Elementary Index Formulae – thoughts and issues

(Version 2: 23/09/24)

Please note this is a draft paper

Jill Leyland

Foreword

This paper is a work in progress so please regard this as a draft. Its purpose is two-fold so it has two main sections. First, in Part A, I aim to provide an explanation of the three elementary indices that is accessible to the lay reader (with little or no algebra other than in potential appendices) and describes the advantages and disadvantages of each in a balanced fashion including the circumstances under which they will perform less well. I will also attempt to explain what happened to the RPI in 2010 and how it came to lose its accredited (national statistics) status.

In my experience, when discussing with lay people, the axiomatic approach can be both confusing and unconvincing. At the moment I have not referred to it in Part A.

I would welcome comments on this part of the paper that will help make it both more accurate and more accessible.

The second aim - in Part B - is as a contribution to the discussion of the elementary formulae in the Advisory panels and, in particular, whether there should be more use of Dutot in the Household Costs Indices, for example the sort of mix that was used in RPIJ. It deliberately contains more questions than answers. In due course I intend it to evolve into part of the paper but I need help from other panel members to do this.

The debate over elementary formulae, and what happened to the RPI, has been particularly toxic. And it lingers on underneath the surface bubbling up from time to time. It may well resurface as we approach 2030 when, on current plans, the RPI is set, effectively, to turn into CPIH. An aim of this paper when finalised, as well as informing, will be to draw (hopefully) some of the poison from this debate.

Confidence in all official statistics is important – but arguably confidence in consumer price indices is most important of all given their use in uprating, or influencing the uprating of, certain incomes and expenditures. The importance that should be paid to this was neatly summed up in the 1986 Report of the Retail Prices Index Advisory Committee:

"Underlying much of our reasoning in this report is the firmly-held view that it is important to sustain and promote public confidence in the RPI. For the index to be of value it must be generally regarded as relevant to people's concerns and a fair reflection

of their experience. This is partly a question of presentation-ensuring that results are readily accessible and understood-but it also concerns the methods of compilation. For the index to carry conviction these should be understandable and seem reasonable to "the man in the street" as well as to professional analysts or academic experts. Therefore, while we have consciously sought to clarify the concepts and principles underlying the index in a way which is intellectually rigorous, we attach equal importance to the simple test of public acceptability."

I have long thought that a paper which explains the issue in full and in balanced fashion for lay readers is needed in order to meet the aim set out by the 1986 committee. But while I have thought about it for a long time I have not found it easy. Indeed, an earlier attempt by Chris Payne and myself to write one ran into the sand. Part of the problem is the plethora of relevant research papers which touch on the issue. I frequently found myself distracted by starting to read yet another paper which inspired a change or tweak to what had already been written. Another problem is that much of what has been written is highly technical and I have often struggled to follow the arguments.

I expect, and will welcome, a lot of comments.

We are due to have a discussion at the Advisory Panel joint meeting on the 27th September. This may be lively. While Part B lists the issues I would ask the reader not to skip Part A (other than the introduction) since much of the text there will be relevant. There will be differing views but the eventual decisions need, in my view, to be justifiable and explicable to the majority of the lay audience that takes a key interest in, and is directly affected by, the decisions made on the compilation of these indices.

Finally, I am aware that there are places which need references but do not yet have them or where references do not yet indicate urls. These will be added in due course and I would welcome any suggestions as to what to add.

Part A

Introduction *(my current readers may want to skip this section which is intended for the ultimate audience)*.

The compilation of consumer price indices is complex and a number of different techniques are used. However, for most items the basic model is essentially simple and is as follows.

Survey data and other information provide "weights" – the proportion of the household spending on different goods and services in a particular year. The percentage of overall household spending (or household spending for a particular group) that goes to, for example, first apples, then to fruit as a whole and then to food as a whole can be obtained for a "base" year. This information is updated annually in the UK.

Calculation of an index is a hierarchical process. An index is first calculated for, for example, apples[1]. This is combined with indices for, e.g. oranges, bananas etc to make an index for fruit. To do this a "Laspeyres-type[2]" formula is used. January is taken as a base month. When calculating the index for fruit for, say March, the proportion of spending on fruit that was devoted to apples in the latest year for which data are available (typically two years earlier) is multiplied by the growth in the price of apples between January and March. Similar calculations are carried out for oranges, bananas, etc and the results for each category are added together and then divided by the sum of the weights for all fruit to obtain the overall change in fruit prices between January and March. This gives a price index value for March for fruit as a whole.

Subsequently the index for fruit is combined with that for other foodstuffs in the same way, using the proportion of household spending in the base year devoted to fruit, to generate an index for food. That is then combined in the same way to give the overall price index.

By repeating for subsequent months we end up with indices from January of year 1 to January of year 2. This is linked to the indices from the previous and subsequent years to provide the final index.

However, at the very lowest level of aggregation expenditure weights are not always available, and so price indices for many items must be constructed using sub-optimal methods that do not rely on expenditure weights. These are 'unweighted' methods.

