidence, but we have not received permission thus far to share this with the ONS. Their most recent publication in November 2020 sets out that "A detailed methodology paper will be published at a future date": so hopefully this should set out the details we have been provided with.

However, the UK and Scottish Governments have now agreed that VAT Assignment will be reviewed as part of the Fiscal Framework Review which is due to kick off in 2021. This is following concerns about the volatility of the estimates produced from the model: the Scottish Government and Parliament were concerned that this would introduce unreasonable volatility to the Scottish Budget.

This is a demonstration, if one were needed, of the continuing challenges in estimating household expenditure at the regional level in the UK for any meaningful policy making purpose.

This is likely to delay or limit the production of more detail on the model, and limit development. Therefore overall this may not be a fruitful avenue for improving the regional CPIH estimates.

5 Conclusions

In this paper we use the methodology outlined in Dawber and Smith (2017) to update the CPIH estimates of Consumer Price Indices including owner occupiers housing cost (CPIH) for the 12 NUTS-1 region regions of the UK. Previous work focuses on the price quote strata and small area estimates whereas in this paper we investigate using a different set of source data from which to calculate the expenditure weights.

With access to the LCFS database two types of weights – adjusted and unadjusted – were developed with regional CPIH and inflation estimates given for both. Similar to previous work this paper finds somewhat unreliable results – so in short, the adjustments have not led to a significant improvement in the quality of estimates produced.

However, the use of Regional Household Expenditure Estimates produced by the ONS is much more promising. Given the interventions that have already been made in the production of these statistics, much of the volatility has been removed. Benchmarking the weights at the published level smooths out the weights and leads to the production of much more sensible estimates overall. We would recommend that this is the approach taken in regular statistical production.

Overall, we find that regional CPIH and inflation estimates follow similar patterns to national indices. However, there are still issues to deal with, especially in the transition between years, driven by the price quote dataset sample size.

6 References

Dawber,J. & Smith,P., (2017). Feasibility study into producing CPIH consistent inflation rates for UK regions. Report for ONS available at: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/feasibilitystudyintoproducingcpihconsistentinflationratesforukregions

Dawber,J., Wurz, N & Smith,P., (2019). The development of regional CPIH: assessment of small area estimation for regional expenditure weights. Report for ONS available at: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/thedevelopmentofregionalcpihassessmentofsmallareaestimationforregionalexpenditureweights

Duran, H.,(2016). Inflation Differentials across Regions in Turkey. South East European Journal of Economics and Business, 11, 7–7. https://doi.org/10.1515/jeb-2016-0001

HM Treasury. 2018. Scottish VAT assignment: summary of assignment model. Available from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/758498/VAT_assignment_model_publication_final_1.pdf

HM Treasury 2020 Scottish VAT Assignment: 2018 estimate https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938006/November_2020_VAT_Assignment_publication_FINAL_TO_BE_PUBLISHED.pdf

Nagayasu, J.,(2011). Heterogeneity and convergence of regional inflation (prices). Journal of Macroeconomics, 33, 711–723. https://doi.org/10.1016/j.jmacro.2011.07.002

ONS, (2017). Consumer/Retail Price Indices.

ONS, (2018a). Relative regional consumer price levels of goods and services, UK: 2016, Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/relativeregionalconsumerpricelevelsuk/2016

ONS, (2018b). Development of regional household expenditure measures. Available from: https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/articles/developmentofregionalhouseholdexpendituremeasures/2018-09-26

ONS, (2019a). Using alternative data sources in consumer price indices: May 2019 Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/usingalternativedatasourcesinconsumerpriceindices/may2019

ONS, (2019b). Consumer Prices Indices Technical Manual, 2019, Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/consumerpricesindicestechnicalmanual2019

ONS, (2019c). Living Costs and Food Survey technical report: financial years ending March 2017 and March 2018, Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/methodologies/livingcostsandfoodsurveytechnicalreportfinancialyearendingmarch2017

Wurz,N., (2017). Model-based Estimation of Regional Baskets based on the UK Living Cost and Food Survey.