

## Statisticians' arguments in favour of the RPI

Only a few statisticians have pronounced explicitly on the merits of the RPI in the heavily politicised political debate in the UK since 2010. The RPI was so well accepted and its statistical basis so well understood, that that it would have been unnecessary and beyond their strictly statistical competence to have commended its general use. Instead, their predominant contribution has been to defend the use of the Carli (arithmetic average of price relatives) index that the RPI employs in first-stage aggregation, since criticism of the Carli was the only statistical reason given first by the Government and later by the ONS for turning instead to the CPI, which uses the Jevons (geometric average of price relatives).

To put the contribution by individual statisticians in defence of the Carli in perspective, it is necessary to bear in mind that the Carli index has been attacked by the UK Statistics Authority/Office for National Statistics on a number of grounds, with the basis of attack shifting as each attack has been discredited. These are, in order of appearance in the UK debate:

### The Economic Approach

This was the one statistical reason given by the Government when introducing the CPI in 2003 as a monetary target and extending its use in 2010, and was also the reason why the USA and Australia had previously reduced the use of the Carli index in their consumer price indices. It is the most respectable criticism statistically, in that it can legitimately be used as a basis for preferring the Jevons to the Carli index *provided* three conditions are met:

- i. We want our consumer price index not to be a cost-of-goods index like the RPI (measuring the price change of a typical basket of goods and services), but a constant-utility index (that can be used to compensate consumers to leave their welfare unchanged once they have adjusted their purchasing patterns in reaction to price changes).
- ii. Price changes are caused only by changes in the supply of goods, rather than by the interaction of changes in supply and demand (i.e. it is reasonable to assume that consumer tastes are constant, uninfluenced by novelty, fashion or advertising).
- iii. Consumer demand is highly "elastic": i.e. the quantity demanded changes in an opposite direction and at least in proportion to changes in its price.

Thus, the defence of the Carli and of the RPI rests on showing that one or more of these conditions does not hold in the UK. Conditions ii and iii have been shown by the statisticians noted below not to hold. Condition i. is a matter of user choice rather than of statistics.

### The Statistical (Stochastic or Sampling) Approach

In contrast to the Economic Approach, the Statistical Approach has almost universally been agreed to support the use of the Carli over the Jevons for elementary aggregation, since both theory and available empirical evidence show that the Carli is an unbiased estimator of the weighted Lowe index used as a target index in both the RPI and CPI, whereas the Jevons is downward-biased. However, critics of the RPI (ONS in 2012, the Johnson Review in 2015) argued that this result might depend on the distribution of price relatives, even though John Greenlees – see below – had already in 2001 demonstrated this to be false.

## The Axiomatic or Test Approach

This approach, proposed in the 1920's by Irving Fisher, looks at possible aggregation formulae and ranks them on whether they have various mathematical properties that are thought to be desirable. In fact, there has always been disagreement among statisticians over which of these properties are theoretically desirable or empirically important, and it had very little influence on the construction on price indices, so that by 1975 Professor Sir Roy Allen – see below – could dismiss it as “mainly of historical interest.” The approach was revived, however, in the 1990's because, by a choice of desirable criteria, it could be cited as a statistical justification for switching from one formula to another that yielded a lower inflation rate figure. It was taken up by the ONS in 2012 when its own research division and the statisticians cited below convinced it that the hitherto predominant Economic Approach provided no reason to discard the Carli. By 2018 the ONS had to admit that, given how the RPI is constructed, the lack of transitivity of the Carli index is irrelevant, and therefore the test approach provided no argument against using the Carli in the RPI. This removed the last statistical reason for the UK Statistics Authority not to restore National Statistic status to the RPI.

## The Significance of Clothing

Having had to admit that none of the recognised statistical arguments justifies discontinuing the RPI, the ONS has, since 2016, increasingly shifted to emphasising that, following the change it made to clothing price sampling in 2010, the RPI's clothing sub-index has shown an implausible upward trend. It is, therefore, significant that many statisticians – even those generally in favour of using the Jevons index – have pointed out the special difficulties with sampling clothing prices, regardless of the aggregation index used, and that mitigating this problem requires adjustments to clothing sampling and/or ad hoc adjustments applied just to the clothing sector – a conclusion that also figured prominently in the “Measuring Inflation” report of the House of Lords Economic Affairs Committee.

