Anna Janicka

Mathematical Statistics 2020/2021
Additional Lecture: Index Numbers

Index numbers are meant to illustrate changes in time of the value of a certain characteristic or characteristics. If we look at one characteristic only, we will talk about simple indices; if we look at more than one, we will talk about aggregate indices.

1. Simple indices.

Let us assume that y_t denotes the value of the characteristic we are looking at at time t, and we are interested in the rate of change of this value. By

$$i_{t|s} = \frac{y_t}{y_s}$$

we will denote the simple index number, describing the change between the value of y between periods s and t. The rate of change (in per cent) of the value of y between periods s and t is then equal to $(i_{t|s} - 1) \cdot 100\%$. For example, if we compare the average wage level in Poland in 2016 ($y_{2016} = 4277$) to the wage level in 2015 ($y_{2015} = 4121$), we see that the index number representing the yearly change is equal to $i_{2016|2015} = \frac{4277}{4121} \approx 1.038$, so the rate of growth of wages between 2015 and 2016 is equal to 3.8%.

The example above is a case where we look at changes from one period ($t-1$) to the next (t). Indices of the type $i_{t|t-1}$ are called **chain indices**. We could also look at the chain index describing the wage level changes between 2014 and 2015 ($i_{2015|2014} = \frac{4121}{3980} \approx 1.035$), the index describing the change between 2013 and 2014, etc. We may also be looking, however, at changes between specific periods with respect to a base year (t^*); we will then be looking at indices of the type $i_{t|t^*}$ for different values of t and for a fixed value of the base year t^*. This will occur if, for some reason, we want to compare subsequent values to one specific value. For example, we could wish to compare wage levels in Poland for the years 2005-2016 to the base level of 2004, corresponding to the date of the EU accession of Poland, in order to visualize the changes that happened since that moment in time.

Chained simple indices may easily be transformed into indices with a fixed base, and vice versa, using basic properties of fractions. For example, if we wanted to transform the two chain indices describing wage level growth in Poland between 2014 and 2016 into an index with a base of 2014, describing the change between 2014 and 2016, we would have

$$i_{2016|2014} = \frac{4277}{3980} = i_{2016|2015} \cdot i_{2015|2014} = \frac{4277}{4121} \cdot \frac{4121}{3980}.$$

If we have several chain indices (or one index number) representing a change between periods t and $t+n$, we may also be interested in the so-called average rate of change between these periods. For example, given different values of wage levels in Poland since 2004, we could be interested in the average yearly rate of change since that time. The average rate of change for value y for a time span of $n+1$ periods from t to $t+n$ is defined as

$$r = \sqrt[n]{\left(\prod_{i=1}^{n} i_{t+i|t+i-1}\right)} - 1 = \sqrt[n]{i_{t+n|t}} - 1 = \sqrt[n]{\frac{y_{t+n}}{y_t}} - 1.$$

Note that although we have $n+1$ periods, we have n changes during the time span (n chain indices in the product), and that is why the root is of order n. This average rate of change tells us what is the fixed rate of change that corresponds to the overall changes observed throughout the period, i.e. instead of looking at real changes from t to $t+1$, form $t+1$ to $t+2$, etc., we decompose the overall change between t and $t+n$ into equal increments for the whole period. In effect, we have

$$y_{t+n} = y_t \cdot (1+r)^n.$$

2. Aggregate indices

If we are looking at more than one characteristic simultaneously, we may look at simple indices for each of the characteristics separately, but that will not be a clear indication of the aggregate changes for the whole phenomenon. A very common example of such a multidimensional case are the changes in the value of sales of a group of products (considered, for example, in the macro-scale when we analyze changes in the GDP or price indices). The values depend on the quantities and on the prices, and if we consider more than one product and are interested in the effects prices and quantities may have on the value, the situation requires a careful analysis.

Let us consider a simple example of a company producing/selling (at least) two different products, A and B, and let us assume that the prices (p) and quantities (q) sold of both of these products may vary in time. If we look at the total sales worth in period t with respect to period $t-1$, this sales changes from $p_{A,t-1}\cdot q_{A,t-1}+p_{B,t-1}\cdot q_{B,t-1}(+...)$ to $p_{A,t}\cdot q_{A,t}+p_{B,t}\cdot q_{B,t}(+...)$. In order tu summarize this change, we may introduce an aggregate index for the value of sales between $t-1$ and t:

$$I_V = \frac{\sum_{i=1}^{k} p_{i,t} \cdot q_{i,t}}{\sum_{i=1}^{k} p_{i,t-1} \cdot q_{i,t-1}},$$

which describes the change in the total sales worth for products numbered $i = 1, ..., k$. Note that based on the value of this index only, we are not able to say if the change in the value of sales was driven by price changes or by quantity changes of the products sold (especially if prices or quantities of different products changed in different directions).

Let us now assume we want to monitor price changes in time in the economy (inflation). We are then interested in the changes of the value of the basket of goods a consumer is buying; these changes may arise from both price changes and changes in the composition of the basket of goods. In this case, too, we may wish to introduce a measure describing the relative importance of price changes (inflation) in the purchasing power changes of consumers. This means that we should compare the changes in the value of a fixed basket of goods for the two periods. But *which* basket of goods should we take for the comparison? The one for the initial period (let's denote it by 0) or the one for the next period (let's denote it by 1)? Depending on the choice, we will either calculate the **Laspeyres price index**, ${}_LI_p$, (if we use the base period quantities), or the **Paasche price index**, ${}_PI_p$, (if we use the next period quantities):

$${}_LI_p = \frac{\sum_{i=1}^{k} p_{i,1} \cdot q_{i,0}}{\sum_{i=1}^{k} p_{i,0} \cdot q_{i,0}},$$

$${}_PI_p = \frac{\sum_{i=1}^{k} p_{i,1} \cdot q_{i,1}}{\sum_{i=1}^{k} p_{i,0} \cdot q_{i,1}}.$$

Obviously, the value of the index will depend on the choice of the set of quantities used. If we want to get rid of this influence, we could take an (geometric) average of the two indices, and calculate the so-called **Fisher price index**:

$${}_FI_p = \sqrt{{}_LI_p \cdot {}_P\, I_p}.$$

Each of the above indices tells us what is the effect the change of prices had on the change of aggregate values (assuming quantities from the base period, next period, or averaging the two, for the three formulas respectively).

Similarly, if we were interested in the aggregate effect of quantity changes on the aggregate product values, we can calculate the **Lasypeyres, Paasche and Fisher quantity indices** as:

$${}_LI_q = \frac{\sum_{i=1}^{k} q_{i,1} \cdot p_{i,0}}{\sum_{i=1}^{k} q_{i,0} \cdot p_{i,0}},$$

$$
{}_P I_q = \frac{\sum_{i=1}^{k} q_{i,1} \cdot p_{i,1}}{\sum_{i=1}^{k} q_{i,0} \cdot p_{i,1}},
$$

$$
{}_F I_q = \sqrt{{}_L I_q \cdot {}_P I_q}.
$$

In this case, the formulas follow the same pattern, with the Laspeyres index using fixed prices from the base period, Paasche index using fixed prices from the next period, and the Fisher index being an average of the two. These indices tell us what is the effect of the change of quantities on the aggregate product values (assuming prices from the base period, next period, or averaging the two, for the three formulas respectively).

By comparing the fractions, we can easily show that the following relationships between the indices hold:

$$
I_V = {}_L I_p \cdot {}_P I_q = {}_P I_p \cdot {}_L I_q = {}_F I_p \cdot {}_F I_q.
$$