[1] In practice this may be an even finer grouping such as apples in a certain type of outlet and in a certain region.

[2] Technically the index used in the UK is a Lowe index since the weights are from a different period to the base.

In future years data for some of these categories will be replaced by data with weights. such as supermarket scanner data. However, unweighted indices will still need to be widely used for a substantial proportion of food and drink outlets as well as for many other expenditure categories.

All unweighted index number methods are suboptimal. They are a necessary compromise that is used when there is no appropriate data on expenditure to inform us of the relative economic importance of the different products being aggregated. The choice of method at the elementary aggregate level has been shown to sometimes make a significant difference to the results and so these methods remain the subject of debate.

The importance of the choice of method has been highlighted in the UK by issues related to the treatment of clothing data in both RPI and CPI, the developments which led to the RPI losing its national statistics (accredited) status in 2013 and the plan to effectively turn it into CPIH from 2030.

The aim of this paper is to provide an explanation of the three commonly used elementary aggregate formulae, their strengths and their weaknesses, that is comprehensible to lay people. It will also trace what happened with the RPI and why it started to give too high estimates of inflation from 2010. It concludes with a discussion of issues around the subject.

The three formulae

We start by looking at each of the three elementary indices commonly used in the calculation of consumer price indices. Each of these is named after the statistician or economist who first invented or popularised them.

Dutot

Dutot - or the ratio of averages - is the simplest and most intuitive of the elementary indices. To calculate it you add up all the observations in the base period, all the observations in the current period (the period for which you are calculating the index) and divide the current period sum by the base period sum.

Its advantage is its simplicity and consequent public credibility.

It has the one problem that all arithmetic averages face in that it is sensitive to high values. To take an extreme example, let us assume that we have a sample of 5 observations in the base and the current period as follows:

	Price January	Price March
Product A	2.0	2.2
Product B	2.2	2.3
Product C	2.5	2.5
Product D	10.0	9.5
Total	**16.7**	**16.5**

As can be seen three of the prices either remain the same or increase slightly. But the fall in the price of product D dominates and an overall Dutot index would show a fall in prices. While an example this extreme is unlikely to happen in practice without D being rejected as an outlier, less extreme examples could still lead to a sub-optimal result. "Tables" is an example of a category where a minority of prices could easily be markedly higher than the majority.

For this reason, there has generally been a convention that Dutot should be used only in circumstances where products are relatively homogeneous so the range of prices is small.

Carli

Carli is the other index that was widely used historically in the compilation of consumer price indices and is still used in the RPI. It is also known as the "average of relatives". A price relative for any item is the ratio of the price in the current month to a previous month. The relatives are averaged to produce the Carli. A key – and often overlooked and misunderstood – issue with Carli is that there are two different versions. In one – the direct version – the relatives are the ratio of the current price for any item in the sample to its price in the base period. In the other – the chained version – the price relative is the ratio of the current price for the item to its price in the previous period. (In both Dutot and Jevons there is no difference between a chained and a direct version.)

Confusion over the two systems has given rise to one of the biggest myths about the use of Carli in the UK. It is often alleged that if prices rise in one period - say from month 1 to month 2 - and then fall back to their month 1 level in month 3 a Carli index will show a rise between month 1 and month 3. This – an event known as "chain drift" - is true of the chained index. A little thought will show that it cannot be true of the direct index. Only the direct version is used in the UK.

Nevertheless, the Carli remains an arithmetic average and consequently can be overly influenced by high values. It seems strange that while it has been well understood that

Dutot should not be used when price **levels** were very heterogeneous, no corresponding practice appeared for Carli in respect of the price **relatives**.

The following table shows an example of how Carli can be skewed.

	January price	**March price**	**Price relative**
Product A	1.0	6.0	6.0
Product B	5.5	5.5	1.0
Product C	6.0	6.0	1.0
Product D	4.5	4.8	1.1
Carli			**2.3**

The low price of Product A in January compared to the others means that its price relative is very high and this dominates the calculation giving an unreasonable result. This is the sort of scenario that could arise if Product A is on sale in January with the other items at normal price.

Chaining in the UK occurs only at higher levels of aggregation when the Laspeyres-type index is used. "Chain drift" can be an issue with weighted indices too although normally the effects are relatively minor. There is, however, some evidence *(ref to add)* that when a Carli is used at the elementary level the resulting chain drift in the higher-level index is worse than when other elementary indices are used.

Finally, it can be shown that provided certain conditions are satisfied Carli is an unbiased estimator of the Lowe (Laspeyres-type) index traditionally used at higher aggregation stages in index compilation.

Jevons

The third average used is the geometric mean known as the Jevons. To calculate a geometric mean, multiply data together and then take the nth root where n is the number of items. Thus for two items, multiply them and take the square root, for three items the cube root and so on.

