

# IndexNumber: An R Package for Measuring the Evolution of Magnitudes

by Alejandro Saavedra-Nieves and Paula Saavedra-Nieves

**Abstract** Index numbers are descriptive statistical measures useful in economic settings for comparing simple and complex magnitudes registered, usually in two time periods. Although this theory has a large history, it still plays an important role in modern today's societies where big amounts of economic data are available and need to be analyzed. After a detailed revision on classical index numbers in literature, this paper is focused on the description of the R package `IndexNumber` with strong capabilities for calculating them. Two of the four real data sets contained in this library are used for illustrating the determination of the index numbers in this work. Graphical tools are also implemented in order to show the time evolution of considered magnitudes simplifying the interpretation of the results.

## 1 Introduction

The problem of reducing a large amount of available microeconomic data is common in dynamic and modern economies. Individuals of today's societies consume services of hundred of commodities over a year, and most producers use and produce hundreds of individual products and services. This overwhelming abundance of data is usually summarized through index numbers theory. *Index numbers* are descriptive statistical measures useful in order to compare or measure changes in simple and complex magnitudes over time. The goal is usually to determine possible increases or decreases and, more generally, trend changes. The situations to be compared are in no way restricted; they may be two time periods (hours, days, months, or years); two places (two cities or countries); or two groups of people (economically active and inactive population). For simplicity in the exposition, we refer to temporal index numbers in this paper. The rest of the situations mentioned could be considered modifying the notation slightly.

Although index numbers are considered a classical statistical tool, the problem of how to construct them is as much one of economic theory as of statistical technique, see [Frisch \(1936\)](#). This can be checked by analyzing their considerable history. Initial works involving index numbers date back to the early 18th century. During the creation of a college in 1440–1460, it was stipulated that any member had to leave it if his richness exceeded five pounds per year. The Anglican Bishop Fleetwood desired to know if, according to the price evolution, this promise could be kept three centuries later. Then, he studied the evolution of prices corresponding to four products (wheat, meat, drink, and clothing) from 1440 to 1700. He concluded that five pounds in 1440–1460 had the same value as 30 pounds in 1700. More details can be found in [Fleetwood \(1707\)](#). [Dütot \(1754\)](#) studied the diminution of the money value analyzing the incomes of kings Louis XII and Louis XV. In order to know which of them had the largest disposable income, he considered the prices of several goods of different nature as such as a chicken, a rabbit, a pigeon, or the day's work value. From the discovery of the Americas, the astronomy professor of Padoue analyzed the evolution of prices in 1764. He considered the prices of three commodities (grains, wine, and cooking oil), and he studied their variation from 1500 to 1750. [Evelyn \(1798\)](#) can be seen as a precursor of index numbers establishing a price index number from 1050 to 1800. In this work, notions as the year of reference and relative prices were introduced. According to [Kendall \(1977\)](#), the real father of index numbers is Joseph Lowe. Many problems dealing with their construction were presented in [Lowe \(1822\)](#). In fact, Lowe's measurement would be known as the Laspeyres index later. In the second half of the 19th century, statisticians developed many advances in this setting. [Jevons \(1863\)](#) recommended considering the geometrical mean in order to construct an index number. Between 1864 and 1880, [Laspeyres \(1864, 1871\)](#), [Drobisch \(1871\)](#), and [Paasche \(1874\)](#) worked on the evolution of prices for material goods from the approach of weighted index numbers. [Palgrave \(1886\)](#) proposed to weigh the relative prices by the total amount of the considered good. [Fisher \(1922\)](#) defined a new index number calculated as the geometric mean of the Laspeyres and the Paasche index numbers. In the same period, Marshall and Edgeworth proposed to calculate weighted means. For an in-depth review on this topic, see [Kendall \(1969\)](#), [Allen \(1982\)](#) or more recently, [CPI Manual \(2004\)](#), or [Dodge \(2008\)](#).

Nowadays, economists continue to use index numbers to make comparisons over time despite their long background. In fact, the main applications of index numbers are either economic, or they occur in related fields as demography or technology. In such settings, the magnitudes to be compared through index numbers usually come in pairs, one of price and the other a matching one of quantity. This pair may be designed to account for the variation in an aggregate value, as when movements in the aggregated expenditure of consumers are analyzed into the two components of changes in

prices and in real consumption. Some more recent contributions in index numbers theory are exposed next. Barnett (1980) focused on economic monetary aggregates from this approach. Changes in food prices were analyzed in Lamm (1980). Index numbers in chain, or more commonly chain indices, are considered in Forsyth and Fowler (1981). Boyle (1988) analyzed the volume of Irish agricultural output from 1960 to 1982. Scanner data on coffee sales are studied in de Haan and Opperdoes (1997). Hill (1999) shows how a comparison of price levels across a group of countries can be made by chaining Fisher index numbers across a spanning tree. Inequality and poverty in several regions of Thailand are studied through the construction of urban and rural cost of living and welfare indices in Kakwani and Hill (2002). Dumagan (2002) showed that the Fisher index could be numerically approximated by other superlative index numbers. In Reinsdorf et al. (2002), additive decompositions of the Fisher index are derived in order to know how much each item contributes to its overall change. From data from United States, Canada, France, Germany, Italy, Japan, and the United Kingdom, drug price and quantity index numbers are considered in Danzon and Chao (2000). Ang et al. (2004) used a generalized version of the Fisher index to analyze CO<sub>2</sub> emission. In Boyd and Roop (2004), the structural change in energy intensity is studied. Exploring the duality between a return to dollar definition of profit and the generalized distance function, the relationship between the Laspeyres, Paasche and Fisher productivity index numbers is established in Zoffio and Prieto (2006). Hill (2006) showed an illustration on index numbers also using scanner data. An application to major crops in Manitoba is presented in Coyle (2007). According to Diewert and Nakamura (2007), Paasche, Laspeyres, or Fisher index number formula is useful in order to manage the total factor productivity growth. The importance of the hedonic imputation method in price index numbers is analyzed in Hill and Melsner (2008) from a data set containing house prices for three regions in Sydney over a three years period. The impact of time aggregation on price change estimates for several supermarket item categories is considered in Ivancic et al. (2011). Białek (2012) proposed a general price index formula with the Fisher, Laspeyres, and Paasche indices as its particular cases. Białek (2014) presented an original price index, and its performance is analyzed through a simulation study where it is compared to several classical price index numbers. A generalized version of the Fisher index is considered in Su and Ang (2014) in order to analyze changes in the carbon emissions embodied in China's exports. O'Donnell (2018) analyzed the productivity change defined as measures of output quantity change divided by measures of input quantity change. Zhen et al. (2019) constructed panel price index numbers using retail scanner data in order to compare consumption costs across space and time.

This paper is focused on the description of the R package **IndexNumber** (Saavedra-Nieves and Saavedra-Nieves, 2021), available from the *Comprehensive R Archive Network* at <https://CRAN.R-project.org/package=IndexNumber>, and its capabilities for calculating classical index numbers. It is organized as follows. Index numbers are formally defined in Section 2.2. Section 2.2.1 introduces simple index numbers. Concretely, simple index numbers in series and in chain are distinguished. In Section 2.2.2, non-weighted and weighted complex index numbers are presented. Details on the usage of **IndexNumber** package are considered from Section 2.3. The four real data sets contained in the library are also described briefly. Note that two of them are used in this paper in order to illustrate the calculation of index numbers. They are available on the website of the *Spanish Statistical Office (INE)*, <http://www.ine.es>.

## 2 Preliminaries on index numbers

*Index numbers* are statistical measures that are useful to compare single and multiple magnitudes for the same interval of time. In both cases, this comparison is made with respect to an element of the mentioned series that is called base period or reference. Some examples of simple magnitudes are prices of a good, sold amounts of a product, or other general individual values. However, most of the time, comparing these simple quantities has not practical interest. If the goal is to analyze some real phenomena where many variables are involved, complex indices must be considered. Using these ideas, index numbers are usually classified into the following two groups:

- An index number is said to be simple if it corresponds to the ratio of two values of the same variable, measured in two different instants. Therefore, a *simple index number* provides the variation that the single magnitude has suffered between two different time periods.  
For instance, a simple index number of the price will give the relative variation of the price between the current period and the period taken as reference.
- Most of the time, comparing prices, amounts, or values of a single product individually is not of interest in practice. If the goal is to analyze some real situations where different variables have influence, a *complex number index* has to be considered. It globally summarizes the information of the different magnitudes involved in the problem.  
For instance, the evolution cost of life in a country is a common case where it is necessary to select a set of goods or variables that give information about it. The relative importance of each of the goods considered must be measured and taken into account.

A wider overview of both classes of index numbers is included in the next sections. In particular, we distinguish the different subclasses belonging to each of them and their possible relationships.

**Simple index numbers**

A *simple index number* is a statistical indicator of the percentage of variation of a single magnitude in two different instants. Simple index numbers are usually classified according to the element that we take as reference. In particular, we distinguish two types of simple index numbers. First, we describe simple index numbers in series, when the first value of the series is taken as the reference value, and simple index numbers in chain (or chain indices), when the reference is the immediately previous value in the serie.

In what follows, we assume that  $X = \{x_0, x_1, \dots, x_T\}$  denotes the observations of the magnitude  $X$  for the  $T + 1$  time instants considered. Besides,  $x_0$  is usually taken as the base period. Most common simple index numbers are individually referred to variables in real-world situations as the followings:

- the *price of a good*, denoted by  $p$ ;
- the *amount of produced or sold product*, denoted by  $q$ ; or
- the *value of a good*, denoted by  $v$ . This value is usually obtained as the product of the price and the amount variables.

In this section, we illustrate the usage of simple index numbers on a real example. Table 1 shows the number (thousands) of economically active women and men in Spain from the first trimester of 2002 (taken as a reference value). Remark that four trimesters of each year are denoted by T1, T2, T3 and T4, respectively.

Stages	2002 (T1)	2002 (T2)	2002 (T3)	2002 (T4)	2003 (T1)	2003 (T2)	...
Total of women	7442.70	7580.80	7670.20	7751.50	7868.70	7977.80	...
Total of men	11192.30	11289.40	11445.10	11472.80	11552.50	11661.40	...

**Table 1:** Number (thousands) of economically active women and men in Spain from first trimester of 2002.

This dataset, included in **IndexNumber** package, can be obtained from the Economically Active Population Survey (EPA) elaborated by the Spanish Statistics National Institute (INE).

