

## Box 1

# The Measurement Bias in the CPI: an Update

Edgar Caicedo G.  
Sergio Andrés Malagón P.  
Andrea Salazar D.\*

### 1. General Considerations

The National Bureau of Statistics (DANE) is planning a new change in the methodology used to calculate the consumer price index (CPI), effective as of 2019 when the weights assigned to the items that make up the basket of consumer goods and services will be reviewed and updated. This will be an opportunity to arrive at a more accurate estimate of the increase in the cost of living during the last ten years, correcting what is known in specialized literature as the substitution bias in consumption, which is the primary subject of this analysis.

For the past several decades, but particularly since 1996 when the Boskin Commission Report was published in the United States<sup>1</sup>, it has been acknowledged that the traditional approach to measuring the CPI (with the Laspeyres index) implies a variety of different biases that misrepresent the true cost of living.<sup>2</sup> The biases identified in that report are consumption substitution, quality-change in products, new products and new outlets.<sup>3</sup>

The consumption-substitution bias is the best known and most important, and perhaps the easiest to estimate. It occurs for two reasons. The first is because the weights now assigned to the basket of consumer goods and services fail to reflect the current structure of household spending, having been established for a given period of years, and

without the possibility of including, in the CPI computation, the dynamics of household consumption in response to changes in relative prices. The second reason is that the CPI cannot be calculated with a functional form that approximates it to the true cost of living index.

International organizations such as the International Monetary Fund (IMF, 2006) have developed specific recommendations to reduce the influence of this bias on the CPI. There also are several works in international literature that quantify this and the other biases in computation of the CPI. See the work of Lizardi (2008), Boldsen (2011), Kalish (2017) and Karsaulidze (2018).

It is recommended that a superlative index (e.g., the Fisher, Walsh or Törnqvist-Theil indexes) be used to estimate the substitution bias and to calculate a cost of living index that is closer to reality. The difference between the official CPI (usually calculated with the Laspeyres index) and one of the superlative indices mentioned above is the substitution bias for goods and services.

The IMF (2006) suggests Fisher's ideal index can provide a satisfactory approximation to the unobservable index of the true cost of living: "More and more economists and users conclude that, in principle, the preferable, ideal index for the purposes of the CPI would have to be a superlative index, such as Fisher's. This idea is reinforced by the fact that Fisher's index also is a very attractive one from an axiomatic standpoint." In response to that recommendation, we have adopted this superlative index number for our analysis, as a satisfactory proxy for the true cost of living index.<sup>4</sup>

In this box, we will estimate the substitution bias for the upper level of the CPI (from basic spending upwards), because the lower level (found within each basic expenditure) cannot be approximated, as it is the statistical reserve of DANE. Moreover, by applying the geometric mean (functional form that tends to approximate a superlative index) to calculate up to the basic expenditure level in the CPI, DANE minimizes the substitution bias.<sup>5</sup>

Similarly, with the current CPI methodology (December 2008 base year = 100), DANE significantly reduced the other biases by making all goods and services that are part of the basic expenditure in the consumer basket flexible. The possibility of including and excluding references or varieties of new products, from the basic expense downward, reduces the quality-change bias. Also, the new sources of information that are incorporated or replaced reduce the bias for new outlets. Moreover, the possibility of including new items at this level reduces the new-product bias.

\* The authors are, respectively, a senior professional and student interns with the Programming and Inflation Department. The opinions and possible errors or omissions in this document are not binding on *Banco de la República* or its Board of Directors.

1 See <https://www.finance.senate.gov/imo/media/doc/Prt104-72.pdf>

2 A consumer price index quantifies the variation in the cost of a particular "basket" of consumer goods and services, while a cost of living index measures the change in the cost of maintaining a certain standard of living or utility.

3 For a more detailed analysis of the different biases, see Karsaulidze (2018) and FMI (2006, Chapter 11).

4 An axiomatic analysis of superlative indices can be found in IMF (2006), Chapter 16.

5 In this respect, see IMF (2006, p. 249).

## 2. Calculating the Substitution Bias in Colombia

As mentioned, the substitution bias in the upper level (which goes from basic to total expenditure) is the difference between the official DANE CPI and the cost of living index. The latter is approximated by a superlative index, which in our case will be Fisher's ideal index ( $Fp_{t,0}$ ),<sup>6</sup>

where,  $(Fp_{t,0}) = [Lp_{t,0} \times Pp_{t,0}]^{1/2}$

$(Fp_{t,0})$  is the geometric mean of the product of the Laspeyres ( $Lp_{t,0}$ ) and Paasche ( $Pp_{t,0}$ ) indexes, so its path will be found in the middle of these two indexes. The theory states, precisely, that the true cost of living follows a path between ( $Lp_{t,0}$ ) and ( $Pp_{t,0}$ ); so, Fisher's ideal is a good approximation to the cost of living index.