The Jevons index can be calculated in two ways which conveniently give the same mathematical result. Either take the geometric mean of items in the current period and divide by the geometric mean of items in the base period. Or calculate the price relatives for each item, as in the first stage of the Carli, and then take their geometric mean.

There is one clear flaw in the Jevons index. If any item has a price of zero then the whole index becomes zero. This may be a rare occurrence but it can, for example, apply to car parking charges; also a product may be offered for free if another product is purchased.

More frequently though if one price falls to a very low level then that can influence the results. For example:

	January price	**March price**	**Price relative**
Product A	5.0	6.0	1.20
Product B	5.5	8.0	1.45
Product C	6.0	6.7	1.12
Product D	4.5	0.5	0.11
Geo mean	**5.22**	**3.56**	**0.68**
Carli			0.97
Dutot			1.01

Here we have three products which rise in price and one which falls sharply. A jevons index is clearly overly affected by product D. While we would expect that a fall of this magnitude would be considered as an outlier and disregarded (or modified) a less extreme example could still affect the result. Dutot and Carli, particularly Dutot, produce more reasonable looking results in this case.

I have deliberately used extremes in the three examples above to illustrate the points. While examples this extreme may rarely occur in practice, or the outlier may be disregarded or modified, they do point to where dangers may lie and where values that are less extreme could still affect the results.

The different outcomes have implications for price validation procedures. For Jevons it is important to look for suspiciously low price relatives, for Carli suspiciously high ones and for Dutot it is important to look out for high initial values.

Relationship between the three indices

It can be shown mathematically that Carli will always produce higher index growth than Jevons (equal at the extreme if all price relatives are identical) and the greater the variability in price relatives, the greater the gap. Dutot may be higher or lower than Jevons depending on whether prices in the current or the base period are more variable. The relationship between Carli and Dutot also varies; if price relatives are positively correlated with base prices (that is the higher the base price the higher the price relative tends to be) then Dutot is higher, but if (as is usually the case in practice) they are negatively correlated then Carli is higher.

Appendix A tests the robustness of the different indices to outliers. It shows a stylised example of how the indices react when presented with one item with a variable value, in

either the base period or the current period, when 9 other items in the notional database are stable. Whether each formula performs well or badly under these circumstances is partly a matter of judgement but some conclusions are fairly clear:

- Carli clearly performs badly when you have a base period sample where one item (or a minority of items) are much lower than the others. Dutot is least influenced by the varying item(s) in these circumstances but, as seen earlier, can be distorted by an item priced significantly higher in the base period.
- Jevons can have a small downward bias in extreme cases; in three of the four examples it gives lower results than the other two.

Empirical assessments using real data tend to show that indices calculated with either Dutot or Jevons are likely to show broadly similar overall rates of inflation, provided care is taken when exceptionally high or low values appear, while Carli will often tend to give somewhat higher results[3]. (The footnoted paper by Schultz – also known as Szulc – is a particularly comprehensive example of how different indices, in both chained and unchained versions, perform using real data. See the appendices in particular.)

We have seen therefore that all three formulae can, under certain circumstances, generate results that seem perverse. At other times they generate results that seem fully acceptable. As the Royal Statistical Society has consistently pointed out, it is the interaction of the data set with the formula that is crucial. Thus the likely characteristics of a data set need to be considered before deciding which formula, or formulae, could be appropriate.

Jevons and consumer behaviour; COLI and COGI

In the 1990s the US Senate Finance Committee commissioned a study into the question of whether the US Consumer Prices Index overestimated inflation. At the time there were concerns that the US CPI overestimated inflation and that that was impacting certain items of Federal spending, notably social security. The Commission – known generally as the Boskin Commission after its Chair - found four reasons why they considered that the USA CPI overestimated inflation. In particular they considered that it was affected by various forms of "substitution bias" – that it did not allow for occasions when consumers were able to reduce their expenditures without reducing their welfare by purchasing new products, switching to different retail outlets or simply purchasing cheaper, but equally satisfactory, brands of the same product. One of the reasons identified was the arithmetic formula then used at the lowest level of aggregation. The Commission argued a geometric mean was superior as it allowed for

[3] An early example of the way different elementary indices, perform was given in Schultz "Choice of Price Index Formulae at the Micro Aggregation Level" f4.pdf (unece.org), 1994 revised 1995. This remains one of the most comprehensive assessments of the behaviour of different elementary indices using actual data.

the fact that consumers tended to buy more of any brand that increased less in price or decreased more than its competitors.

While discussion about the relative merits of the different elementary formulae was far from new, the Boskin report generated substantial interest in other countries and its arguments gained traction. A number of countries were persuaded by the argument that Jevons allowed for consumers substituting towards products where price changes were smaller or more negative[4]. Previously most countries had used either Dutot or Carli or a mixture of both. Among other factors, the limited computer power of the past favoured relatively simple calculations. There had always been theorists – some influential - who advocated the use of a geometric mean but initially this was little used in either the elementary levels of aggregation or the higher levels[5]. But by the 1990s the formerly limited computer power was no longer an impediment.