**Simple index numbers in series**

Let  $x_0$  and  $x_t$  be the values of the variable  $X$  where  $x_0$  corresponds to the base period and  $t \in \{0, 1, \dots, T\}$ , respectively. Thus, the value of the simple index number in series for  $X$  in  $t$  is defined as follows:

$$I_0^t(X) = \frac{x_t}{x_0} \cdot 100. \tag{1}$$

For each  $t \in \{0, 1, \dots, T\}$ , this measure has a natural interpretation. Fixed a certain variable of interest  $X$ , the index number in series in  $t$  shows the percentage of variation of the magnitude in this instant of time with respect to the reference value (in this case, the initial one).

Using this type of index number, some usual magnitudes can be formally defined as it is indicated next:

- When prices are considered, the *relative price* of a product  $i$ , also called simple price index, can be determined as

$$p_0^t = \frac{p_{it}}{p_{i0}},$$

where  $p_{it}$  denotes the price at instant  $t$  and  $p_{i0}$ , the price in the base period.

- The *relative amount* of a product  $i$  can be written as

$$q_0^t = \frac{q_{it}}{q_{i0}},$$

where  $q_{it}$  denotes the produced or sold amount at instant  $t$  and  $q_{i0}$ , the amount for the base period.

- Finally, the *relative value* of a product  $i$  has the next expression:

$$v_0^t = \frac{v_t}{v_0} = \frac{p_{it}q_{it}}{p_{i0}q_{i0}} \cdot 100 = p_0^t \cdot q_0^t \cdot 100,$$

where  $v_t$  denotes the value of the good at instant  $t$  and  $v_0$ , the value of the base period.

Below, we illustrate the usage of simple index numbers on the two series included in Table 1. Thus, the simple index numbers in series for the economically active women and men in Spain from the first trimester of 2002 is given in Table 2.

Stages	2002 (T1)	2002 (T2)	2002 (T3)	2002 (T4)	2003 (T1)	2003 (T2)	...
Index number for women	100.00	101.86	103.06	104.15	105.73	107.19	...
Index number for men	100.00	100.87	102.26	102.51	103.22	104.19	...

**Table 2:** Simple index numbers in series for number (thousands) of economically active women and men in Spain from the first trimester of 2002.

Comparing the evolution of the index numbers in series for the population of women and men shown in Table 2 has an interest, for instance, to analyze the effect of variable sex in the Spanish labor market. Note that the number of economically active women and men in the second trimester of 2003 is 7.2% and 4.2% larger than in the reference time, respectively. Therefore, women increasing is slightly larger than men.

### Simple index numbers in chain

Below, we introduce another approach of simple index numbers. Contrary to our previous assumptions, this new setting arises when the reference value is not the initial one; rather we take the value immediately preceding. Let  $x_t$  and  $x_{t-1}$  be two values of a variable  $X$  observed in two consecutive instants  $t$  and  $t - 1$ , being  $t \in \{1, 2, \dots, T\}$ . Thus, the value of the *index number in chain* or *chain index* (cf. Forsyth and Fowler, 1981) that corresponds to an instant  $t$ , with  $t \in \{1, 2, \dots, T\}$ , is defined as follows:

$$IC^t(X) = \frac{x_t}{x_{t-1}} \cdot 100. \tag{2}$$

Again, this index number can be naturally interpreted. It is worth mentioning that these measures the variation of the characteristic under study with respect to the previous value in a fixed instant  $t$ . For instance, these index numbers reflect the percentage variation that the variable experiments between two consecutive values in time series settings.

To illustrate this definition, we take again the example considered in the previous section. For this subset of values, we obtain again the evolution of the amount of economically active people (per sex) in Spain under this new approach.

Stages	2002 (T1)	2002 (T2)	2002 (T3)	2002 (T4)	2003 (T1)	2003 (T2)	...
Index number for women	100.00	101.86	101.18	101.06	101.51	101.39	...
Index number for men	100.00	100.87	101.38	100.24	100.70	100.94	...

**Table 3:** Simple index numbers in chain for number (thousands) of economically active women and men in Spain from the first trimester of 2002.

Table 3 shows the evolution in time of the number (thousands) of economically active women and men in Spain from the trimester of 2002 (see Table 1). According to the obtained results, we emphasize as relevant that the number of economically active women and men in the second trimester of 2003 increases 1.4% and 0.9%, respectively, with respect to the previous trimester.

### Relationship between simple index numbers in series and in chain

This section briefly introduces some comments on the relations between simple index numbers in series and in chain. In this way, one can be obtained from another (and vice versa) without having to use the exact values of the magnitude under study.

Take  $x_t$  the value of the variable  $X$  in instant  $t$ , with  $t \in \{1, 2, \dots, T\}$ . Thus, a simple index in chain can be obtained from a simple index in series due to the relation

$$IC^t(X) = \frac{x_t}{x_{t-1}} \cdot 100 = \frac{\frac{x_t}{x_0}}{\frac{x_{t-1}}{x_0}} \cdot 100 = \frac{I_0^t(X)}{I_0^{t-1}(X)} \cdot 100.$$

For example, from simple index numbers in series for women shown in Table 2, it is possible to obtain the corresponding simple index number in chain for women also for the first trimester of 2003 (contained in Table 3). That is,

$$IC^{2003(T1)}(X) = \frac{7868.70}{7751.50} \cdot 100 = \frac{105.72}{104.15} \cdot 100 = 101.51.$$

Besides, simple indices in series can be equivalently obtained for each  $t \in \{1, \dots, T\}$  from indices

in chain:

$$I_0^t(X) = \frac{x_t}{x_0} \cdot 100 = \frac{x_t}{x_{t-1}} \cdot \frac{x_{t-1}}{x_{t-2}} \cdot \dots \cdot \frac{x_2}{x_1} \cdot \frac{x_1}{x_0} \cdot 100 = \frac{IC^t(X)}{100} \cdot \frac{IC^{t-1}(X)}{100} \cdot \dots \cdot \frac{IC^2(X)}{100} \cdot \frac{IC^1(X)}{100} \cdot 100.$$

For instance, if we consider the simple index numbers in chain shown in Table 3, it is possible to obtain the simple index number in series for the first trimester of 2003 contained in Table 2. That is, we check that

$$I_0^{2003(T1)}(X) = \frac{100.70}{100} \cdot \frac{100.24}{100} \cdot \frac{101.38}{100} \cdot \frac{100.87}{100} \cdot 100.$$

**Variation Rate**

In this section, we formally introduce another measure of the evolution of a magnitude. Furthermore, we relate it to the simple index numbers previously defined.

The *variation rate* of the observations in the instants  $t$  and  $t - 1$ , with  $t \in \{1, \dots, T\}$ , is denoted by  $Rate^t(X)$  for each  $t \in \{1, \dots, T\}$ . It can be calculated from the simple index in chain as follows:

$$Rate^t(X) = \frac{x_t - x_{t-1}}{x_{t-1}} \cdot 100 = IC^t(X) \cdot 100.$$

This definition can be extended to any pair of instants in  $\{0, 1, 2, \dots, T\}$ . Take  $t_1, t_2 \in \{0, 1, 2, \dots, T\}$  such that  $t_1 < t_2$ . Let  $x_{t_1}$  be the value of a variable measure in instant  $t_1$  and let  $x_{t_2}$  be the value of a variable measure in a later instant  $t_2$ . Formally, the *variation rate* of  $X$  in  $t_2$  with respect to  $t_1$  is

$$Rate_{t_1}^{t_2}(X) = \frac{x_{t_2} - x_{t_1}}{x_{t_1}} \cdot 100.$$

Note that

$$Rate_{t_1}^{t_2}(X) = \frac{x_{t_2} - x_{t_1}}{x_{t_1}} \cdot 100 = \left( \frac{x_{t_2}}{x_{t_1}} - 1 \right) \cdot 100 = I_{t_1}^{t_2}(X) - 100,$$

where  $I_{t_1}^{t_2}(X) = \frac{x_{t_2}}{x_{t_1}} \cdot 100$ , also satisfying

$$I_{t_1}^{t_2}(X) = IC^{t_2}(X) \cdot IC^{t_2-1}(X) \cdot \dots \cdot IC^{t_1+1}(X).$$

In particular, if the consecutive observations correspond to two different years, months, or trimesters the variation rate is called interannual variation rate, monthly variation rate, and quarterly variation rate, respectively. From the data shown in Table 2, the quarterly variation rates have been calculated in Table 4.

Stages	2002 (T2)	2002 (T3)	2002 (T4)	2003 (T1)	2003 (T2)	...
Rate for women	1.86	1.18	1.06	1.51	1.39	...
Rate for men	0.87	1.38	0.24	0.70	0.94	...

**Table 4:** Quarterly variation rate for number (thousands) of economically active women and men in Spain.

**Average Variation Rate**

Finally, we introduce a third method to measure the evolution of a given magnitude  $X$  between the instants  $t$  and  $t + k$ , with  $t \in \{0, \dots, T - k\}$  and  $k > 0$ . If  $x_{t+k}$  denotes the observation at instant  $t + k$  and  $x_t$  the corresponding to instant  $t$ , the *average variation rate* of the variable  $X$  between instants  $t$  and  $t + k$  is defined as the constant rate  $T_k$  that allows obtaining the observation  $x_{t+k}$  at time  $t + k$  from observation  $x_t$  at time  $t$ .

Then, it is possible to write:

$$\begin{aligned} x_{t+1} &= x_t + \frac{T_k}{100} \cdot x_t = \left( \frac{100 + T_k}{100} \right) \cdot x_t \\ x_{t+2} &= x_{t+1} + \frac{T_k}{100} \cdot x_{t+1} = \left( \frac{100 + T_k}{100} \right)^2 \cdot x_t \\ x_{t+3} &= x_{t+2} + \frac{T_k}{100} \cdot x_{t+2} = \left( \frac{100 + T_k}{100} \right)^3 \cdot x_t \\ &\vdots \\ x_{t+k} &= x_{t+k-1} + \frac{T_k}{100} \cdot x_{t+k-1} = \left( \frac{100 + T_k}{100} \right)^k \cdot x_t \end{aligned}$$

Therefore,  $x_{t+k} = \left(\frac{100+T_k}{100}\right)^k \cdot x_t$  and, as consequence,  $\frac{x_{t+k}}{x_t} = \left(\frac{100+T_k}{100}\right)^k$ . Then,

$$T_k = 100 \cdot \left(\frac{x_{t+k}}{x_t}\right)^{1/k}.$$

From data shown in Table 2, we have calculated the quarterly average variation rate in 2002 for women and men as

$$\left(\sqrt[3]{\frac{7751.50}{7442.70}} - 1\right) \cdot 100 = 1.36 \text{ and } \left(\sqrt[3]{\frac{11472.80}{11192.30}} - 1\right) \cdot 100 = 0.83, \text{ respectively.}$$

According to the results obtained, the average variation rate for women is considerably bigger than the corresponding for men in this period.