## Statisticians, in Alphabetical Order

**Professor Sir Roy Allen, (RGD Allen), 1906-1983.** Roy Allen was Professor of Statistics at the LSE from 1944 to 1973 and – among many other public appointments – was a member of the Cost of Living Advisory Committee (renamed the Retail Prices Advisory Committee in 1968) from its inception in 1946 to his death in 1983. His 1975 book, *Index Numbers in Theory and Practice*, is the most comprehensive and still probably the best text ever written on this statistically intractable field. The most eminent British Statistician of his generation, the modern RPI is very largely his creation and his book is still very topical, since it shows that the statistical weaknesses more recently alleged against the RPI, including its use of the Carli, had already been anticipated and discounted. Two results in particular, if they had been assimilated by the ONS, would have prevented reliance on either the economic approach or the test approach to discredit the use of the Carli:

- i. In applying an economic-theoretic analysis to index numbers, Roy Allen makes a distinction often neglected by his successors, pointing out (Section 2.7) that there are two plausible situations: (i) where the market is supply-dominated in that suppliers adjust their supply to shifts in demand, and observed movements in prices and quantities tend to go in the same direction, in which case the Paasche index of price is greater than the Laspeyres; (ii) where the market is demand-dominated and prices and

quantities tend to move in opposite directions, in which case the Laspeyres form of the retail price index is generally the greater. Only case (ii) is relevant for a constant-utility price index, which is the basis of the Economic Approach. **Courtney** (2013) draws out the implication that it is only in case (ii) that the Geometric Mean might fall between the Laspeyres and Paasche indices and the Jevons might therefore be a better estimator for a Fisher index, interpreted as a constant-utility index, than the Carli, whereas in case (i), the Carli has to be the better estimator of the Fisher. These theoretical results were demonstrated in simulations by **Winton et al** (2012), while empirical work by them and others showed that case (i) – the existence of “fashion goods” in modern terminology – is empirically important and that in the UK the Carli is at least as good an estimator of the Fisher index as is the Jevons.

- ii. Roy Allen has an illuminating chapter (Chapter 5) on chain index numbers. This deals entirely with chaining weighted index numbers, since he knew that in the RPI there was no question of chaining the unweighted elementary aggregation indices. However, the mathematical properties of the weighted and unweighted arithmetic average of price relatives (the Laspeyres and Carli indices) are similar. In both cases, because the arithmetic indices do not have the statistical property of “transitivity” there can be chain drift, with the chained index diverging from the direct index. Roy Allen establishes that there is no theoretical presumption that the drift will be in either direction (contrary to claims by the ONS of an upward bias), and his empirical results for the RPI indicated that with annual chaining of the RPI at the section level, the chained Laspeyres form drifts moderately (two percentage points in ten years) below its direct counterpart. Roy Allen observes that it is likely, but not certain, that the chain Laspeyres is closer to the Fisher index, (the geometric average of the two, taken as a good proxy for a constant-utility index), than is the direct index.

**Professor Bert Balk**, Emeritus Professor at the Rotterdam School of Management, Erasmus University and formerly Senior Researcher at Statistics Netherlands is probably the most eminent practising statistician active in the field of index numbers. His book, *Price and Quantity Index Numbers*, CUP, 2008 is the current authoritative work in the field. It covers substantially the same theoretical ground as Roy Allen, and reproduces his results in a more rigorous and systematic way, although without his practical and empirical extensions. Three of his results are particularly relevant:

- i. Bert Balk shows that the choice of aggregation formula for a price index depends on the objective or target that one is trying to measure. There is no single “best” measure of price inflation, contrary to ONS assertions.
- ii. He reaches clear-cut conclusions (Section 5.7) about the appropriate aggregation index to be used depending on the target index. Specifically, “When the target is a Lowe price index, the sample Lowe and Carli price indices exhibit, dependent on the sampling design (pps sampling relative to period price-updated value shares), appropriate behaviour.” Since the target for the UK’s CPI and the RPI is indeed, in both cases, a Lowe index, this is an unambiguous recommendation for the RPI, which uses the appropriate sample price indices, ahead of the CPI, which does not.
- iii. Professor Balk also reaffirms Professor Allen’s result that nothing can be said with certainty about whether a chained index will be above or below the direct index, since

it all depends on the covariance of price changes and quantity changes in the relevant periods.

**Dr Arthur Barnett**, a former Government statistician and a committee member of the RPI/CPI User Group, is one of a number of statisticians who have pointed out that the perceived problems with the RPI since 2010 are the result primarily of an error in clothing price collection methodology. This proposition was accepted by the House of Lords Economic Affairs Committee in its report, *Measuring Inflation*, where the Summary lists it as the sole problem in the RPI that needs fixing:

**“Fixing RPI:** The present position of the Authority is untenable. Rather than pre-empting the decision of the Chancellor, it should fulfil its statutory duty to promote and safeguard the quality of official statistics and to do that, it should request a fix to the clothing problem.”

Dr Barnett, in his submission to the House of Lords Economic Affairs Committee inquiry also provided a concise explanation of why the alleged problems with chain-linking the Carli index are irrelevant (a point that is now accepted even by the ONS): “The RPI avoids many of the potential issues by implementing Carli in the following way: the price relatives use a single base month of January; Carli is only used in annual direct indices which are not chain linked.”