A further push away from the use of Carli and towards Jevons came from the development of the EU Harmonised Indices of Consumer Prices (HICPs) during the 1990s. The spur for the development of these indices was the provision in the 1992 Treaty on European Union (Maastricht Treaty) concerning the tests candidate countries for the Euro had to undergo. One of these stipulated that candidates should have a rate of inflation no higher than 1.5 percentage points above the average of the three best performing countries. Inter-country comparability was therefore crucial. Consequently countries are permitted to use either Jevons or Dutot in their HICPs but not Carli unless it can be shown that Carli will generate broadly similar results[6].

Returning to the issue of Jevons proxying consumer behaviour, there were counter arguments. First, substituting towards products that have become relatively cheaper is only part of consumer behaviour which is far more complex. Second, as Courtney[7] has pointed out, this ignores the fact that price is set by the interaction between supply and demand and consumers, while they can act as Boskin suggested, are not simply price takers. In practice they can influence prices (e.g. popular brands can rise in price, products which do not generate sufficient sales are reduced in price). Third, demand for some products can be considered as not very price elastic in which case an arithmetic average would be more appropriate. Fourth, research by ONS[8] did not support the theory. Fifth, Professor Erwin Diewert, an international expert commissioned to examine the UK formula effect by the ONS in 2012, has argued that the idea that an unweighted

[4] See the appendix to International-comparison-of-the-formula-effect-between-the-cpi-and-rpi/2012/ by Bethan Evans. This summarised the findings from ONS research into the formulae used by comparable countries.

[5] If my memory from my time of working at the OECD in the 1970s is correct, Ireland was the only member country to use geometric means in its CPI.

[6] See the section on Computation Issues in Eurostat - Statistics Explained: HICP Methodology

[7] Mark Courtney Consumer Price Indices in the UK

[8] Joseph Winton, Robert O'Neill and Duncan Elliott *Elementary Aggregate Indices and Lower Level Substitution Bias*, ONS, March 2012

Jevons formula proxies consumer behaviour is misconceived and this only applies to a weighted formula. So this argument for Jevons is no longer used in the UK although it was widely used in the past.

There is also a more fundamental point. Unusually, the US Bureau of Labor Statistics now aims to make its CPI a Cost of Living Index or COLI. The phrase "Cost of Living" in this context has a very specific meaning - differing from the ways of everyday usage – and a COLI is an index which aims to measure the change in costs which a household would have to make in order to hold constant some specified standard of utility or well-being. While the meaning is fairly precise its interpretation can be wide. Many papers have been written about the meaning and the compilation of COLIs[9].

Most countries, including the UK, and the European Union with regard to its Harmonised Index of Consumer Prices (HICP), do not aim to make their consumer price indices COLIs. Instead they are Cost of Goods Indices (COGIs) – also known as Pure Price Indices – measuring the price evolution of a fixed basket of goods and services.

While the basket of goods and services is typically updated each year to reflect the change in composition of household purchases, the basket is fixed within the year.

Insofar as Jevons can be said to proxy consumer substitution, therefore, it is not consistent with the COGI concept. Some experts - notably Sir Roy Allen, a leading expert in the past and heavily involved in the early development of the RPI – argued that while a geometric mean might have superior mathematical properties, it should not be used as it was incompatible with the aims of a consumer price index based on the "cost of goods" principle[10].

What happened to the RPI?

As mentioned, the RPI uses, and always used, a mixture of Carli and Dutot for its elementary indices where weights were not available.

The HICPs were first published in the late 1990s. It soon became clear that the UK HICP – which from 2003 was known in this country as the CPI – showed lower inflation figures than the RPI. Various factors (including differences in coverage or weighting) contributed to the difference, or "wedge" as it is normally referred to, between the two series but the main factor, and one which was always in one direction, was the difference caused by the use of different elementary formulae. The CPI mainly used the Jevons index, with a small use of Dutot, while the RPI used a mixture of Dutot and Carli. Up to 2009 the annual difference caused by the formula effect was around or below 0.5 percentage points.

[9] See, for example, US National Research Council, "At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes". National Academy Press, 2002.

[10] "The geometric index does not make economic sense; the arithmetic index does." R.G.D. Allen: *Index Numbers in Theory and Practice, Macmillan Press, 1975 ISBN: 0-33-16916-6 Page 18.*

In 2008- 09 Eurostat carried out one of its periodic assessments of the UK's HICP (the CPI). Eurostat queried the index for clothing in the CPI/HICP, which showed negative growth, out of line with other countries. Clothing inflation also looked low in the RPI but much less so than in the CPI. The reason appeared as follows. A new piece of clothing normally enters the market at full price. If it does not sell well the price is reduced, often substantially. Eventually the item disappears and a replacement has to be selected. If the replacement is deemed similar to the previous item, its price is taken as is. If not, it is treated as non-compatible and its price is "linked" which means that its price (normally at full value) was taken to be equivalent to the final, discounted, price of the former item. This created a downward rachet. This was exacerbated in the CPI by the tendency of Jevons to under - rather than over - estimate while the Carli's tendency to produce a higher value partly offset it.