### Complex index numbers

Most of the time, comparing variables marginally does not provide real information. There are many phenomena in the real world that involve several variables. Of course, simple index numbers described in the previous section could be naturally applied for each of these variables separately. However, comparing these magnitudes can be not realistic and, for this reason, *complex index numbers* have to be introduced. A complex index number summarizes the information of the different marginal index numbers related to the set of variables of interest.

One of the most relevant examples that illustrates the use of complex index numbers is briefly described next. The evolution of living cost in a country is a common case where it is necessary to select a set of goods or variables that give information about it. The relative importance of each of the goods considered may be taken into account. This is the case of the *Consumer Price Index* (CPI) in Spain. The sets of good considered for calculating it follows the *International Classification of Consumption according to Purpose* (COICOP) prepared by the Statistical Division of the United Nations. This classification is also used by other countries. In this way, comparisons between different geographical areas make sense.

When all variables are not of equal relevance, it is possible to add complementary information for weighting each magnitude corresponding to its degree of importance. Depending on the use of this additional information, two classes of complex index numbers are distinguished in literature:

- In several practical cases, the relative weight of each involved variable has no interest. Fixed a magnitude, the class of required index numbers to be used in this setting are named *non-weighted complex index numbers*.
- The use of *weighted index numbers* allow greater importance to be attached to some items. For instance, real information other than simply the change in price over time can be used. Factors as quantity sold or quantity consumed for each item can also be considered.

In what follows, we assume that  $X = (X_1, \dots, X_n)$  denotes the collection of  $n$  magnitudes registered for  $n$  different products. For each  $j = 1, \dots, n$ ,  $X_j = \{x_{j0}, x_{j1}, \dots, x_{jT}\}$  denotes the observations of the magnitude  $X_j$  for the  $T + 1$  instants of time considered. Analogously to the simple case, we refer  $x_{j0}$  as the base period.

In practice, the available information can be summarized in a table such as Table 5.

Time	Products			
	1	2	...	$n$
0	$x_{10}$	$x_{20}$	...	$x_{n0}$
1	$x_{11}$	$x_{21}$	...	$x_{n1}$
⋮	⋮	⋮	⋮	⋮
$T$	$x_{1T}$	$x_{2T}$	...	$x_{nT}$

**Table 5:** Evolution of a set of magnitudes  $X$  from time 0 to  $T$ .

The most common complex indices jointly involve (some of) the variables in real-world situations that we enumerate below:

- the *price of a collection of  $n$  goods*, denoted by  $p = (p_1, \dots, p_n)$ , where  $p_j = \{p_{j0}, p_{j1}, \dots, p_{jT}\}$  denotes the prices for  $T + 1$  instants and for each product  $j = 1, \dots, n$ ;
- the *amount of produced or sold products*, denoted by  $q$ , where  $q_j = \{q_{j0}, q_{j1}, \dots, q_{jT}\}$  denotes the amounts of product for  $T + 1$  instants and for each product  $j = 1, \dots, n$ ; or



- the value of  $n$  goods is given by  $v = (v_1, \dots, v_n)$ . In this case,  $v_j = \{v_{j0}, v_{j1}, \dots, v_{jT}\}$  denotes the values of product  $j$ , with  $j \in \{1, \dots, n\}$  for  $T + 1$  instants.

To illustrate the usage of complex index numbers, we take the example described in Table 6. It shows the unitary value (euros) of prices of combustibles and other energy resources for the main home in Spain from 2005 to 2015. In this case, the data source is again the Spanish Statistics National Institute (INE).

	Electricity (Kwh)	Natural municipal gas ( $m^3$ )	Liquified gas (kilo)	Liquified combustibles (litre)	Solid combustibles (kilo)
2006	0.14	0.70	1.00	0.69	0.12
2007	0.14	0.78	0.97	0.66	0.10
2008	0.15	0.83	1.03	0.81	0.11
2009	0.16	0.87	0.93	0.68	0.10
2010	0.17	0.79	1.04	0.72	0.12
2011	0.19	0.77	1.15	0.91	0.12
2012	0.22	0.82	1.19	0.96	0.12
2013	0.23	1.00	1.34	0.99	0.13
2014	0.24	1.07	1.35	0.89	0.15
2015	0.25	1.04	1.22	0.77	0.13

**Table 6:** Unitary value (euros) of combustibles and other energy resources for the main home in Spain from 2005 to 2015.

Table 7 shows the (thousands of units) consumption of combustibles and other energy resources for the main home in Spain from 2005 to 2015. This dataset is closely related to the prices presented in Table 6. Of course, this data set can also be obtained from the Spanish Statistics National Institute (INE).

	Electricity (Kwh)	Natural municipal gas ( $m^3$ )	Liquified gas (kilo)	Liquified combustibles (litre)	Solid combustibles (kilo)
2006	50623635	3617285	1057488	2297923	1306920
2007	51990501	3266575	1066857	2454265	1602799
2008	54990338	3473851	1210607	2274326	1556673
2009	59749470	3730349	1113642	2505711	1724222
2010	69751162	3954065	987112	2345215	1584123
2011	67574654	4466072	926824	1974662	1414234
2012	62878557	4576052	943632	2029733	1733591
2013	56017871	4116079	867695	1952593	2071152
2014	49177739	3653055	868743	2180866	2077766
2015	48541712	3795339	818183	2176533	2161208

**Table 7:** Consumption (thousands of units) of combustibles and other energy resources for the main home in Spain from 2005 to 2015.

Next, we formally describe non-weighted and weighted index numbers.

### Non-weighted complex index numbers

Complex index numbers analyze several magnitudes that measure the evolution of a set of goods or services. The goal here is to find a statistical measure in order to summarize the information shown, for instance, in Table 5. In particular, knowing the variation of a magnitude in time  $t$  with respect to the base period has an interest. In this sense, it is worth mentioning that non-weighted complex index numbers can be easily obtained as the arithmetic, geometric and harmonic means of the simple index numbers in series for the considered magnitudes. Anyway, their mathematical expressions are described below:

- The *Sauerbeck index* (cf. Sauerbeck, 1895) at time  $t$  for  $X$ ,  $S^t(X)$ , is calculated as the arithmetic mean of the simple index in series for the  $n$  involved magnitudes at  $t$ :

$$S^t(X) = \frac{1}{n} \sum_{i=1}^n \frac{x_{it}}{x_{i0}} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{3}$$

- The *Geometric mean index* at time  $t$ ,  $G^t(X)$ , is calculated as follows:

$$G^t(X) = \sqrt[n]{\prod_{i=1}^n \frac{x_{it}}{x_{i0}}} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{4}$$

Given a collection of  $n$  magnitudes, the geometric mean index at  $t$  is obtained as  $n^{\text{th}}$ -root of the product of simple index numbers for  $X$  at time  $t$ . See more details in [Jevons \(1863\)](#).

- The *Harmonic mean index* at time  $t$ ,  $H^t(X)$ , is determined by

$$H^t(X) = \frac{n}{\sum_{i=1}^n \frac{x_{i0}}{x_{it}}} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{5}$$

It is initially introduced in [Jevons \(1865\)](#) and [Coggeshall \(1886\)](#).

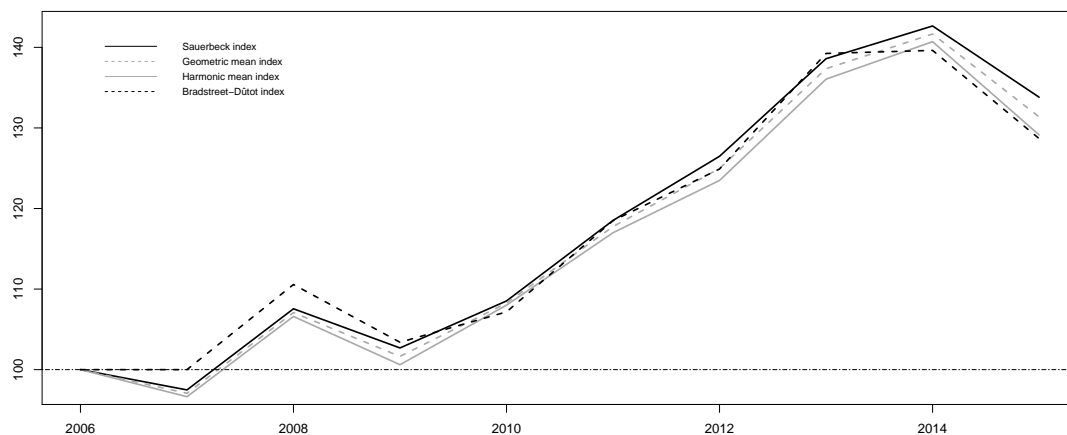
- The *Bradstreet-Dütot index* at time  $t$ ,  $BD^t(X)$ , is introduced in [Walsh \(1901\)](#). Its value is obtained as the ratio between the means of the magnitude in time  $t$  and the magnitude in the base period as follows:

$$BD^t(X) = \frac{\sum_{i=1}^n x_{it}}{\sum_{i=1}^n x_{i0}} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{6}$$

If  $X$  denotes the matrix  $p$  of prices of a set of goods or services along a period of time, these index numbers are specifically considered *non-weighted complex index numbers for prices*. Thus, it arises the *Sauerbeck index* at time  $t$  for prices,  $S^t(p)$ ; the *Geometric mean index* at time  $t$  for prices,  $G^t(p)$ ; the *Harmonic mean index* at time  $t$  for prices,  $H^t(p)$ ; and the *Bradstreet-Dütot index* at time  $t$  for prices,  $BD^t(p)$ . Their usage will be illustrated on the set prices of combustibles and other energy resources for the main home in Spain from 2005 to 2015. From the information in [Table 6](#), the four index numbers previously described are numerically shown in [Table 8](#), and their evolution is depicted in [Figure 1](#).

