

# Methods explained

Methods explained is a quarterly series of short articles explaining statistical issues and methodologies relevant to ONS and other data in a simple, non-technical way. As well as defining the topic areas, the notes explain when, why and how these methodologies are used. Where relevant, the reader is also pointed to further sources of information.

## Index numbers

Peter Goodridge

Office for National Statistics

### SUMMARY

Many of the statistics produced by the Office for National Statistics, particularly economic statistics, are published in the form of indices. However, there are a number of different forms of indices and this article attempts to explain the subtle differences in the methodologies used to construct them, and also factors that feed into the choice of which type of index to use. Hypothetical examples are provided to illustrate the effects of using different index forms.

Index numbers are a common and convenient way of expressing time-series variables. They are useful because the movement from one period to the next is expressed as the percentage of the value in the base period, the choice of the base period depending on the type of index used. The concept of the base period is discussed later. This article summarises the common types of indices used, particularly those used by the Office for National Statistics (ONS), and the differences between them. Some common variables expressed in index form are capital, labour, output, prices, earnings and productivity.

In **Table 1** below is a simple example of an index. Say we are interested in a variable X and its evolution over time. Then the index is simply the growth in X multiplied by the value of the index in the previous period. That is:

$$I_{(t)} = \left( \frac{X_{(t)}}{X_{(t-1)}} \right) I_{(t-1)} \quad (1)$$

or, for the year 2003, the index is equal to  $(37/34) * 113.33 = 123.33$ .

The informed reader will notice that the above index has been chained. The difference between chained and non-chained indices is discussed below.

**Table 1**  
**Simple example**

Time (t)	X	Index (I)
2000	30	100
2001	33	110
2002	34	113
2003	37	123
2004	35	117
2005	36	120
2006	40	133

It should be noted that, when constructing an index, the variables in question need to be in the same unit, or converted into the same unit. Also the index in the above example was made up of just one input variable. If the variables used to create the index are in different categories, then these will need to be weighted together to form the aggregate index, based on their relative importance. Weighting is discussed below in reference to Laspeyres indices.

### Laspeyres

One of the most common forms of indices used at the ONS is the Laspeyres. The main feature of the Laspeyres is that the weights used are taken from the base period. For example, if we consider a price index with a base t-1, then the prices will be aggregated for all periods using weights from that period. However, if the index is chain-linked, then the weight will be taken from the previous period. Chain-linking simply means updating the weights so that, for each period, the base used is the weight from the previous period. ONS output (GDP or GVA) measures are now calculated as chain-linked Laspeyres indices and are referred to as chained volume measures (CVM). For a fuller discussion of chain-linking in relation to National Accounts output measures, see Robjohns (2006).

Below are two simple examples of a weighted price index consisting of two goods, X and Y. The weights used are the quantities of the goods consumed and so represent their relative importance. The price is multiplied by the quantity, to obtain expenditure, and the change is expressed as a percentage of the expenditure in the base period. Equation 2 below shows a standard Laspeyres price index; however, the principle is true for any type of index, not just prices, that is the weight is taken from the base period. For instance, for a quantity index, the quantity in the numerator term would be updated rather than the price, and the denominator would remain the same.

$$I_{i,0}^L = \left( \frac{\sum_i P_{i,t} Q_{i,0}}{\sum_i P_{i,0} Q_{i,0}} \right) 100 \quad (2)$$

**Table 2** presents a Laspeyres index using a base year of 2000, and **Table 3** shows the effect of chain-linking the index.

As can be seen above, chain-linking has the strongest impact on the index when there is a relative shift in composition of the variable. In this case, there has been a relative shift toward consumption of

**Table 2**  
**Example of a weighted Laspeyres index**

Time (t)	P(X)	P(Y)	Q(X)	Q(Y)	Expenditure		Index: base (2000)
					Expenditure: $P_t * Q_0$	in base period: $P_0 * Q_0$	
2000	10	5	30	60			100
2001	14	7	32	60	840	600	140
2002	19	6	38	55	930	600	155
2003	28	7	43	52	1,260	600	210
2004	40	8	47	50	1,680	600	280
2005	44	9	53	50	1,860	600	310
2006	56	11	60	45	2,340	600	390