In 2009 ONS decided to make some "improvements" to clothing price collection from 2010 to correct for the underestimation of clothing inflation in the CPI. Three changes were made:

- Small changes in composition and style were no longer considered to render a new garment non-comparable with its predecessor thus increasing the effective sample size;
- When new items were being selected, or new outlets included, products on sale with reduced prices in the base month of January were included; previously collectors had been instructed to avoid items on sale;
- Prices were collected for all available products in January

Disastrously the changes were considered to be so minor that they were implemented without any prior testing. They were neither discussed in the then Advisory Committee nor referred to the Bank of England.

The new guidelines had the impact of increasing the number of sale prices used in the base month of January. The proportion of sales prices included jumped from around 12% in earlier years to 18% in 2010 and then 22% in 2011[11]. This automatically increased the variability of the price relatives and created precisely the conditions in which Carli gives an exaggerated indication of growth.

The formula effect for clothing surged - in 2012 it averaged 6 percentage points - and that for the RPI/CPI as a whole jumped from around or under 0.5 percentage points to around a whole percentage point. While the changes made the CPI clothing figures look

[11] Source CPAC papers (link to add).

more plausible, those for RPI clearly showed implausibly high growth levels and this impacted the overall index[12].

In 2011 the ONS started a research programme to investigate the matter in full. This consisted of various strands of theoretical research and an investigation into the formulae used by other countries and their experience. It also included testing a number of changes to the sampling pattern, for example: tightening some item descriptions; alternative methods of outlet stratification, the introduction of seasonal items and a change in price collection guidelines. Finally, the impact of changing the base month from January to December was also tested.

Most of these sampling pattern changes appeared to have only a very small effect (although work by the RSS suggested[13] that cumulatively they were more effective). The exception was the change in the base month from January to December which did narrow the formula effect although by raising the CPI inflation rate rather than reducing the RPI rate. It was, however only tested over one year. Meanwhile the ONS was alarmed to discover that no other comparable country was still using Carli[14]. Political and public interest in the whole issue was live, heightened by the fact that in 2010 the incoming Coalition Government had decided to switch from using the RPI to the CPI for the uprating of public sector pensions and some benefits.

In the Autumn of 2012 it was decided to cut short the research programme and carry out a public consultation into whether Carli should be replaced in the RPI by Jevons or Dutot either partly (e.g. just for clothing) or entirely. Over 400 people responded and there was an overwhelming majority in favour of keeping the status quo. This led to the National Statistician accepting that the RPI should remain unchanged and proposing that its methodology should remain broadly unchanged in the future. She also asked for the RPI's national statistics status, which had been confirmed just a short time earlier, to be re-assessed. In 2013 the RPI was stripped of its national statistics status due to a) its use of the Carli formula deemed not in line with international best practice and b) its methodology being "frozen" so that it was not open to methodological improvements.

The ongoing divergence between the inflation rates shown by CPI and RPI remained a live issue. However, legal requirements prevent any fundamental change which would make a material reduction in RPI inflation before 2030 without the agreement of the

[12] It seems strange that the ONS did not realise the danger of what might happen. However, this took place during or just after the relocation of ONS from London to Newport. Only 12-15% of existing ONS staff moved to Newport with the prices division among those particularly affected so there was a substantial loss of knowledge and expertise which took several years to rebuild.

[13] Ref to add

[14] See Bethan Evans, *Op cit*.

Chancellor of the Exchequer[15]. In 2020 the UKSA decided that from 2030 the RPI will be calculated in the same way as CPIH.

[15] Statistics and Registration Service Act, 2007, Section 21. This results from conditions concerning Index-linked government bonds issued before mid-2002. The last of these bonds matures in 2030.

Section B: Questions for discussion

Here I would ask Members of the Panels not just to give their opinions but to advise how to express them in a way that is accessible to the lay person. In the fulness of time this should evolve into another section of the paper. Several Members of the Technical Panel have written on this subject but in general their papers are highly mathematical.

No unweighted formula is perfect. The Royal Statistical Society has always argued for a "horses for courses" approach and that different formulae could be appropriate according to the characteristics of the underlying data set.

Carli.

While its use has almost disappeared from consumer price indices, I understand it is still used by certain countries in producer price indices – is this true?

The direct Carli does not deserve all the opprobrium that has been heaped upon it and it can be an unbiased estimator of a Lowe index. However, under certain circumstances it can give misleading results. It is clearly not suitable for clothing given the current sampling system with January as the base month and sales prices included. But is its use entirely out of the question and if so how to explain this to a lay audience?

Carli is often criticised for failing various of the tests in the axiomatic approach. Statisticians differ both in how many of these tests are defined and also in their estimate of their importance. In general, the following tests that Carli fails are mentioned:

Transitivity. Over a period of two or more months the chained Carli will give a different result than the direct Carli. In contrast with both Dutot and Jevons where the chained and direct versions will give identical results.