$t$	$S^t(p)$	$G^t(p)$	$H^t(p)$	$BD^t(p)$
2006	100.00	100.00	100.00	100.00
2007	97.48	97.06	96.64	100.00
2008	107.56	107.08	106.60	110.57
2009	102.69	101.64	100.61	103.40
2010	108.53	108.26	108.00	107.17
2011	118.52	117.76	116.99	118.49
2012	126.48	124.97	123.48	124.91
2013	138.59	137.35	136.05	139.25
2014	142.65	141.66	140.70	139.62
2015	133.81	131.37	129.14	128.68

**Table 8:** Non-weighted complex price indexes in seriea for the unitary value of combustibles and energy resources for the main home in Spain from 2006 to 2015.



**Figure 1:** Joint evolution of the non-weighted complex price indexes in series for the unitary value of combustibles and energy resources for the main home in Spain from 2006 to 2015.

However, the variation of a given magnitude in time  $t$  with respect to the previous period may also be of interest. Below, we alternatively enumerate the mathematical expressions of the previous indices based on index numbers in chain for the magnitudes.

- The *Carli index* (cf. [Carli, 1804](#)) at time  $t$  for  $X$ ,  $C^t(X)$ , is calculated as the arithmetic mean of the



simple index in chain for the  $n$  involved magnitudes at  $t$ :

$$C^t(X) = \frac{1}{n} \sum_{i=1}^n \frac{x_{it}}{x_{i,t-1}} \cdot 100, \text{ for each } t \in \{1, \dots, T\}. \tag{7}$$

- The *Jevons index* at time  $t$ ,  $J^t(X)$ , is calculated following *Jevons (1863)* as follows:

$$J^t(X) = \sqrt[n]{\prod_{i=1}^n \frac{x_{it}}{x_{i,t-1}}} \cdot 100, \text{ for each } t \in \{1, \dots, T\}. \tag{8}$$

Given a collection of  $n$  magnitudes, the geometric mean index at  $t$  is obtained as  $n^{\text{th}}$ -root of the product of simple index numbers for  $X$  in chain at time  $t$ .

- The *Dùtot index* at time  $t$ ,  $D^t(X)$ , is introduced in *Dùtot (1754)*. Its value is obtained as the ratio between the means of the magnitude in time  $t$  and the magnitude in time  $t - 1$  as follows:

$$D^t(X) = \frac{\sum_{i=1}^n x_{it}}{\sum_{i=1}^n x_{i,t-1}} \cdot 100, \text{ for each } t \in \{1, \dots, T\}. \tag{9}$$

Again, if  $X$  denotes the matrix of prices of a set of goods or services along a period of time ( $p$ ), these index numbers are named as the *Carli index* at time  $t$  for prices,  $C^t(p)$ ; the *Jevons index* at time  $t$  for prices,  $J^t(p)$ ; and the *Dùtot index* at time  $t$  for prices,  $D^t(p)$ . To conclude this section, we obtain them for the set prices of combustibles and other energy resources for the main home in Spain from 2005 to 2015, detailed in Table 6. These index numbers are numerically detailed in Table 9.

$t$	$C^t(p)$	$J^t(p)$	$D^t(p)$
2006	100.00	100.00	100.00
2007	101.11	101.09	101.79
2008	113.51	113.43	115.89
2009	100.94	100.18	126.56
2010	105.75	105.27	144.02
2011	103.58	103.29	155.49
2012	110.65	110.49	168.01
2013	107.00	106.32	163.98
2014	100.72	100.39	154.09
2015	93.08	92.79	153.29

**Table 9:** Non-weighted complex price indexes in chain for the unitary value of combustibles and energy resources for the main home in Spain from 2006 to 2015.

All of the index numbers described are easy to be computed. However, they present an important disadvantage: they do not take into account the relative importance of each product.

### Weighted complex index numbers

For analyzing the evolution of a given magnitude  $X$ , it is very common to use an alternative magnitude  $Y$  through the value of  $Y$  in the reference or the actual period to weight complex index numbers. The information relative to this alternative variable can be summarized in a table such as Table 10. For instance, the use of the amount of production of different products or the use of prices may result of interest, depending on the setting under study.

Time	Products			
	1	2	...	$n$
0	$y_{10}$	$y_{20}$	...	$y_{n0}$
1	$y_{11}$	$y_{21}$	...	$y_{n1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$T$	$y_{1T}$	$y_{2T}$	...	$y_{nT}$

**Table 10:** Evolution of a set of magnitudes  $Y$  from time 0 to  $T$ .

Next, the main weighted complex price index numbers for a given magnitude  $X$ , taking  $Y$  as weight, are formally described:

- The *Laspeyres index* (*Laspeyres, 1871*) analyzes the variations of  $X$  using  $Y$  as weight. In this sense, the weights considered for product  $i$  are  $x_{i0} \cdot y_{i0}$  (note that both values are referred to

the base period). Then, this complex index is defined as the weighted arithmetic means of the simple index numbers:

$$L^t(X, Y) = \frac{\sum_{i=1}^n \frac{x_{it}}{x_{i0}} x_{i0} y_{i0}}{\sum_{i=1}^n x_{i0} y_{i0}} \cdot 100 = \frac{\sum_{i=1}^n x_{it} y_{i0}}{\sum_{i=1}^n x_{i0} y_{i0}} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{10}$$

The main disadvantage of the Laspeyres index is that it assumes that the weights do not vary in time. This hypothesis is not always realistic in some practical settings.

- The *Paasche index* is an alternative index to the Laspeyres index introduced in Paasche (1874), when the weighted criteria is  $x_{i0} \cdot y_{it}$ . Therefore, it can be formally written as:

$$P^t(X, Y) = \frac{\sum_{i=1}^n \frac{x_{it}}{x_{i0}} x_{i0} \cdot y_{it}}{\sum_{i=1}^n x_{i0} \cdot y_{it}} \cdot 100 = \frac{\sum_{i=1}^n x_{it} y_{it}}{\sum_{i=1}^n x_{i0} y_{it}} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{11}$$

- The *Marshall-Edgeworth index* (cf. Marshall, 1887, Edgeworth, 1887) is an agregative weighted measure where weights are  $y_{i0} + y_{it}$ . Therefore, it can be calculated as:

$$E^t(X, Y) = \frac{\sum_{i=1}^n x_{it} (y_{i0} + y_{it})}{\sum_{i=1}^n x_{i0} (y_{i0} + y_{it})} \cdot 100, \text{ for each } t \in \{0, \dots, T\}. \tag{12}$$

- The *Fisher index* is equal to the geometric mean of the index numbers under the approaches of Laspeyres and Paasche:

$$F^t(X, Y) = \sqrt{L^t(X, Y) \cdot P^t(X, Y)}, \text{ for each } t \in \{0, \dots, T\}. \tag{13}$$

For instance, see more details in Fisher (1922).

Note that other values can be defined (as we will see below). The choice of using a specific index formula often relies on the availability of data. According to the previous comments, the Laspeyres index does not require information on the products of the current period. Then, the Laspeyres formula is usually preferred for the calculation of complex indices, which are typically released rapidly before information for the current period could have been collected.

In what follows,  $p$  denotes the matrix of prices of a set of goods or services along a period of time and  $q$  is the matrix of the total amounts of goods in the same period. Thus, the *weighted complex price index numbers* analyze the time evolution of prices by introducing the variation of the physical production or the consumption of a set of goods or services. The weights are obtained by multiplying the price of a product in an instant of time  $t$  by the consumption in the base period or the actual period. Hence, the *Laspeyres price index*,  $L^t(p, q)$ , the *Paasche price index*,  $P^t(p, q)$ , the *Marshall-Edgeworth price index*,  $E^t(p, q)$ , and the *Fisher price index*,  $F^t(p, q)$ , are naturally defined in prices settings.

Under this approach, a new complex index for  $v$  can be naturally introduced under the approach of the Bradstreet-Dütot index. It is based on the notion of the value of good indicated by  $v$ . It can be calculated as follows:

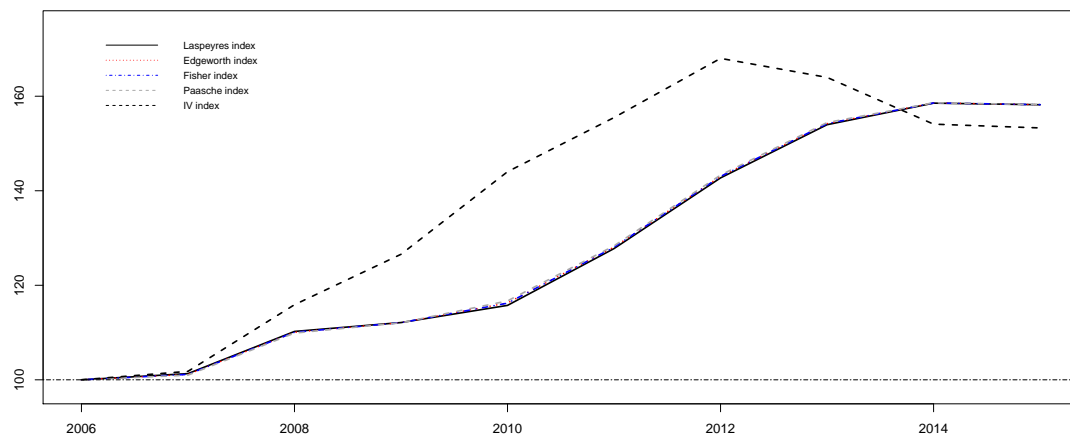
$$IV_0^t(p, q) = \frac{V_t(p, q)}{V_0(p, q)} = \frac{\sum_{i=1}^n p_{it} q_{it}}{\sum_{i=1}^n p_{i0} q_{i0}}, \text{ for each } t \in \{0, \dots, T\}. \tag{14}$$

It satisfies that  $IV_0^t(p, q) = L^t(p_0, q) \cdot P^t(p_0, q) = L^t(p, q_0) \cdot P^t(p, q_0) = F^t(p_0, q) \cdot F^t(p, q_0)$ .

From data contained in Tables 6 and 7, these five index numbers were determined. They are in Table 11 (from tenth to twelfth column), and they are graphically depicted in Figure 2.