**Michael Baxter** is currently Director of Statistics, Media at Kantar Group. He was a member of the Government Statistics Service from 1997 to 2008 and from 2009 was a committee member and sometime Chairman of the Official Statistics Section of the Royal Statistical Society. From 1993 to 1998 he was Head of Methodology of the RPI at the ONS. This gives authority to his evidence to the Public Administration and Constitutional Affairs Committee (PACAC) Inquiry into the Governance of Official Statistics regarding the political pressure exerted on the ONS by the Treasury (following the publication of the Boskin report in the USA) to make similar changes that would have lowered the RPI, and the ONS’ subsequent refusal to do so “bolstered by the views of many former members of these [a series of RPI Advisory Committees].” See also references to **Fenwick** and **O’Donoghue & Wilkie**. Baxter’s publications in *Economic Trends* during this period showed the official, detailed ONS response to the Boskin report and an active research programme to improve it.<sup>1</sup>

Baxter was a co-author of the Royal Statistical Society’s response to the 2012 UKSA consultation on Improving RPI: on the basis of his experience in RPI methodology, he advised that the ONS could within three days provide a good estimate of the extent of the upward bias to the RPI caused by the reaction of the Carli index to “price bouncing” – which was the only statistical fault alleged against the RPI. The failure of the ONS then, or later, to respond to the Royal Statistical Society’s request to produce such an estimate demonstrated that their claim of an upward bias was based on wilful blindness and not just on ignorance (previous theoretical work – **Allen, Balk** – and subsequent indirect empirical evidence – **Jones** – showed that no such upward bias would be expected or actually existed).

---

M. Baxter, “Implications of the US Boskin Report for the UK Retail Prices Index,” *Economic Trends* No. 527, October 1997 56-62; M. Baxter and D. Camus, “Three Year Research Programme on RPI Methodology,” *Economic Trends*, No. 543 February 1999 25-29<sup>1</sup>

**Simon Briscoe** is a consultant in statistics and economics, an adviser to parliament's Public Administration and Constitutional Affairs Committee and the director of a data science start-up. He is a trustee of the fact-checking charity Full Fact and a council member of the Royal Statistical Society. He was Statistics Editor at the Financial Times for a decade and, before that worked as a statistician and economist in the civil service and then in investment banking. He appeared as an expert witness, defending the appropriateness of the RPI in the landmark case, *British Telecommunications PLC v BT Pension Scheme Trustees Limited & Bruce-Watt* (High Court, 19 January 2018), in which the Court ruled that RPI remained appropriate as a metric by which pensions were increased to keep pace with increases in the cost of living under the Scheme rules.

Briscoe was called to give oral evidence and also submitted written evidence to the House of Lords inquiry into the use of the Retail Prices Index (report published 17 January 2019). In doing so, he provided evidence that convinced the Committee that "In publishing an index which it admits is flawed but refuses to fix, the Authority could be accused of failing in its statutory duties." Briscoe also provided an outline of a feasible, incremental way in which the UKSA could provide such a fix, including to the problems of measuring clothing inflation.

**Dr Mark Courtney** was Deputy Director and Head of Economics in the Regulatory Impact Unit in the Cabinet Office from 1999 to 2007 and, before that Senior Lecturer in Economics at Rhodes University and Economic Adviser at the Treasury. He was a member of the *ad hoc* Consumer Price Indices Technical Group set up by the Office for National Statistics in 2011/12, and was a co-author of the Royal Statistical Society's response to the ONS consultation on improving the RPI, October 2012.

Courtney's widely-cited paper, *CPI and RPI Differences: the Formula Effect and the Identification Problem*, August 2011, was the first to identify the behaviour of the UK clothing price indices before and after the 2010 changes as an example of price behaviour influenced by changes in fashion and consumer taste, and drew out the implication that this made it untenable to use estimates of consumer price elasticity – as in the conventional Economic Approach – to decide on the choice of aggregation index. A shortened and revised version was published as: M M Courtney, *Consumer Price Indices and the Identification Problem*, *Statistical Journal of the IAOS*, vol 29, no 1, 2013, pp.41-52.

In November 2014 (with an updated and revised version in August 2016), Courtney published a comprehensive review article, *Consumer Price Indices in the UK*, which, drawing on theory and recent empirical research by the ONS and others, established that if – as the ONS maintained was still the case – one was aiming at a cost-of-goods consumer price index, the formula-effect difference between the RPI and CPI was due almost entirely to underestimation by the CPI. Even if, instead, the target was an American-style constant-utility index, the RPI was at least as likely as the CPI to provide the best estimate. At the request of the independent members of the Stakeholder Advisory Panel on Consumer Prices (APAC-Stakeholder), the ONS agreed to invite Dr Courtney to present the paper to the first meeting of the ONS-appointed APAC-Technical, whose minutes, "noted that panel members found the mathematical properties of the elementary aggregate formulae finely balanced." This was the first admission within the ONS that the theoretical basis of the RPI was legitimate, although it

had no effect on pronouncements by the ONS and UKSA leaders, who denounced the RPI in terms that grew increasingly shrill as the evidence in its favour accumulated.