**Table 3**  
**Example of a weighted chain-linked Laspeyres index**

Time (t)	P(X)	P(Y)	Q(X)	Q(Y)	Expenditure		Index: chain-linked
					Expenditure: $P_t * Q_{t-1}$	in base period: $P_{t-1} * Q_{t-1}$	
2000	10	5	30	60			100
2001	14	7	32	60	840	600	140
2002	19	6	38	55	968	868	156
2003	28	7	43	52	1,449	1,052	215
2004	40	8	47	50	2,136	1,568	293
2005	44	9	53	50	2,518	2,280	324
2006	56	11	60	45	3,518	2,782	409

good X, the more expensive good, causing the chain-linked index to rise by a greater amount than the standard Laspeyres, since the price of good X has risen more than that of good Y. This is obviously relevant to a number of areas, for instance the changing industrial composition of the economy. Therefore, chain-linking is a way of improving the methodology because it takes account of more information and so provides a more accurate measure. However, the ability to chain-link does depend on the timeliness of the data used for the weights.

The first main benefit of chain-linking is that new items can be added to the ‘basket’ every year. If the index is non-chained, new items can only be added to the base year. The second benefit is that by chaining the series, the comparison is with the previous year, rather than the base year. This is obviously relevant when most of the interest is in the annual change such as with prices, output and most economic indicators.

Another benefit of chain-linking is that it removes ‘substitution bias’. This is a problem that is encountered when there are large shifts in both the weight and in the actual variable that is being indexed. This was encountered in the US with data for computer hardware, but also applies more generally to all technology goods. Prices for such goods have reduced dramatically, causing the quantities consumed to increase. However, when the output index is unchained, it means that the index is being weighted on prices from the base year, when current prices are far removed from those in the base year, that is, the weights used are too high. Therefore, when the index is rebased, which in the US was every five years, the rebasing has the effect of reducing the index, causing large revisions. Chain-linking removes this problem.

The retail prices index (RPI) is an annually chain-linked Laspeyres index.<sup>1</sup> Other data produced as Laspeyres indices are the quality-adjusted labour input measure (QALI), the volume index of capital

services (VICs), both produced by the ONS, and also the house price index, produced by the Department for Communities and Local Government (DCLG).

**Paasche**

The main difference between Laspeyres and Paasche indices is that, with the Paasche, the weights are taken from the current period. The formula for the Paasche is given below in equation 3:

$$I^P_{t,0} = \left( \frac{\sum_i P_{i,t} Q_{i,t}}{\sum_i P_{i,0} Q_{i,t}} \right) 100 \tag{3}$$

So, whereas the Laspeyres calculates what expenditure in the current period would be if the quantity consumed were the same as in the base period (a pure price effect), the Paasche calculates the expenditure needed to buy current year quantities, and is expressed as a percentage of what the expenditure would have been in the base period if the quantity consumed had been at current levels. Examples of a Paasche price index, both unchained and chained, are shown below in **Table 4** and **Table 5** using the same data as in previous examples.

Unlike the Laspeyres, in the case of the Paasche, chain-linking has the effect of reducing the index. This is because growth is not calculated as a percentage of expenditure in the base period but instead is backward-looking. Therefore, the effect of the high increases in the price of X, and substitution from Y to X, is less pronounced when the index is chained together. Put another way, there is an incremental weighting effect, that is, the shift in the weights (quantities) is more pronounced between the starting and

**Table 4**  
**Example of a weighted Paasche index**

Time (t)	P(X)	P(Y)	Q(X)	Q(Y)	Current		Index: base (2000)
					Current expenditure: $P_t * Q_t$	expenditure using base year prices: $P_0 * Q_t$	
2000	10	5	30	60			100
2001	14	7	32	60	868	620	140
2002	19	6	38	55	1,052	655	161
2003	28	7	43	52	1,568	690	227
2004	40	8	47	50	2,280	720	317
2005	44	9	53	50	2,782	780	357
2006	56	11	60	45	3,855	825	467

**Table 5**  
**Example of a weighted chain-linked Paasche index**

Time (t)	P(X)	P(Y)	Q(X)	Q(Y)	Current		Index: chain-linked
					Current expenditure: $P_t * Q_t$	expenditure using base year prices: $P_{t-1} * Q_t$	
2000	10	5	30	60			100
2001	14	7	32	60	868	620	140
2002	19	6	38	55	1,052	917	161
2003	28	7	43	52	1,568	1,129	223
2004	40	8	47	50	2,280	1,666	305
2005	44	9	53	50	2,782	2,520	337
2006	56	11	60	45	3,855	3,045	427

current period than it is between the previous and current period, causing the index to increase more slowly.