$t$	$L^t(p, q)$	$P^t(p, q)$	$E^t(p, q)$	$F^t(p, q)$	$IV_0^t(p, q)$
2006	100.00	100.00	100.00	100.00	100.00
2007	101.31	100.99	101.15	101.15	101.79
2008	110.23	109.89	110.06	110.06	115.89
2009	112.11	112.06	112.09	112.09	126.56
2010	115.76	116.69	116.27	116.22	144.02
2011	127.77	128.35	128.08	128.06	155.49
2012	142.72	143.32	143.04	143.02	168.01
2013	153.98	154.43	154.21	154.20	163.98
2014	158.54	158.62	158.58	158.58	154.09
2015	158.20	158.23	158.21	158.21	153.29

**Table 11:** Weighted complex price indexes for the unitary value of combustibles and energy resources for the main home in Spain from 2006 to 2015.



**Figure 2:** Joint evolution of the weighted complex price indexes for the unitary value of combustibles and energy resources for the main home in Spain from 2006 to 2015.

Otherwise, the *weighted complex production index numbers* analyze the time evolution of the amount of product by introducing the variation of the price of the goods or services as weight. Their obtaining is analogous to the previous one. The weights are obtained by multiplying the amount of a product in an instant of time  $t$  by the price in the base period or the actual period. Thus, we dealt with the *Laspeyres production index*,  $L^t(q, p)$ ; the *Paasche production index*,  $P^t(q, p)$ ; or the *Fisher production index*  $F^t(q, p)$ .

### 3 IndexNumber in practice

This section presents an overview of the structure of the package. **IndexNumber** is a tool that R users can use in order to determine several classical index numbers that describe the evolution of a single magnitude or a set of magnitudes. This software helps the user to calculate faster these statistical measures. Functions in this library automatize the required operations for the computation of index numbers. First, we will describe the real data sets included in the package. Then, the functions implemented are detailed. Of course, other libraries exist in R dealing with index numbers theory. In particular, **micEconIndex** (Henningsen, 2017), **IndexNumR** (White, 2021), and **PriceIndices** (Bialek, 2021) packages also allow to compute complex index numbers but only when the considered magnitudes are prices and quantities. It is worth mentioning that **IndexNumber** library was designed under a more general perspective by extending to any type of magnitude. Moreover, none of the above-referenced packages implement simple index numbers, and they do not offer graphical tools to facilitate the analysis of time evolution series either. Additionally, **IndexNumber** can be seen as an additional basic library that can also be exploited by non-experts R users. For instance, inputs of functions are numeric vectors or matrices containing the magnitude values, much more flexible than data structures that other packages consider. As for computational complexity, it is also relatively simple because, unlike **IndexNumR** that implements multilateral methods, the number of required elementary operations is smaller. Finally, it is convenient to note that **IndexNumber** package also provides four new recent real data sets.

#### Data sets in IndexNumber

Index numbers have been theoretically introduced in previous sections using two real data sets included in the package **IndexNumber**. However, we decide to include them in the package because they could be used directly by the users, avoiding search and download. Besides, two additional data sets were also included. All of them are available in the website of the Spanish Statistical Office (INE), <http://www.ine.es>. These four data sets are briefly described below:

- Firstly, the data set `ActivePeople` was considered as an example in order to illustrate the simple index numbers. It contains information separately on the number (thousands) of economically active women and men in Spain from the first trimester in 2002 to the fourth one in 2019.
- Secondly, `ECResources` is a data set containing as variables, the unitary value (euros) and consumption (thousands of units) of several combustibles and other energy resources for the

main home in Spain from 2006 to 2015. It was used in this paper for illustrating complex index numbers.

- An additional data set called Mortgages was also included in the package. In this case, the variables correspond to the number of mortgages constituted on urban properties in Spain from 2003 to 2018, distinguished between the kind of mortgages entities (banks, saving banks and other types). The corresponding mortgages amounts (thousands of euros) were also included as variables.
- Finally, the variables in the data set Food are the unitary value (euros) and consumed amount (thousands of units) of the main types of food in Spain from 2006 to 2015.

Once the package is installed and loaded, a full description of these data sets is shown through `help(ActivePeople)`, `help(ECResources)`, `help(Mortgages)`, and `help(Food)`, respectively.

### Functions in IndexNumber

**IndexNumber** package includes several functions that enable users to determine the index numbers, simple and complex (weighted and non-weighted), described in previous sections. The functions incorporated in the package are summarized in Table 12.

Function	Description
<code>aggregated.index.number</code>	Function to obtain several non-weighted index numbers: the Sauerbeck index number (3), the Geometric index number (4), the Harmonic index number (5) the Bradstreet-Dûtot index number (6), the Carli index number (7), the Jevons index number (8) and the Dûtot index number in (9).
<code>edgeworth.index.number</code>	Function to calculate the Marshall-Edgeworth index number (12).
<code>fisher.index.number</code>	Function to calculate the Fisher index number (13).
<code>index.number.chain</code>	Function to calculate the simple index number in chain (2).
<code>index.number.serie</code>	Function to calculate the simple index number in series (1).
<code>laspeyres.index.number</code>	Function to determine the Laspeyres index number (10).
<code>paasche.index.number</code>	Function to obtain the Paasche index number (11).

**Table 12:** Summary of functions in the **IndexNumber** package.

Users can obtain different kinds of index numbers by introducing the associated parameters in the corresponding function. Table 13 describes the different options to determine those index numbers whose implementation was included in **IndexNumber** package. However, not all of the mentioned options are required since only some of them are specific for each particular class of index number. Thus, Table 14 summarizes the arguments associated with each function. Examples of usage for the implemented functions are described in the next section.

Argument	Description
<code>x</code>	A matrix that contains the magnitude(s) under study. In each column, it contains the magnitude of a different product considered. Thus, we have <code>nrow(x)</code> values of a magnitude for <code>ncol(x)</code> products. Notice that if we intend to analyze a single magnitude, <code>x</code> corresponds to a vector of length equal to the total instants of time registered.
<code>y</code>	A matrix that contains that magnitude used as weight. In each column, it contains another magnitude associated to each different product along the time. Thus, we have <code>nrow(x)</code> values of magnitudes for the set of <code>ncol(x)</code> products. It is only required for obtaining those weighted index numbers mentioned in the paper.
<code>base</code>	A chain of characters that indicates the nature of the index number. If we introduce <code>base="serie"</code> , we compare each value with respect to the initial one. In this case, it is said to be an index number in series. Otherwise, if we introduce <code>base="chain"</code> , we obtain the index number in chain, by comparing each value with the immediately previous value.

type	A chain of characters to indicate the type of non-weighted index number to evaluate the evolution of a set of magnitudes (even for different products). By considering base="serie", if we introduce type="arithmetic", we obtain the Sauerbeck index number in (3). If we introduce type="geometric", we obtain the Geometric index in (4). If we choose type="harmonic", we obtain the Harmonic mean index in (5). If we write type="BDutot", we will obtain the Bradstreet-Dûtot index in (6). This argument is only required in the function aggregated.index.number. Otherwise, if we take base="chain" and type="Carli", we obtain the Carli index number in (7). If we introduce type="Jevons", we obtain the Jevons index in (8) and if we choose type="Dutot", we obtain the Dûtot index in (9). This argument is only required in the function aggregated.index.number.
name	A chain of characters to indicate the name of the variable under study.
opt.plot	A Boolean variable that indicates if a graphical description of the index number along the different stages is required. If it is desired, opt.plot=TRUE, else opt.plot=FALSE.
opt.summary	A Boolean variable that indicates if a basic statistical summary of the index number is required. If it is desired, opt.summary=TRUE, else opt.summary=FALSE.

**Table 13:** Summary of arguments for functions in the **IndexNumber** package.

Function	x	y	base	type	name	opt.plot	opt.summary
aggregated.index.number	✓	-	✓	-	✓	✓	✓
edgeworth.index.number	✓	✓	-	-	✓	✓	✓
fisher.index.number	✓	✓	-	-	✓	✓	✓
index.number.chain	✓	-	-	-	✓	✓	✓
index.number.serie	✓	-	-	-	✓	✓	✓
laspeyres.index.number	✓	✓	-	-	✓	✓	✓
paasche.index.number	✓	✓	-	-	✓	✓	✓

**Table 14:** Arguments for each function in the **IndexNumber** package.

### Examples of using IndexNumber

In what follows, we describe several examples of the application of the **IndexNumber** package that is used to illustrate its performance. Initially, a user has to incorporate the package from the CRAN in the R Console. After its installation, the next code allows its usage:

```
> library("IndexNumber")
```

Below, it is shown how to use the different functions implemented for determining index numbers on the real data sets presented in the preliminaries section.

### Simple index numbers in series in R

The example that we consider describes the obtaining of the simple index numbers in series (1) in R software on the real data partially given in Table 1. Remember that it depicts the number (thousands) of economically active women and men in Spain. As we mentioned before, this information is also included in the data set ActivePeople in **IndexNumber** package. The first trimester of 2002 is considered as the reference value.

Using index.number.serie() function, we obtain the simple index number in series for the first instants of time in the example.

```
> index.number.serie(ActivePeople$TotalWomen[1:15], name="Woman", opt.plot=TRUE,
  opt.summary = TRUE)
```

```
Index number in serie
```

```
Summary
```

```
Min.=101.855509425343
```

```
Stage=1
```

```

Max.=117.978690528975

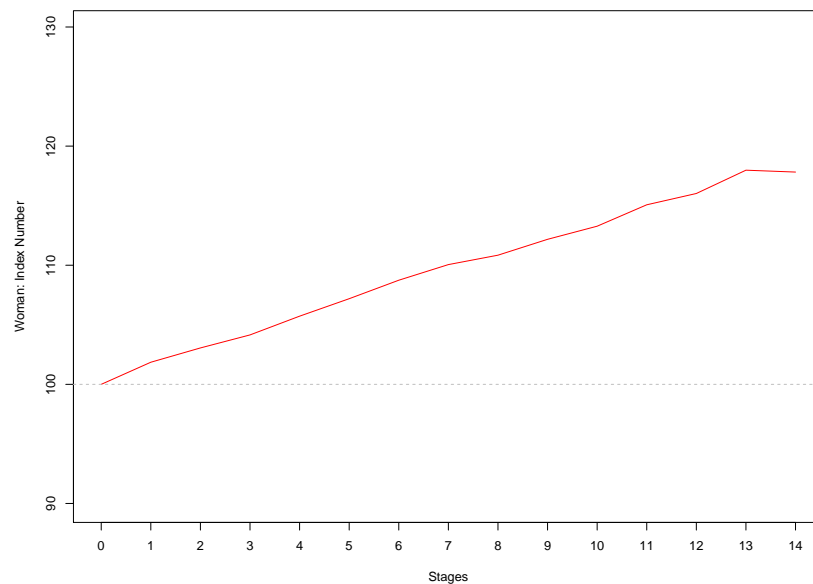
Stage=13

$Summary
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
101.9  106.1   110.4   110.3   114.6   118.0

$`Index number`
  Stages Woman Index number
1      0 7442.7    100.0000
2      1 7580.8    101.8555
3      2 7670.2    103.0567
4      3 7751.5    104.1490
5      4 7868.7    105.7237
6      5 7977.8    107.1896
7      6 8093.3    108.7415
8      7 8190.9    110.0528
9      8 8249.7    110.8428
10     9 8348.9    112.1757
11    10 8430.8    113.2761
12    11 8564.6    115.0738
13    12 8635.2    116.0224
14    13 8780.8    117.9787
15    14 8769.1    117.8215

```

We include a graphical summary of the evolution of this magnitude under the fixed criteria in Figure 3, by using `opt.plot=TRUE`. We also summarize the most relevant information in terms of the instant in which the maximum and minimum values are reached choosing `opt.summary=TRUE`.