Courtney's 2018 paper, *UK Clothing Inflation 1997-2016*, <https://ssrn.com/abstract=3140666>, examined the UK clothing-price indices for the RPI and CPI in the period before 2010 in which use of matched-model resulted in a downward bias in both indices with a greater bias in the CPI, and in the period after 2010 in which looser sampling criteria caused the RPI and CPI sub-indices to diverge. For the latter period, results from the ONS' clothing price pilot study allowed him to make separate estimates of the effects on the indices of clothing mis-sampling and of the formula-effect difference. It turns out that, once the effect of clothing mis-sampling was allowed for, the formula-effect difference for clothing was in line with theory and with empirical evidence from other sectors: the overestimation by the RPI formula is roughly equal to the underestimation by the CPI formula, where the Fisher index is the target. (Where a Laspeyres-type index is the target, as it is for the RPI and CPI, the RPI formulae are generally superior: however, ONS research was in terms of a Fisher-index target and so, for comparability, Courtney used that throughout).

**Professor Erwin Diewert** is a Professor in the School of Economics at the University of British Columbia. It might seem surprising to include him in a list of statisticians who have argued in favour of the RPI, since he was well known as being the world's foremost advocate for replacing the Carli by the Jevons index in calculating a consumer price index, and was hired on that basis as an adviser by the ONS when they decided in May 2012 to make just such a change. His report for the ONS (W Erwin Diewert, *Consumer Price Statistics in the UK*, October 2012) duly recommended replacing the Carli index, but his intellectual honesty meant that, faced with recent theoretical arguments and the empirical evidence in the UK, he admitted that his earlier belief that the Economic Approach provided a weak reason to prefer the Jevons index was mistaken. He also conceded in subsequent comments (Erwin Diewert, *Comments on Consultation Responses*, December 28, 2012) that the Statistical or Stochastic Approach could be used to argue in favour of the Carli, although he thought one should be cautious in applying it.

That left only the Axiomatic or Test approach, which Diewert had always regarded as the strongest argument for use of the Jevons. However, his report reproduced some of his earlier, theoretical work, which (i) pointed out that the price-bouncing test contradicts the principle of comparing like with like and is, therefore, a "suspect" test and (ii) in the case of the circularity or transitivity test, explained that for the Carli, as for other indices that fail this test, this chain drift can be either positive or negative. For the Carli, chaining will result in negative chain drift if prices trend relatively smoothly, which, as Diewert noted, they usually will for annual chaining, (which is the case in the RPI, so there could be no presumption of the upward bias alleged by the ONS). That left only the test of time-reversal, which is failed by the Carli, and Diewert, in his report, relied on that, when coupled with a preference for a symmetric higher-level index rather than the Lowe index used in both the RPI and CPI. However, in his subsequent comments (*Answers to Questions Arising from the RPI Consultation*, 28 December 2012), he made clear that an upward bias from the Carli's lack of a time-reversal property would result only from the use of a chained Carli index in the presence of price-bouncing. It

is now clear, but might not have been evident to Diewert at the time, that the RPI uses only the direct Carli, and therefore this last reason to prefer the Jevons also drops away.<sup>2</sup>

Thus, although his report recommended switching the RPI to use of the Jevons, Diewert's report and comments effectively destroyed any statistical reason to do so. Inconveniently for the ONS, Diewert's report also added additional recommendations, including that the RPI (and by extension the CPI) should exclude a wide range of fashion goods, while his *Comments* saw "definite merit" in the Royal Statistical Society's recommendation of waiting a year before making any changes to the RPI, in order to allow the clothing price pilot research to run its course and facilitate the introduction of more homogeneous elementary aggregates. This accounts for the fact that, while the UKSA consultation report noted Diewert's "headline" recommendation for a switch to the Jevons, they have not relied on him in their campaign to downgrade the RPI.

**D Elliott**, R O'Neill, J Ralph and R Sanderson, wrote a paper, *Stochastic and Sampling Approaches to the Choice of Elementary Aggregate Formula*, ONS Discussion Paper, October 2012, as part of the ONS research department's investigation of the formula effect difference between the RPI and CPI. The first half of the paper took a stochastic approach, looking at the distribution of price relatives for a very large data set covering alcoholic beverages, to see if they had a normal or log-normal distribution, on the premise that a normal distribution would indicate a Carli as an unbiased estimator and a log-normal distribution would indicate a Jevons. They failed to find a conclusive pattern, but in any case, the whole exercise was moot because, as **Greenlees** had already observed in this context in 2001, the Carli gives an unbiased estimate of the population mean of price relatives regardless of their distribution, whereas, by Jensen's inequality, the Jevons yields a downward-biased estimate.