Time reversal. The index for month y with base month x should be the reciprocal of the index for month x with base month y. Both Dutot and Jevons pass this test.

Price Bouncing. Suppose shops a,b,c,d have a certain set of prices in the base month. If in month x they have the same set of prices, but exchanged between themselves, then the index for month x should be 100. (Not everyone agrees with this premise.) This is the case for Jevons and Dutot but not for Carli.

Conversely however, there are properties of the Carli that are not possessed by the other formulae. For example, if the price quotes for an item are split into two groups, a Carli calculated for each and then the two are combined, weighting by the number of price quotes, the result is the same as if the prices are not split. This is not true of Dutot or Jevons.

One can argue that each one of these tests can be disregarded. And there is no consistent view among statisticians as to which of these are important in practice.

While most experts argue that Carli should never be used, Courtney[16] argues that when prices are primarily demand driven it underestimates inflation less than other indices.

Its use is generally not permitted in the HICP which is an important international comparator.

A key question is whether the axiomatic approach is a valid way of deciding on an elementary formula given that different experts differ in their assessment of each axiom's importance. Is it useful or should we agree with Triplett and Reinsdorf: " We have contended that the test approach to index numbers cannot be used for guidance about aggregating the basic components (see paragraph section 1). The tests included and excluded from the tests are arbitrary and the index number that emerges as "best" is very sensitive to the arbitrary choice of tests."[17]

Jevons.

A key question is whether it is compatible with the Cost Of Goods approach. If not, does it matter since there always has to be an element of pragmatism in consumer price indices? It would not be the only major inconsistency in UK consumer price indices: neither the use of rental equivalence in CPIH nor the lagging in the calculation of rents[18] are compatible with the acquisition principle on which CPI is based.

It has convenient mathematical properties.

It passes all the usual axiomatic tests but would fail the little-known axiom quoted above.

It is, however, viewed with some suspicion in some quarters given that it tends to provide lower estimates of inflation than Carli and at times than Dutot.

It is permitted in the HICP.

What can now be said about it proxying consumer reaction to price changes?

Dutot

A great advantage of Dutot is the fact that it is simple and easily explained. It can therefore benefit from high public acceptability. Its disadvantage is that when price levels, particularly those in the base period, are very variable it is overly influenced by

[16] Op cit.
[17] A Review of Reviews: Ninety Years of Professional Thinking about the Consumer Price Index in NBER: Price Index Concepts and Measurement, 2004. University of Chicago Press ISBN : 0-226-14855-6
[18] In contrast the lagging in rents is compatible with the payment principle behind HCIs.

the higher prices. (In the axiomatic approach it fails the commensurability test but this amounts to the same thing in practice.) It fails the little known axiom quoted above.

It is permitted in the HICP.

The Household Costs Indices are intended to measure inflation as households experience it. If in the fullness of time the indices are used for uprating purposes public credibility will be particularly important. Should therefore Dutot be used in the HCIs, at least in areas where prices are reasonably homogenous?

Other elementary price indices.

Other formulae, such as the Carruthers-Sellwood-Ward-Dalen (CSWD) which was at one time used in the Swedish index, are more complicated and harder for the lay person to understand.

Other points for discussion

1. With increasing complexity of calculation (eg GEKS-T) with large data-sets how much will it be possible to explain to lay people? Can we explain the principle of the calculations at least[19]?
2. How treatment of outliers is carried out is clearly important. Are there any particular points to highlight- eg that with Jevons it is more important to recognise low price relatives while in Dutot it is crucial to pay attention to high price levels?
3. As a base month, January has disadvantages. HICP rules stipulate December and in recent years ONS has carried out an adjustment so that December is in many ways the effective base month. But does this overcome the problems with the use of sales prices in January?
4. How crucial is homogeneity of prices when choosing samples?
5. Can one argue that there is an advantage in using different formulae so that any bias from a particular formula is offset?

[19] I have always been surprised at the interest many lay people do show in the calculation – particularly when the index affects their income or major outgoings. I was once asked to explain hedonics to a group of pensioners.

Appendix A

I would welcome suggestions for any other empirical or stylised tests that would be useful.

Robustness

To investigate how sensitive the three indices are to outliers we can look at a stylised example[20] to assess how sensitive each formula is and thus how likely to generate a result which seems to be skewed away from what one would expect. The aim of this is simply to understand better how the various formulae behave under circumstances varying from the banal to the extreme.

Suppose we have a data set of 10 prices. 9 of them take the value 1 in both the base period and the current period (the period being measured). The value of the final pair varies. In the first example, case a, below it takes the value x which varies from 0.1 to 10.0 in the base period with the current period remaining at 1. In the second example, case b, it is 1 in the base period and takes the value x in the current period which again varies from 0.1 to 10.

The values of the elementary aggregates are easily calculated.