**Figure 3:** Evolution of the simple index number in series for the number (thousands) of economically active women in Spain.

Analogously, the user can obtain the corresponding results for the data associated to TotalMen. The required code in this case is the shown below:

```
index.number.serie(ActivePeople$TotalMen[1:15],name="Man",opt.plot=TRUE,opt.summary = TRUE)
```

Results contained in Table 2 have been obtained using both functions of **IndexNumber** package in R.



## Simple index numbers in chain in R

Again, we take the data set `ActivePeople` to determine the corresponding simple index number in chain (2) for the number (thousands) of economically active women in Spain. Note that the reference value in each instant of time is the immediately previous one in the series.

Alternatively, we use `index.number.chain()` function for obtaining the simple index number in chain for the first instants of time for the variable considered in the example.

```
> index.number.chain(ActivePeople$TotalWomen[1:15], name="Woman", opt.plot=TRUE,
  opt.summary = TRUE)
```

```
Index number in chain
```

```
Summary
```

```
Min.=99.8667547376093
```

```
Stage=14
```

```
Max.=101.855509425343
```

```
Stage=2
```

```
$Summary
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
99.87 101.00 101.20 101.18 101.50 101.86
```

```
$`Index number`
```

```
  Stages  Woman  Index number
1      0 7442.7   100.00000
2      1 7580.8   101.85551
3      2 7670.2   101.17930
4      3 7751.5   101.05995
5      4 7868.7   101.51197
6      5 7977.8   101.38651
7      6 8093.3   101.44777
8      7 8190.9   101.20594
9      8 8249.7   100.71787
10     9 8348.9   101.20247
11    10 8430.8   100.98097
12    11 8564.6   101.58704
13    12 8635.2   100.82432
14    13 8780.8   101.68612
15    14 8769.1    99.86675
```

Also in this case, the option `opt.summary=TRUE` summarizes the most relevant information about the corresponding simple index number. The option `opt.plot=TRUE` provides a graphical representation of the evolution of the magnitude as Figure 4 depicts.

Table 3 also includes the numerical analysis of the evolution of economically active men in Spain. To determine these values, we use the following code:

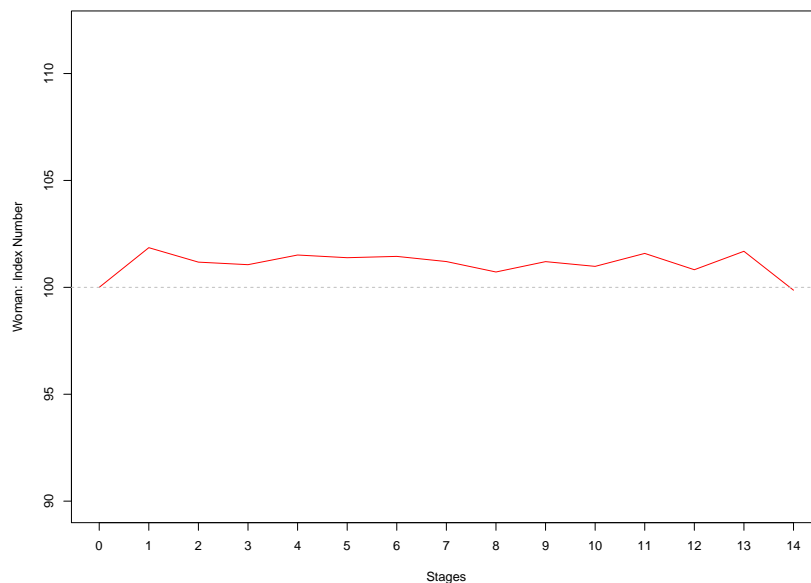
```
index.number.chain(ActivePeople$TotalMen[1:15], name="Man", opt.plot=TRUE, opt.summary = TRUE)
```

## Non-weighted complex index numbers in R

In this section, we illustrate the usage of `IndexNumber` to determine non-weighted complex index numbers.

As in its theoretical description, we take the example described in Table 6 to show its performance in practice. That table describes the unitary value of prices of several energy resources for the period 2005-2015. As we have mentioned, this data set is available in library `IndexNumber` under the name `ECResources` (in particular, from the second to the sixth column).

```
> ECResources[,2:6]
  ElectricityPrice NaturalGasPrice LiquefiedGasPrice LiquefiedCombustiblesPrice SolidCombustiblesPrice
```



**Figure 4:** Evolution of the simple index number in chain for the number (thousands) of economically active women in Spain.

1	0.14	0.70	1.00	0.69	0.12
2	0.14	0.78	0.97	0.66	0.10
3	0.15	0.83	1.03	0.81	0.11
4	0.16	0.87	0.93	0.68	0.10
5	0.17	0.79	1.04	0.72	0.12
6	0.19	0.77	1.15	0.91	0.12
7	0.22	0.82	1.19	0.96	0.12
8	0.23	1.00	1.34	0.99	0.13
9	0.24	1.07	1.35	0.89	0.15
10	0.25	1.04	1.22	0.77	0.13

First, we describe the R procedures that provide the variations of a magnitude in time  $t$  with respect to the base period by using non-weighted index numbers. For this purpose, we have to introduce the option `base="serie"`. Notice that all the values included in Table 8 were obtained by using the code in R that we show in the current section.

#### Sauerbeck index in R

Next, we determine the Sauerbeck index (3) in R software. To this aim, we use the corresponding function `aggregated.index.number()` by adding, as option, `type="arithmetic"`. Recall that it corresponds to an average by stages. In this case, we also include a graphical description of the joint evolution of prices in Figure 5 with `opt.plot=TRUE`, and a numerical summary of such magnitude (with `opt.summary=TRUE`).

```
> aggregated.index.number(ECResources[,2:6],base="serie",type="arithmetic",
                           name="Prices",opt.plot=TRUE,opt.summary=TRUE)
```

```
Aggregate index number
```

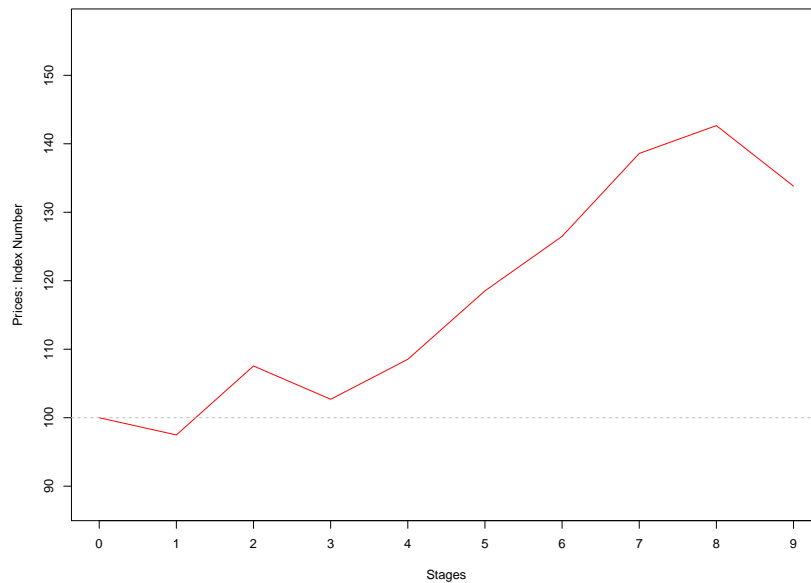
```
Arithmetic
```

```
Summary
```

```
Min.=97.4828157349897
```

```
Stage=1
```

```
Max.=142.654244306418
```



**Figure 5:** Sauerbeck index evolution for the prices of several energy resources for the period 2005-2015.

Stage=8

\$Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
97.48	107.55	118.52	119.59	133.81	142.65

\$^Agg. index number^

	Stages	Prices 1	Prices 2	Prices 3	Prices 4	Prices 5	Agg. index number
1	0	0.14	0.70	1.00	0.69	0.12	100.00000
2	1	0.14	0.78	0.97	0.66	0.10	97.48282
3	2	0.15	0.83	1.03	0.81	0.11	107.55445
4	3	0.16	0.87	0.93	0.68	0.10	102.69110
5	4	0.17	0.79	1.04	0.72	0.12	108.52671
6	5	0.19	0.77	1.15	0.91	0.12	118.51967
7	6	0.22	0.82	1.19	0.96	0.12	126.48323
8	7	0.23	1.00	1.34	0.99	0.13	138.59089
9	8	0.24	1.07	1.35	0.89	0.15	142.65424
10	9	0.25	1.04	1.22	0.77	0.13	133.81408

**Geometric mean index in R**

The second approach that we consider is the one given by the Geometric mean index (4). To obtain for the case of prices of the energetic resources, we slightly change the parameters of the function aggregated.index.number(). We have to introduce the parameter type="geometric" on it.

```
aggregated.index.number(ECResources[,2:6],base="serie",type="geometric",
                        name="Prices",opt.plot=FALSE,opt.summary=FALSE)
```

The results have the same structure as the previous case that we have explained. For this reason, they have not already been included here.

**Harmonic mean index in R**

The third option of a non-weighted complex index is the Harmonic mean index (5). With regard to the previous cases, the main difference is the parameter type that, in this case, has to take the value "harmonic".

```
aggregated.index.number(ECResources[,2:6],base="serie",type="harmonic",
                        name="Prices",opt.plot=FALSE,opt.summary=FALSE)
```

The scheme of showing the results maintains also in this scenario.