The second part of the paper by Elliott et al. looked at which of the unweighted elementary indices provided a better estimate for various weighted target population indices in each of the nineteen items in their alcohol data set, which contained matched price and quantity information. Where a Laspeyres index was the target, both the Carli and Jevons had a negative bias, but that of the Carli was less: the Carli estimate also had a lower mean-squared-error (MSE) indicating a lower variance of the estimator about the target. The Carli was, therefore unambiguously preferred (c.f. the similar results reported by **Mehrhoff**). Where the symmetric, Fisher, index was the target, the Carli and Jevons had roughly equal and opposite biases, with the Carli overestimating the target by slightly less than the Jevons underestimated

---

<sup>2</sup> Diewert still makes a theoretical case for the Jevons in his *Answers to Questions* paper, when he expresses a preference for a symmetric target index like the Fisher index, for which the Jevons could be a good estimator. He recognises that the empirical conditions for this are restrictive but asserts that in the absence of empirical estimates it is reasonable to assume a consumer price elasticity of (minus) unity, when the estimate will be exact. This is the same assertion made by Peter Levell (whom he quotes) but, unlike him, Diewert recognises in Section 9 of his paper that one also has to specify that all fashion goods are excluded from the index. These are exactly the results from the simulations in the paper by **Winton** et al, who find that the Jevons is the best elementary aggregation index for estimating the Fisher in about half the cases when the elasticity of substitution is relatively high – between 0.5 and 1.5 – if all fashion goods are excluded, but is the best estimator in only a tiny proportion of cases if fashion goods are included. This is also the theoretically tenable Economic Approach case for the Jevons explained right at the beginning of this paper.

it: the MSE in this case was not reported but the authors concluded that for this data there was no clear preference for either estimator.

**David Fenwick** was Director, Consumer Prices and General Inflation Division at the ONS until 2008. His paper, “The impact of choice of base month and other factors on the relative performance of different formulae used for aggregation of Consumer Price Index data at an elementary aggregate level, in *proc. Ottawa Group fifth meeting*, 1999,” was the first systematic investigation of the difference between the RPI and HICP (renamed CPI in 2003) indices and set the tone of the subsequent debate. Fenwick attributed the large difference between them – compared to similar formula-effect differences in other countries – to sampling procedures and the definition of elementary aggregates, being conscious of the recent, 1996, broadening of the UK’s definitions and revisions to sampling methodology.

This paper does not explicitly touch on the reason why Fenwick did not even consider bowing to Treasury pressure to lower the RPI by switching to the geometric mean elementary aggregation formula used in the HICP, but his exposition of the different formulae in terms of differing assumptions about consumer behaviour indicates his thinking. He is more explicit in his later paper, *System of Price Indices and Supporting Frameworks*, joint ECE/ILO meeting, ECE/CES/GE.22/2006/9 21 February 2006 in which he takes as his starting point, “Price indices serve many different purposes and their definition, coverage and construction depends on the precise use for which they are being constructed,” and, among other observations, goes on to make a sharp theoretical distinction between a cost-of-goods index (such as the RPI) and a cost-of-living index, which measures a particular level of welfare. See also **Baxter and O’Donoghue**.

**Professor Robert Gordon** is Professor of Economics at Northwestern University. His 2004 paper, *Apparel Prices 1914-93 and the Hulten/Bruegel Paradox* (reprinted in W Erwin Diewert et al (eds) *Price Index Concepts and Measurement*, Chicago, 2009, p 85-128) was an early and path-breaking study of the peculiarities of clothing prices, comparing the US CPI apparel series (i) with a matched model index of many different types of apparel, based on 10,000 exact comparisons taken from the Sears catalogues of 1914-93 and (ii) with a hedonic price index for women’s dresses over the period 1914-88 based on another 6,500 comparisons. The main conclusion of the paper is based on the hedonic study of women’s dress prices, which exhibits a price index increase of many orders of magnitude faster than either the Sears matched-model index developed from the same source data or than the CPI. Series. Thus, the CPI for apparel incorporates a significant downward bias, at least for the 50 years since 1945 and perhaps over a longer period, at odds with the perception in the USA at the time that the CPI was overstating inflation.

Gordon attributes this result to the fact that price changes in apparel, particularly in women’s apparel, coincide with model (i.e. new fashion) changes, something that the matched-model and CPI series do not take account of but that is captured by the hedonic pricing model. This anticipated the belated discovery in 2009 in the UK of a downward bias in clothing prices and also indicated a sensible way of adjusting for it (similar but opposite to the hedonic pricing applied to computer prices).



Gordon relates this result to a paradox that had been pointed out by Charles Hulten, namely that projecting recent estimates of CPI bias backwards in time leads to ridiculous results. Thus, the estimates of CPI upward bias in the Boskin report, if projected backwards would result in impossibly low living standards in colonial America and to the starvation of the well-fed peasants depicted in Breugel's paintings.

**John Greenlees** is a senior researcher in the Bureau of Labor Statistics in the US Department of Labor. In that capacity, he has loyally defended the composition of the US CPI, e.g. in his paper, "The BLS Response to the Boskin Commission Report", *International Productivity Monitor*, No 12, 2006, where, however, he is explicit that the adoption of the geometric mean (Jevons) aggregation formula "was justified explicitly on the basis that it would reflect consumer response to relative price changes." This contrasts with the ONS unwillingness to admit that the use of the Jevons necessarily implies an assumption about consumer behaviour. His article is also, to some extent a re-writing of history, since the BLS, in its initial reaction to the Boskin Commission Report, before the political decision was taken to accept it, was far more critical of its analysis and recommendations: see Bureau of Labor Statistics U.S. Department of Labor "Measurement issues in the consumer price index". *Statistical Journal of the UN Economic Commission for Europe*, 1998, Vol. 15, Issue 1, which, in Section 3, identifies important weaknesses in the Commission's case for adoption of the Jevons formula: "As described in Appendix A, the Commission fails to mention several strong assumptions about the distribution of price changes that they implicitly use when claiming that the geometric means index is unbiased, or to note that, under these same assumptions, the Laspeyres formula currently used by the BLS also is unbiased." This was an important early statement in view of later attempts by, e.g. Diewert to argue that the Geometric (Jevons) index provides an unbiased estimate of a symmetric aggregate index like the Fisher index.