The choice of whether to use a Laspeyres or Paasche is fairly arbitrary. The decision will probably make very little difference to the final index unless there has been a substantial change to the weighting of the variable, as there has been in the examples above. Therefore, the decision tends to be based on practicalities. To construct a Paasche index, as explained above, weights for the actual year of the series, or current weights, are required. Such data are often unavailable. Therefore, the Laspeyres tends to be used, as weights from a previous period are more readily available. Also, using a Paasche index means that the denominator changes every year, so different years can only be compared with the base year and not with each other.

An example of a Paasche index produced by ONS is the implied GDP deflator, which is also chain-linked.

### Fisher

The Fisher index, or the Fisher ideal index, is a form of compromise between the Laspeyres and Paasche. Its formula is a geometric mean of the Laspeyres and Paasche, as shown in equations 4 and 5. In general, the Laspeyres is always greater than or equal to the Fisher, and the Paasche is always less than or equal to the Fisher (6):

$$I^F_{t,0} = \sqrt{\left(\frac{\sum_i P_{it} Q_{i,t}}{\sum_i P_{i0} Q_{i,t}}\right) \cdot \left(\frac{\sum_i P_{it} Q_{i,0}}{\sum_i P_{i0} Q_{i,0}}\right)} 100 \tag{4}$$

$$I^F_{t,0} = \sqrt{I^L_{t,0} \cdot I^P_{t,0}} \tag{5}$$

$$\text{Laspeyres} \geq \text{Fisher} \geq \text{Paasche} \tag{6}$$

It should be noted that the reason the inequality given in equation 6 does not hold in the above examples is that the numbers used for price and quantity suggest a perverse demand relationship. That is, in general, if there is an increase in the price ratio ( $P_X/P_Y$ ) of the two goods, we would expect the quantity ratio ( $Q_X/Q_Y$ ) to decrease. This is not the case in the hypothetical numerical examples given.

In general, the Fisher is used when prices and quantities in the base and observation periods are substantially different. Examples of a Fisher index, both unchained and chained, are provided below in **Table 6** and **Table 7**, again using the same data as used previously.

**Table 6**  
**Example of a weighted Fisher index**

Time (t)	Laspeyres	Paasche	Product	Fisher: square root of product
2000	100	100	10,000	100
2001	140	140	19,600	140
2002	155	161	24,895	158
2003	210	227	47,722	218
2004	280	317	88,667	298
2005	310	357	110,567	333
2006	390	467	182,236	427

**Table 7**  
**Example of a weighted chain-linked Fisher index**

Time (t)	Laspeyres	Paasche	Product	Fisher: square root of product
2000	100	100	10,000	100
2001	140	140	19,600	140
2002	156	161	25,076	158
2003	215	223	47,969	219
2004	293	305	89,429	299
2005	324	337	109,032	330
2006	409	427	174,555	418

The Fisher index is termed ‘ideal’ because it gets around the practical problem of ‘time reversal’. This is particularly relevant when the data are seasonal and was encountered when attempts were made to construct a quarterly Laspeyres version of QALI. Basically, what happens is that when growth in the variable increases, but in a subsequent period decreases, the index fails to decrease all the way back and remains at a higher level. Superlative<sup>2</sup> indices, such as the Fisher, and also the Törnqvist, which is discussed in the next section, overcome this problem and this is the reason why they are sometimes preferred.

Another convenient property is that multiplying a Fisher price and Fisher quantity index produces a Fisher expenditure index.

### Törnqvist

The main feature of the Törnqvist index is that the weight used is an average of the weight in the current and base period. Therefore, like the Paasche, it tends to be used on historic data sets, as current period information is needed to weight the series. Another difference is that it is calculated geometrically rather than arithmetically. The formula for a Törnqvist index is shown below in equations 7 and 8, and examples of both unchained and chained Törnqvist indices are shown below in **Table 8** and **Table 9** (the price and quantity data are the same as those used previously, but are omitted to save space).

$$I^T_{t,0} = \prod_i \left( \frac{P_{i,t}}{P_{i,0}} \right)^{\frac{(w_{i,t} + w_{i,0})}{2}} \tag{7}$$

where:

$$w_{i,t} = \frac{P_{i,t} Q_{i,t}}{\sum_i P_{i,t} Q_{i,t}} \tag{8}$$

For ease of composition, the Törnqvist is often written as:

$$I^T_{t,0} = \exp \left[ \sum_i \left( \frac{w_{i,t} + w_{i,0}}{2} \right) \ln \left( \frac{P_{i,t}}{P_{i,0}} \right) \right] \tag{9}$$