### Bradstreet-Dûtot index in R

The Bradstreet-Dûtot index (6) is determined in R software by using `aggregated.index.number()` with the parameter `type="BDutot"`. We illustrate the case of obtaining the indicated index for the prices of energetic resources. Again, the output has the same structure as the previous cases.

```
aggregated.index.number(ECResources[,2:6],base="serie",type="BDutot",
                        name="Prices",opt.plot=FALSE,opt.summary=FALSE)
```

Secondly, we describe examples of usage of **IndexNumber** package that involves non-weighted and weighted index numbers in chain. Specifically, we show those ones required for obtaining the results in Table 9. They involves the usage of `base="chain"`.

### Carli index in R

The Carli index (7) is obtained in R software through `aggregated.index.number()` with `type="Carli"`. The following R code determines the indicated index for the prices of energetic resources with an analogous structure for the output.

```
aggregated.index.number(ECResources[,2:6],base="chain",type="Carli",
                        name="Prices",opt.plot=FALSE,opt.summary=FALSE)
```

### Jevons index in R

The Jevons index (8) can be determined in R software by using `aggregated.index.number()` with the parameter `type="Jevons"`. Again, we illustrate the case of obtaining the indicated index for the prices of energetic resources. The required code is the one displayed below:

```
aggregated.index.number(ECResources[,2:6],base="chain",type="Jevons",
                        name="Prices",opt.plot=FALSE,opt.summary=FALSE)
```

### Dûtot index in R

The Dûtot index (9) is determined in R with `aggregated.index.number()` with additionally including `type="BDutot"`. Finally, we illustrate the case of obtaining this index for the prices of energetic resources.

```
aggregated.index.number(ECResources[,2:6],base="chain",type="Dutot",
                        name="Prices",opt.plot=FALSE,opt.summary=FALSE)
```

### Weighted complex index numbers in R

Next, we enumerate the capabilities of **IndexNumber** package in R software to determine weighted complex index numbers. The results in Table 11 were also obtained by using the functions of R that we show below.

Again, we pretend to obtain the evolution of the unitary value of prices of several energy resources for the period 2005-2015. However, in this case, we weight their values by the total amount of consumed energy resources given in Table 7. This information is also included in the data set `ECResources` of the package **IndexNumber**.

```
> ECResources[,7:11]
  ElectricityConsumed NaturalGasConsumed LiquifiedGasConsumed LiquifiedCombustiblesConsumed SolidCombustiblesConsumed
1          50623635          3617285          1057488          2297923          1306920
2          51990501          3266575          1066857          2454265          1602799
3          54990338          3473851          1210607          2274326          1556673
4          59749470          3730349          1113642          2505711          1724222
5          69751162          3954065          987112          2345215          1584123
6          67574654          4466072          926824          1974662          1414234
7          62878557          4576052          943632          2029733          1733591
8          56017871          4116079          867695          1952593          2071152
9          49177739          3653055          868743          2180866          2077766
10         48541712          3795339          818183          2176533          2161208
```

### Laspeyres index in R

Using weights, the first alternative that we describe is devoted to obtain the Laspeyres index (10) in R. **IndexNumber** package allows this through `laspeyres.index.number()` by adding these as parameters: the matrix of the magnitudes to be evaluated, the matrix containing the weights, and several options of graphical and numerical representation for the results.

```
> laspeyres.index.number(ECResources[,2:6],ECResources[,7:11],
  name="Price",opt.plot=TRUE,opt.summary=TRUE)
```

Laspeyres index number

Summary

Min.=101.309108829536

Stage=1

Max.=158.535309198466

Stage=8

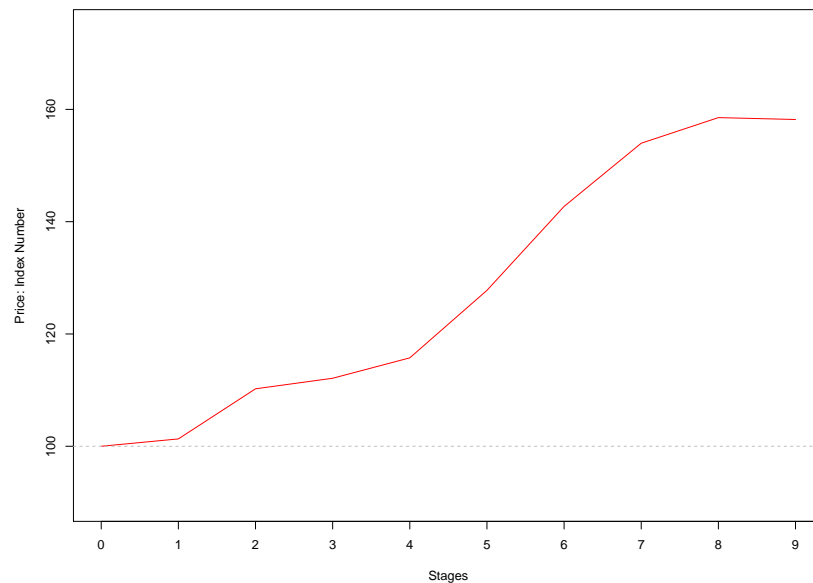
\$Summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
101.3	112.1	127.8	131.2	154.0	158.5

\$^Agg. index number^`

Stages	Price 1	Price 2	Price 3	Price 4	Price 5	Agg. index number	
1	0	0.14	0.70	1.00	0.69	0.12	100.0000
2	1	0.14	0.78	0.97	0.66	0.10	101.3091
3	2	0.15	0.83	1.03	0.81	0.11	110.2332
4	3	0.16	0.87	0.93	0.68	0.10	112.1124
5	4	0.17	0.79	1.04	0.72	0.12	115.7457
6	5	0.19	0.77	1.15	0.91	0.12	127.7677
7	6	0.22	0.82	1.19	0.96	0.12	142.7184
8	7	0.23	1.00	1.34	0.99	0.13	153.9749
9	8	0.24	1.07	1.35	0.89	0.15	158.5353
10	9	0.25	1.04	1.22	0.77	0.13	158.2000

As before, if the option `opt.plot=TRUE` is considered, the output of the function also includes a graphical representation in which the joint evolution can be analyzed as Figure 6 depicts.



**Figure 6:** Laspeyres index evolution for the prices of several energy resources for the period 2005-2015.

### Paasche index in R

Analogously, the joint evolution of magnitudes under the Paasche index (11) can be also determined by using `IndexNumber` package in R software. More specifically, `paasche.index.number()` provides

the mentioned weighted index number. The results of the function follows a similar structure as the function previously mentioned, and using the same graphical and numerical options.

```
paasche.index.number(ECResources[,2:6],ECResources[,7:11],
                    name="Price",opt.plot=TRUE,opt.summary=TRUE)
```

### Marshall-Edgeworth index in R

The obtaining of the Marshall-Edgeworth index number (12) is also possible in R software through the use of **IndexNumber** package. In particular, we have to use the function `edgeworth.index.number()` with the above-mentioned options for obtaining graphics and summaries.

```
edgeworth.index.number(ECResources[,2:6],ECResources[,7:11],
                      name="Price",opt.plot=FALSE,opt.summary=FALSE)
```

### Fisher index in R

Here, we consider the case of determining the Fisher index (13). In this case, `fisher.index.number()` provides a measure of the considered magnitude. Again, the options `opt.plot` and `opt.summary` allow the obtaining of additional information that may be of interest to the user.

```
fisher.index.number(ECResources[,2:6],ECResources[,7:11],
                   name="Price",opt.plot=FALSE,opt.summary=FALSE)
```

### A complex index for $v$ in R

The complex index for  $v$  given in (14) can be easily obtained as the Bradstreet-Dûtot index for the value in each instant of time. Recall that the value is obtained as the product of the amount of good by its price in each instant of time. Thus, the function `aggregated.index.number()` provides the value of this new index number as follows.

```
aggregated.index.number(ECResources[,2:6]*ECResources[,7:11],
                       base="serie",type="BDutot",name="Prices",opt.plot=FALSE,
                       opt.summary=FALSE)
```

## 4 Concluding Remarks

This paper discusses the implementation in R of classical index numbers used for comparing magnitudes mainly in economic contexts. Therefore, the **IndexNumber** package provides R users a set of functions to calculate index numbers. Concretely, this library allows the calculation of simple index numbers in series and in chain. Furthermore, complex index numbers are also implemented. In particular, the non-weighted index numbers included are the Sauerbeck, Geometric mean, Harmonic mean, Bradstreet-Dûtot, Carli, Jevons, and Dûtot indexes; as weighted index numbers, the Laspeyres, Paasche, Marshall-Edgeworth, Fisher, and Bradstreet-Dûtot indexes were considered. Additionally, this package contains graphical tools in order to facilitate the results visualization and four real data sets that can be used as illustrative examples. Moreover, the use of this library could be easily combined with other classical packages focused on time series analysis. Future research and development plans for forthcoming versions of the package include the addition of new index numbers already considered in the literature that can be dealt with in the framework presented above. Of course, the corresponding graphical tools should also be implemented.

## 5 Acknowledgments

A. Saavedra-Nieves acknowledges the financial support of FEDER/Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación under grant MTM2017-87197-C3-3-P. P. Saavedra-Nieves acknowledges the financial support of Ministerio de Economía y Competitividad of the Spanish government under grants MTM2016-76969-P and MTM2017-089422-P, and by the Xunta de Galicia through the European Regional Development Fund (Grupos de Referencia Competitiva ED431C-2017/38).