Greenlees, in an early paper investigating possible formula effect bias at the aggregate level, (*Random Errors and Superlative Indexes*, BLS Working Paper 343 March 2001) established, using Jensen's inequality, that the Jevons index introduce a downwards bias when estimating the expected change in prices, regardless of the distribution of price relatives, whereas the Carli did not, which reduced the difference which the BLS had expected to see when using a Laspeyres-type or Tornqvist aggregation index at the upper level. Since, in the UK, both RPI and CPI use a Laspeyres-type index at the upper level, his overall result was not directly applicable here. But the confirmation that the Jevons estimate of the expected value of price relatives is downwards-biased regardless of the distribution of price relatives confirmed the superiority of the Carli for lower-level aggregation relative to the Jevons under the stochastic approach (see **Elliott** et al).

More recently, and also relevant to the RPI/CPI debate, Greenlees and Robert McClelland anticipated the UK's 2017 "Strappy Tops" controversy in their paper, "RGEKS Superlative Consumer Price Indexes for Apparel Using U.S. Scanner Data", Presentation at the 31<sup>st</sup> General Conference of the International Association for Research in Income and Wealth, St. Gallen, Switzerland, August 27 2010. Their study used a dataset that contained every transaction for a single apparel good – Misses' tops – of a large retail chain in a major U.S. metropolitan area. The data confirmed their expectation that the usual matched-model assumption would yield implausible drops in price indexes for misses' tops during their study period. They therefore investigated various alternative approaches to solving this problem. They conclude, "The

operational consumer price indexes for apparel indexes published by national statistical agencies rely on careful matching of identical items, usually combined with special procedures to deal with short product life spans. The BLS, for example, uses hedonic regressions to facilitate comparisons of the prices of exiting and entering items. None of the approaches we tested in this paper demonstrated any superiority to those statistical agency procedures, despite our large and detailed data set.”

This result might be slightly disappointing to Greenlees but should be encouraging for the ONS, whose ten-year failure to develop a sensible clothing index might be solved by a visit to the BLS, for example, to study their hedonic regression adjustments to matched-model comparisons.

**Peter Hill**, 1929-2017, was Head of the Statistics Division of the OECD from 1980 to 1994. His paper, “Recent Developments in Index Number Theory and Practice”, *OECD Economic Studies*, vol 10, 1988, p.124-148, is notable for its clarity of exposition and its recognition that no index number is perfect but must be used in appropriate circumstances. He notes Eichorn and Voeller’s proof that there is no possible index that satisfies all the desirable axioms and that in some circumstances transitivity must be abandoned – as Fisher himself finally abandoned it – if the product test is to be satisfied.

Hill’s paper is also notable for its emphasis on chaining as a relatively new and, in the right circumstances, desirable development. (In 1987 only four countries, among them the UK, used a chain index with annual links, and US still does not do so). Hill concludes that, “in general, chaining will tend to minimise the traditional index number problem by greatly reducing the spread between alternative formulae, provided that changes in prices and quantities proceed in a fairly smooth or regular fashion and that the chaining is done frequently.” He notes the possibility of price “bouncing,” e.g. as a result of seasonal variations with monthly chaining, which would tend to cause the Laspeyres and Paasche indices to diverge, in which case chaining is to be avoided.

**Dr Gareth Jones** has been one of the earliest and most consistent defenders of the RPI. In a letter of 5 July 2011 to Sir Michael Scholar, he made the point that under the GM formula, the degree and direction of brand substitution depend only on the percentage price changes of different brands and that price levels (and utility) are not relevant, which he characterised as an inherently implausible description of consumer behaviour. Since then he has been active in exposing the fallacies of ONS attacks on the RPI, mostly in postings and papers deposited in the Royal Statistical Society’s online User Net library. Notable examples have been:

(i) a paper deposited on 12 April 2013, in which he used the published ONS figures for the formula effect difference between the CPI and RPI to estimate what part of the formula effect was due to price bouncing in the RPI. He found that, contrary to ONS claims that bouncing accounted for most of the formula effect, in fact, over fifteen years, the average difference attributable to bouncing over fifteen years was 0.02 percentage points, within the rounding error of the RPI and CPI.