Dutot is the average of the prices for the current period divided by the average for the base period. In case a) the average of the current period would be 1 and that for the base period would be (9+x)/10 so the value of Dutot would be 10/(9 + x).

Carli is the average of relatives. The relatives for the 9 pairs that are stable would all be 1 while that for the remaining pair would be 1/x. Thus the value of Carli would be (9 + 1/x)/10.

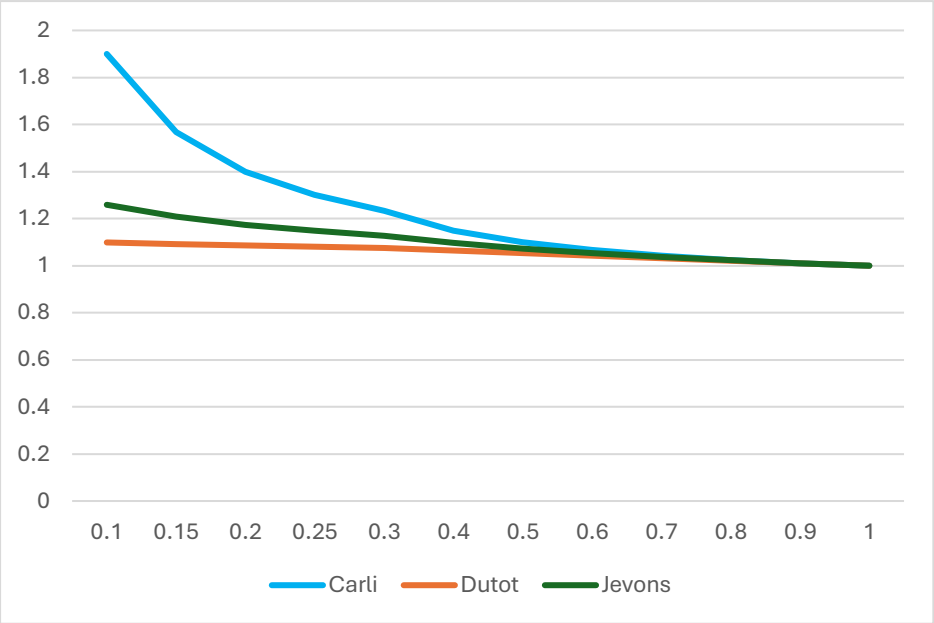
Jevons would be the 10^{th} root of the product of the values in the current period, which would be 1, divided by the product of the values in the base period; the latter being x since all other values are 1. Jevons would therefore be $(1/x)^{0.1}$.

Similarly, in case b we have:

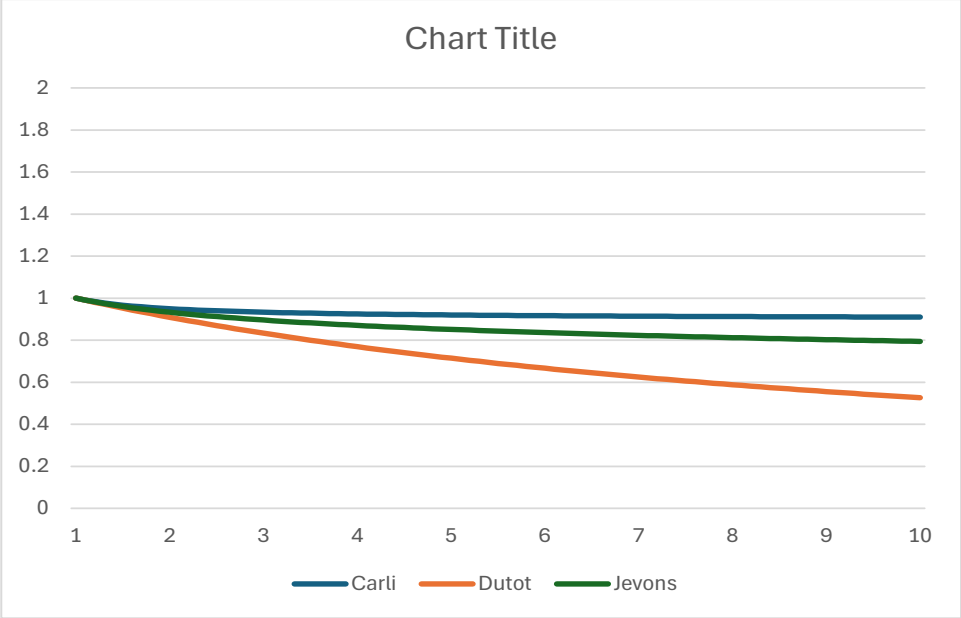
Dutot: (9+x)/10
Carli: (9+x)/10
Jevons: $x^{0.1}$

[20] I am indebted to Michael Baxter for suggesting this approach.