**Table 8**  
**Example of a weighted Törnqvist index**

Time (t)	Relative change in P(X)	Share of Good X	Share of X averaged over both periods	X <sup>share</sup>	Relative change in P(Y)	Share of Good Y	Share of Y averaged over both periods	Y <sup>share</sup>	(X <sup>share</sup> )*(Y <sup>share</sup> )	Index: base (2000)
2000	1.0	0.50	0.50	1.00	1.00	0.50	0.50	1.00	1.00	100
2001	1.4	0.52	0.51	1.19	1.40	0.48	0.49	1.18	1.40	140
2002	1.9	0.69	0.59	1.46	1.20	0.31	0.41	1.08	1.58	158
2003	2.8	0.77	0.63	1.92	1.40	0.23	0.37	1.13	2.17	217
2004	4.0	0.82	0.66	2.50	1.60	0.18	0.34	1.17	2.94	294
2005	4.4	0.84	0.67	2.69	1.80	0.16	0.33	1.21	3.27	327
2006	5.6	0.87	0.69	3.26	2.20	0.13	0.31	1.28	4.18	418

**Table 9**  
**Example of a weighted chain-linked Törnqvist index**

Time (t)	Relative change in P(X)	Share of Good X	Share of X averaged over both periods	X <sup>share</sup>	Relative change in P(Y)	Share of Good Y	Share of Y averaged over both periods	Y <sup>share</sup>	(X <sup>share</sup> )*(Y <sup>share</sup> )	Index: chain-linked
2000	1.00	0.50	0.50	1.00	1.00	0.50	0.25	1.00	1.00	100
2001	1.40	0.52	0.51	1.19	1.40	0.48	0.49	1.18	1.40	140
2002	1.36	0.69	0.60	1.20	0.86	0.31	0.40	0.94	1.13	158
2003	1.47	0.77	0.73	1.33	1.17	0.23	0.27	1.04	1.38	219
2004	1.43	0.82	0.80	1.33	1.14	0.18	0.20	1.03	1.37	299
2005	1.10	0.84	0.83	1.08	1.13	0.16	0.17	1.02	1.10	330
2006	1.27	0.87	0.85	1.23	1.22	0.13	0.15	1.03	1.27	417

The Törnqvist has a number of useful features. As well as removing the problem of 'time reversal', it can also be used to show the contribution of each component to aggregate growth because it is log-linear. Therefore, the contributions can be broken down into additive form. Another feature is they correspond to translog production and cost functions and are hence widely used in empirical work. For instance, using a standard Cobb-Douglas production function where:

$$Y = K^\alpha L^{(1-\alpha)} \quad (10)$$

and:

Y = output,

K = capital,

L = labour

$\alpha$  and  $(1-\alpha)$  = income shares of K and L, respectively

In general, the income shares are assumed to be constant. However, the use of Törnqvist indices means these are more flexible and can change according to the returns to labour or capital.

## Conclusion

There are, therefore, various forms of indices and their construction affects the result and how they should be interpreted. The choice of which to use tends to depend on what is being compared over time, the timeliness of the available data, how much flexibility is needed and what its final use is, for example, is it going to be used as an input into another piece of analysis which requires an index of a certain form.

## Notes

- 1 Technically the RPI is not a true Laspeyres index, but it is a very close approximation. For further detail and reasons consult the Retail Price Index Technical Manual, available at [www.statistics.gov.uk/statbase/Product.asp?vlnk=2328](http://www.statistics.gov.uk/statbase/Product.asp?vlnk=2328)
- 2 A superlative index is one that uses more information in its construction than a base index, and is more flexible. The OECD definition is as follows: 'Superlative indices are price or quantity indices that are 'exact' for a flexible aggregator. A flexible aggregator is a second-order approximation to an arbitrary production, cost, utility or distance function. Exactness implies that a particular index number can be directly derived from a specific flexible aggregator.'

## REFERENCES

- Caswell F (1995) *Success in Statistics* Third Edition, Coventry University
- Harper M (1991) *Statistics* Sixth Edition (1991), Longman Group UK Limited
- Robjohns J (2006) 'Methodology Notes: Annual chain-linking' *Economic Trends* 630, pp 25–8 and at [www.statistics.gov.uk/cc/article.asp?ID=1554](http://www.statistics.gov.uk/cc/article.asp?ID=1554)