(ii) a paper *Chain Drift Synopsis*, deposited in the Royal Statistical Society User Net Library on 6 February 2015, completely refuted a paper prepared by the ONS (*Comparing Class level*

*Chain Drift for Different Elementary Aggregate Formulae Using Locally Collected CPI Data*) prepared by the ONS for the Johnson review, which had purported to show high chain drift from using a Carli index for elementary aggregation. Jones showed in detail that the classes in their analysis with higher chain drift in fact used the Dutot index in the RPI and that in any case their analysis used monthly chaining over three summer months where seasonal fluctuations were significant, whereas the RPI uses annual chaining. He therefore stuck to his previous estimate of there being no significant formula-effect difference on chain-drift in the RPI.

Jones has also provided an annual exercise in Benchmarking Clothing Inflation, where he uses ONS data to draw out the implications for e.g. the implied volume of clothing purchases, which provides a reality check on claims about the plausibility of clothing price indices.

Gareth Jones' current view is that the ideal uprating index would be a development of the Household Inflation Index (aka Household Costs Index) advocated by Jill **Leyland** and John Astin, which has many of the features of the RPI. The chances are almost nil that the ONS will produce a version of such an index to replace the RPI, because, among other reasons, they have for the last nine years argued for just a single consumer price index, and so, in the meantime, Jones would regard the RPI as the best uprating index available.

**Jill Leyland** has been Vice President of the Royal Statistical Society (RSS) and the Chair of its National Statistics Advisory Group and is currently on the National Statistician's Advisory Panel on Consumer Prices (Stakeholder). In those positions she has maintained an even-handed approach to the construction of consumer price indices but has been forthright in defending statistical rigour. In particular she has always made the case that the choice of index depends on the use to which it will be put and is unlikely to be satisfied by a single index. She has consistently pointed out that the only current problem with the RPI is the need to correct the problems in clothing price sampling introduced in 2010 and thereafter to remove the "freeze" on incremental changes introduced in 2013. And she has vigorously countered the ONS assertion that the chained Carli introduces an upward bias to the RPI by pointing out that the RPI uses only the direct and not the chained Carli.

In her evidence to the House of Lords Economic Affairs Committee, Leyland was explicit that there is nothing really wrong with the RPI, other than the clothing issue, which can be corrected, and that, "In addition there is nothing currently which could replace it." She mentions the Household Costs Indices, (which she was instrumental in developing) as a possible ultimate replacement for the RPI, but points out that they would take years of development and would then need to have proved themselves fully satisfactory.

**Professor David Mayston** is Professor of Public Sector Economics, Finance and Accountancy at the University of York. In 2012 he provided a detailed statistical response to the ONS Consultation on Improving RPI. Like many independent statisticians, he explained how the Jevons formula is a downwards-biased estimator if used in the first stage of the construction of an upper-level index designed to estimate the average price change in the average basket of goods and services consumed. In addition, from the public welfare point of view, he pointed out that, "Once the variance of price relatives is not simply zero, the impact of inflation will not be uniform across individual products and there will be important distributional impacts of

inflation across individual groups of consumers. As noted above, under reasonable assumptions the Carli index can indeed take account of such distributional impacts better than the alternatives suggested in the Consultation document.” Given these distributional effects and therefore the certainty that no changes to the RPI could be an unambiguous improvement, Mayston questioned the whole basis of the consultation, concluding that the ONS was failing in its statutory duty ‘to promote and safeguard the production and publication of official statistics that serve the public good.’

So far as can be established, Professor Mayston has not been active in this area since then.

**Dr Jens Mehrhoff** works in the Statistics Department and Research Centre of the Deutsche Bundesbank. He has recently been appointed to the ONS’ CPAC-Technical. His 2010 article, “Aggregate Indices and Their Corresponding Elementary Indices,” *Journal of Economics and Statistics*, No 230, 709-725, uses a “generalised means” approach to investigate theoretical conditions under which a particular index formula at the elementary level exactly equals the desired aggregate index, with the solution depending only on the item’s price elasticity. He then presents an empirical application using German foreign trade statistics, finding that different elementary aggregates should be used for different product groups, depending on their average elasticity: for example if a Laspeyres is the target aggregate index – as he argues it should be for a price index – the Carli should be used for 70% of exports and 72% of imports and the Jevons for 14% of exports and 17% of imports (and other elementary indices for the rest).

Mehrhoff applied a similar methodology, which he calls the consistency in aggregation approach, to the very large matched price and quantity data set covering the alcoholic drinks sector in the UK that was obtained by the ONS in 2011 for its investigation of the Economic Approach. The results were presented in his paper, *The CIA (Consistency in aggregation) approach, a new economic approach to elementary indices*, Ottawa Group paper, 2 May 2013. He found that the Carli had a smaller bias when estimating the Laspeyres than did the Jevons, although both tended to underestimate it. For every one of the nineteen sub-classes the Jevons has a more negative bias than the Carli and in all but one case a greater absolute bias. Mehrhoff concludes that “In particular, the Carli index performs remarkably well at the elementary level of a Laspeyres price index, questioning the argument of its “upward bias” – in fact, it is the Jevons index that has a downward bias.” This is a similar result to that obtained by **Winton et al**, using a different methodology.