Looking at case a when the tenth item varied from 0.1 to 1.0 we obtain the following chart:



And when the tenth item varies from 1 to 10:

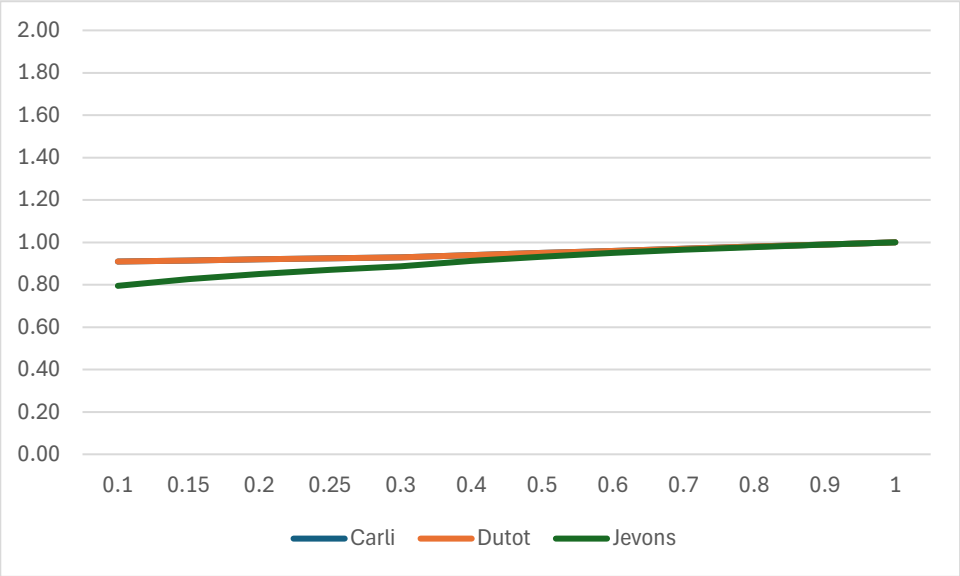


It is a matter of judgement as to how far one would wish the variable item to affect the overall result. But the charts do suggest that when the price of the tenth item is substantially less than 1.0 then Carli gives an overly high result while Dutot performs badly when the price varies is much higher. This is not entirely surprising. Both being arithmetic averages they are sensitive to high values. When the tenth item is high in price one would expect Dutot to perform poorly as it does. Note though that Carli in this case is little affected by the variable item and produces a credible looking result. When the tenth item is very low it will provide a

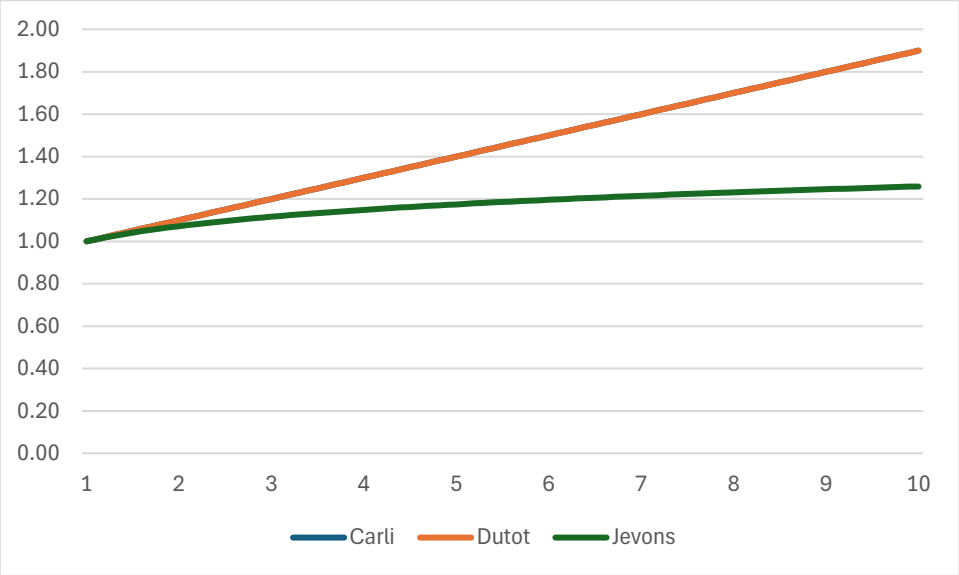
high price relative so Carli performs badly – in act it produces a completely unacceptable result. And in this case Dutot is the most robust. Jevons lies between the two.

For **case b**, we have the following charts:

When the 10th item varies from 0.1 to 1.0:



When the 10th item varies from 1 to 10:



Carli and Dutot give identical results. Jevons is less robust for one low price and more robust for one high one. In both cases it is below Carli and Dutot (except when the 10^{th} item equals one).

In some of these cases it is a matter of judgement as to which index gives the better or best results or whether and when the 10^{th} items would have been identified as an outlier and either eliminated from the calculation or had its weight much reduced. However, the first two charts show up the weakness of Carli and Dutot respectively. Charts 1 and 3 show that Jevons becomes questionable when the tenth item approaches its extreme values although in Chart 1 it remains far more acceptable than Carli. In Chart 4 it gives the most robust answer,