## Bibliography

- R. G. D. Allen. *Index numbers in theory and practice*. Springer, 1982. URL <https://doi.org/10.4324/9780203788677>. [p253]
- B. W. Ang, F. Liu, and H.-S. Chung. A generalized Fisher index approach to energy decomposition analysis. *Energy Economics*, 26(5):757–763, 2004. URL <https://doi.org/10.1016/j.eneco.2004.02.002>. [p254]
- W. A. Barnett. Economic monetary aggregates an application of index number and aggregation theory. *Journal of econometrics*, 14(1):11–48, 1980. URL [https://doi.org/10.1016/0304-4076\(80\)90070-6](https://doi.org/10.1016/0304-4076(80)90070-6). [p254]
- J. Bialek. Proposition of a general formula for price indices. *Communications in Statistics-Theory and Methods*, 41(5):943–952, 2012. URL <https://doi.org/10.1080/03610926.2010.533238>. [p254]
- J. Bialek. Simulation study of an original price index formula. *Communications in Statistics-Simulation and Computation*, 43(2):285–297, 2014. URL <https://doi.org/10.1080/03610918.2012.700367>. [p254]
- J. Bialek. *PriceIndices: Calculating Bilateral and Multilateral Price Indexes*, 2021. URL <https://CRAN.R-project.org/package=PriceIndices>. R package version 0.0.4. [p263]
- G. A. Boyd and J. M. Roop. A note on the Fisher ideal index decomposition for structural change in energy intensity. *The Energy Journal*, 25(1), 2004. URL <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol125-No1-5>. [p254]
- G. Boyle. The economic theory of index numbers: Empirical tests for volume indices of agricultural output. *Irish Journal of Agricultural Economics and Rural Sociology*, pages 1–20, 1988. URL <https://www.jstor.org/stable/25556589>. [p254]
- G.-R. Carli. Del valore e della proporzione dei metalli monetati. *Scrittori Classici Italiani Di Economia Politica*, 13:297–366, 1804. [p260]
- F. Coggshall. The arithmetic, geometric, and harmonic means. *The Quarterly Journal of Economics*, 1(1): 83–86, 1886. URL <https://doi.org/10.2307/1883111>. [p260]
- B. T. Coyle. Aggregation of price risk over commodities: An economic index number approach. *American Journal of Agricultural Economics*, 89(4):1085–1097, 2007. URL <https://doi.org/10.1111/j.1467-8276.2007.01023.x>. [p254]
- CPI Manual. *Consumer Price Index Manual: Theory and Practice*. OECD, UN, Eurostat, and The World Bank by ILO, Geneva, 2004. [p253]
- P. M. Danzon and L.-W. Chao. Cross-national price differences for pharmaceuticals: how large, and why? *Journal of health economics*, 19(2):159–195, 2000. URL [https://doi.org/10.1016/S0167-6296\(99\)00039-9](https://doi.org/10.1016/S0167-6296(99)00039-9). [p254]
- J. de Haan and E. Opperdoes. Estimation of the Coffee Price index Using scanner data: the choice of the Micro index. In *third meeting of the Ottawa Group*, pages 16–18, 1997. [p254]
- W. E. Diewert and A. O. Nakamura. The measurement of productivity for nations. *Handbook of econometrics*, 6:4501–4586, 2007. URL [https://doi.org/10.1016/S1573-4412\(07\)06066-7](https://doi.org/10.1016/S1573-4412(07)06066-7). [p254]
- Y. Dodge. *The concise encyclopedia of statistics*. Springer Science & Business Media, 2008. URL <https://doi.org/10.1007/978-0-387-32833-1>. [p253]
- M. Drobisch. Ueber Mittelgrößen und die Anwendbarkeit derselben auf die Berechnung des Steigens und Sinkens des Geldwerths. *Berichte der mathematisch-physicalischen Classe der königlich Sächsischen Gesellschaft der Wissenschaften*, 1, 1871. [p253]
- J. C. Dumagan. Comparing the superlative Törnqvist and Fisher ideal indexes. *Economics Letters*, 76(2): 251–258, 2002. URL [https://doi.org/10.1016/S0165-1765\(02\)00049-6](https://doi.org/10.1016/S0165-1765(02)00049-6). [p254]
- C. Dûtot. *Reflexions politiques sur les finances, et le commerce*, volume 2. les freres Vaillant [and] N. Prevost, 1754. [p253, 261]
- F. Y. Edgeworth. Measurement of change in value of money i. *First Memorandum presented to the British Association for the Advancement of Science. Reprinted in his Papers Relating to Political Economy*, 1: 198–259, 1887. [p262]

- S. G. Evelyn. An account of some endeavours to ascertain a standard of weight and measure. *Philos. Trans*, 113, 1798. [p253]
- I. Fisher. *The making of index numbers: a study of their varieties, tests, and reliability*, volume 1. Houghton Mifflin, 1922. URL <https://doi.org/10.2307/1883934>. [p253, 262]
- W. Fleetwood. *Chronicon Preciosum: Or, an Account of English Money: The Price of Corn, and Other Commodities, for the Last 600 Years. In a Letter to a Student in the University of Oxford*. Charles Harper, 1707. [p253]
- F. Forsyth and R. F. Fowler. The theory and practice of chain price index numbers. *Journal of the Royal Statistical Society: Series A (General)*, 144(2):224–246, 1981. URL <https://doi.org/10.2307/2981921>. [p254, 256]
- R. Frisch. Annual survey of general economic theory: The problem of index numbers. *Econometrica: Journal of the Econometric Society*, pages 1–38, 1936. URL <https://doi.org/10.2307/1907119>. [p253]
- A. Henningsen. *micEconIndex: Price and Quantity Indices*, 2017. URL <https://CRAN.R-project.org/package=micEconIndex>. R package version 0.1-6. [p263]
- R. J. Hill. Comparing price levels across countries using minimum-spanning trees. *Review of Economics and Statistics*, 81(1):135–142, 1999. URL <https://doi.org/10.1162/003465399767923881>. [p254]
- R. J. Hill. When does chaining reduce the Paasche–Laspeyres spread? An application to scanner data. *Review of Income and Wealth*, 52(2):309–325, 2006. [p254]
- R. J. Hill and D. Melser. Hedonic imputation and the price index problem: an application to housing. *Economic Inquiry*, 46(4):593–609, 2008. URL <http://doi.org/10.1111/j.1465-7295.2007.00110.x>. [p254]
- L. Ivancic, W. E. Diewert, and K. J. Fox. Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1):24–35, 2011. URL <https://doi.org/10.1016/j.jeconom.2010.09.003>. [p254]
- W. S. Jevons. *A Serious Fall in the Value of Gold Ascertained: And Its Social Effects Set Forth*. E. Stanford, 1863. [p253, 260, 261]
- W. S. Jevons. On the variation of prices and the value of the currency since 1782. *Journal of the Statistical Society of London*, 28(2):294–320, 1865. URL <https://doi.org/10.2307/2338419>. [p260]
- N. Kakwani and R. J. Hill. Economic theory of spatial cost of living indices with application to Thailand. *Journal of Public Economics*, 86(1):71–97, 2002. URL [https://doi.org/10.1016/S0047-2727\(00\)00174-2](https://doi.org/10.1016/S0047-2727(00)00174-2). [p254]
- M. Kendall. Studies in the history of probability and statistics, XXI. The early history of index numbers. *Revue de l'Institut International de Statistique*, pages 1–12, 1969. URL <https://doi.org/10.2307/1402090>. [p253]
- M. Kendall. *The early history of index numbers*, volume II. Studies in the History of Statistics and Probability, Griffin, London, 1977. URL <https://doi.org/10.2307/1402090>. [p253]
- R. Lamm. Index numbers and changes in food prices. *Agricultural Economics Research*, 32(2):41–43, 1980. [p254]
- E. Laspeyres. Hamburger Waarenpreise 1851–1863 und die californisch-australischen Goldentdeckungen seit 1848. *Jahrbücher für Nationalökonomie und Statistik*, 3(1):81–118, 1864. [p253]
- K. Laspeyres. IX. Die Berechnung einer mittleren Waarenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik*, 16(1):296–318, 1871. [p253, 261]
- J. Lowe. The Present State of England in Regard to Agriculture. *Trade & Finance*, 1822. [p253]
- A. Marshall. Remedies for fluctuations of general prices. *The Contemporary review, 1866-1900*, 51: 355–375, 1887. [p262]
- C. J. O'Donnell. Measures of Productivity Change. In *Productivity and Efficiency Analysis*, pages 93–143. Springer, 2018. URL <https://doi.org/10.1007/s11123-012-0311-1>. [p254]
- H. Paasche. Ueber die Preisentwicklung der letzten Jahre nach den Hamburger Börsennotirungen *Jahrbücher für Nationalökonomie und Statistik*, 23, 1874. [p253, 262]

- R. Palgrave. Currency and Standard of Value in England, France and India and the Rates of Exchange Between these Countries. *Memorandum submitted to the Royal Commission on Depression of Trade and Industry*, pages 312–90, 1886. [p253]
- M. B. Reinsdorf, W. E. Diewert, and C. Ehemann. Additive decompositions for Fisher, Törnqvist and geometric mean indexes. *Journal of Economic and Social Measurement*, 28(1-2):51–61, 2002. URL <http://dx.doi.org/10.3233/JEM-2003-0194>. [p254]
- A. Saavedra-Nieves and P. Saavedra-Nieves. *IndexNumber: Index Numbers in Social Sciences*, 2021. R package version 1.3.2. [p254]
- A. Sauerbeck. Index Numbers of Prices. *The Economic Journal*, 5(18):161–174, 1895. URL <https://doi.org/10.2307/2955755>. [p259]
- B. Su and B. Ang. Attribution of changes in the generalized Fisher index with application to embodied emission studies. *Energy*, 69:778–786, 2014. URL <https://doi.org/10.1016/j.energy.2014.03.074>. [p254]
- C. M. Walsh. *The measurement of general exchange-value*, volume 25. Macmillan, 1901. [p260]
- G. White. *IndexNumR: Index Number Calculation*, 2021. URL <https://CRAN.R-project.org/package=IndexNumR>. R package version 0.2.0. [p263]
- C. Zhen, E. A. Finkelstein, S. A. Karns, E. S. Leibtag, and C. Zhang. Scanner data-based panel price indexes. *American journal of agricultural economics*, 101(1):311–329, 2019. URL <http://doi.org/10.1093/ajae/aay032>. [p254]
- J. L. Zofío and A. M. Prieto. Return to dollar, generalized distance function and the Fisher productivity index. *Spanish Economic Review*, 8(2):113–138, 2006. URL <http://dx.doi.org/10.1007/s10108-006-9004-0>. [p254]

Alejandro Saavedra-Nieves  
Departamento de Estatística, Análise Matemática e Optimización  
Universidade de Santiago de Compostela  
Spain  
ORCID: 0000-0003-1251-6525  
[alejandro.saavedra.nieves@usc.es](mailto:alejandro.saavedra.nieves@usc.es)

Paula Saavedra-Nieves  
Departamento de Estatística, Análise Matemática e Optimización  
Universidade de Santiago de Compostela  
Spain  
ORCID: 0000-0002-3224-8858  
[paula.saavedra@usc.es](mailto:paula.saavedra@usc.es)