**Jim O’Donoghue** is a statistician at the ONS. The paper that he wrote with Colin Wilkie (J O’Donoghue and C. Wilkie, “Harmonised Indices of Consumer Prices”, *Economic Trends*, 532, 1998), was a straightforward description of the newly introduced UK HICP, but is notable for its unequivocal statement of the ONS position at the time that, “The HICP and the RPI differ in their purpose and construction. The RPI is the best indicator of consumer price inflation in the United Kingdom.” This echoes a similar statement made by **Fenwick** in Fenwick, D (1997): *The Boskin Report from a United Kingdom Perspective; Bias in the CPI: Experiences from five OECD countries*; Statistics Canada, 1997; No. 62 F0014MPB, p45 -52. The Appendix to the paper by O’Donoghue and Wilkie also indicates a reason for preferring the RPI when it demonstrates that the arithmetic aggregation indices used in the RPI “have the

effect of giving implicit fixed weights to each price quote,” whereas the geometric mean index used in the HICP “is equivalent to assuming that expenditure shares remain constant, so that if one price doubles while the others stay the same, the quantity purchased of the former will halve; and thus the implicit weights will vary through time.”

**Professor Denise Osborn** is currently Secretary-General of the Royal Economic Society and Emeritus Professor of Econometrics at the University of Manchester, where she was a Professor of Econometrics from 1992 to 2015. She was asked by the ONS to comment on the responses with a statistical content that they had received in response to their “Consultation on Improving RPI.” In her *Comments on Responses to ONS Consultation on Improving RPI*, 19 December 2012, Osborn provides a succinct summary, drawing on but also critiquing the arguments, among others, of John Wood, Peter Levell and David Mayston, and she concludes that the Statistical Approach indicates a downwards bias of the Jevons; that the Test Approach is inconclusive; and that the Economic Approach justifies the use of a Carli index for lower-level aggregation combined with a Lowe index for upper-level aggregation, as in the RPI, and does not justify using the Jevons for lower-level aggregation and a Lowe for upper-level aggregation.

Given such a damning assessment of the Jevons, Osborn’s overall conclusion is rather mild in echoing what she calls a ‘quite measured’ response by the Royal Statistical Society, when she concludes that: “The consultation responses made available to me provide a number of coherent arguments in support of retention of Carli aggregation at the elementary level within the [RPI], at least until further work is undertaken within ONS.”

So far as I know, Professor Osborn has not been active in this area since then.

**M. Sabag and Y. Finkel**, are senior statisticians in the Israeli Central Bureau of Statistics. Their paper, “The Israeli Consumer price index: alternative computing methods for an ‘Item Price Index’”, *Statistical Journal of the United Nations ECE*, vol 11, 1994, pp. 95-118, is notable for its description of an ad hoc method used by the Central Bureau of Statistics to correct for the inherent downward bias in any clothing index using the usual matched-model sampling. They apply this to the Carli clothing indices which were, and still are, used in Israel’s CPI.

**Ria Sanderson** is a statistician at the ONS and does not express an opinion on the RPI. Her paper, “Do Consumers Substitute Between Clothing Brands? Estimating the Elasticity of Substitution for a Selection of Clothing Products,” *Survey Methodology Bulletin*, No 72, Spring 2014, pp 57 -72, found that estimates for the elasticity of substitution of the clothing items peaked at around zero, with a large proportion of estimates showing a positive correlation between price and quantity. Her conclusion was that there was no evidence of substitution between brands and is powerful empirical evidence that, if clothing sampling were satisfactory, there would no case for using the Jevons in constructing clothing item sub-indices. It also is evidence for the conjecture in **Jones’** 2011 letter to Sir Michael Scholar about the lack of brand substitution.

**Joseph Winton**, Robert O'Neill and Duncan Elliott were working at the ONS when their important working paper, *Elementary Aggregate Indices and Lower Level Substitution Bias*, ONS, March 2012 was written and circulated to a meeting of the Consumer Prices Advisory Committee as Annex A to CPAC(12)15. A shortened version was published by the authors in the *Statistical Journal of the International Association for Official Statistics*, vol 29, no 1, 2013, pp. 11-20, with the usual disclaimer that "the analysis and views in this paper are those of the authors and not necessarily those of the Office for National Statistics." So far as can be established, it is the only paper on consumer price indices research conducted in the ONS that has been published in a peer-reviewed journal. Elliott is still at the ONS. Dr O'Neill left the ONS for academia in 2013 and is currently a lecturer at the University of Manchester. Winton was at the ONS until recently and currently seems to be working in New Zealand (Reserve Bank/Statistics New Zealand).

Their paper was significant, since it contradicted what had hitherto been the official Economic Approach justification for a switch from the RPI to the CPI. The main conclusion is worth quoting in full: "As mentioned above it is commonly asserted in the index numbers literature that where consumers are thought to be substituting between products as prices change the Jevons index is the most appropriate elementary aggregate for use in constructing price indices. Our empirical investigation identifies several cases where elasticity is estimated to be greater than 1 but the Carli is better able to approximate the Fisher index than the Jevons. This is counter to the assertion that Jevons is the most appropriate formula when substitution behaviour is present and is a result worthy of further investigation."