The Laspeyres index is defined as:

$$\begin{aligned} Lp_{t,0} &= \sum_{i=1,n} (P_i^t Q_i^0 / \sum_{i=1,n} P_i^0 Q_i^0) \\ &= \sum_{i=1,n} [\alpha_i^0 (P_i^t / P_i^0)] \end{aligned}$$

By having fixed weights ( $\alpha_i^0$ ), this index tends to overestimate (but not always)<sup>7</sup> the true cost of living, since its functional form does not allow households to maximize their utility by replacing expensive goods with inexpensive products when it comes to their consumption.

For its part, the Paasche corresponds to:

$$\begin{aligned} Pp_{t,0} &= \sum_{i=1,n} (P_i^t Q_i^t / \sum_{i=1,n} P_i^0 Q_i^t) \\ &= 1 / \sum_{i=1,n} [\alpha_i^t (P_i^0 / P_i^t)] \end{aligned}$$

The Paasche index has a formula that uses weights in the current period ( $\alpha_i^t$ ). This allows consumers to substitute products in response to changes in relative prices, with a downward calculation in the CPI.

To estimate the dynamic weights ( $\alpha_i^t$ ), we used the weight structure published by DANE for the CPI base year 2008 = 100. Subsequently, with a linear progression, they were transformed into the weights obtained from the National

Household Budget Survey (ENPH in Spanish) for 2016-2017<sup>8</sup> (footnote no. 7 and Table B1.1).

## 3. Results

Graph B1.1 summarizes the behavior of the different index numbers calculated between December 2008 and September 2018.<sup>9</sup> The results corroborate what the theory and the empirical evidence indicate; namely, the Laspeyres index<sup>10</sup> tends to behave as an upper limit of the true cost of living, while the Paasche index generally follows a path below the true cost of living index.

Table B1.2, for its part, summarizes the calculation of the change in the substitution bias of the upper level of the CPI (from basic expenditure to the national total), which is estimated as the difference between the annual variation in the official CPI calculated by DANE (with the Laspeyres index) and the annual adjustment from Fisher's ideal index. According to the results, not having an updated consumption basket of goods and services and, hence, the impossibility of calculating the CPI with a superlative index, means the official CPI computed by DANE (Laspeyres) has been overestimating the true cost of living.

Additionally, the trend in the substitution bias was found to be positive. This indicates it has increased over the years, having gone from a bias of 0.10 percentage points (pp) in 2009 to a bias of 1.35 pp so far during 2018, with an average of 0.75 pp for the entire period under study.

## 4. Conclusions

In view of the foregoing, a complete update of all aggregate levels in the household basket of goods and services would eliminate, to a large extent and for a time, biases with respect to quality, outlets and new products. Updating the weight structure also would place DANE's official CPI calculation closer to the true cost of living index, by reducing the substitution bias.

6 Where  $P_i^t$  = price in the current period  
 $P_i^0$  = price in the base period  
 $Q_i^0$  = quantities in the base year  
 $Q_i^t$  = quantities in the current period  
 $\alpha_i^0$  = weights in the base period  
 $\alpha_i^t$  = weights in the current period

7 Sometimes, the Laspeyres numerical index is lower than the Paasche index. This occurs when the weighted variations in price and quantity correlate positively. When they are negatively correlated, the level of the Laspeyres index exceeds that of the Paasche index.

8 DANE recently published the results of this survey with the international classification of consumption according to purpose (COICOP). To do so, it was necessary to construct a correlation between the 2016-2017 ENPH and the weight structure of the CPI base December 2008 = 100. Once the CPI was constructed with the descriptions from the ENPH, each of the sub-categories resulting from that correlation was indexed up to September 2018, so as to ultimately obtain the definitive weights for the CPI, using the COICOP classification from the ENPH. This correlation involved eliminating several categories from the COICOP, since they did not correspond to the basic expenses included in the CPI. This left only eighty sub-categories. Due to space limitations, Table B1.1 only shows the subsequent weights for the total, food and non-food CPI. It should be noted that these weights may be very different from DANE's definitive weights as of 2019.

9 This is the period corresponding to the calculation methodology for the current basket of goods consumer and services.

10 The Laspeyres index corresponds exactly to that observed and published by DANE.

**Table B1.1**  
Weights Used to Calculate the Index Numbers

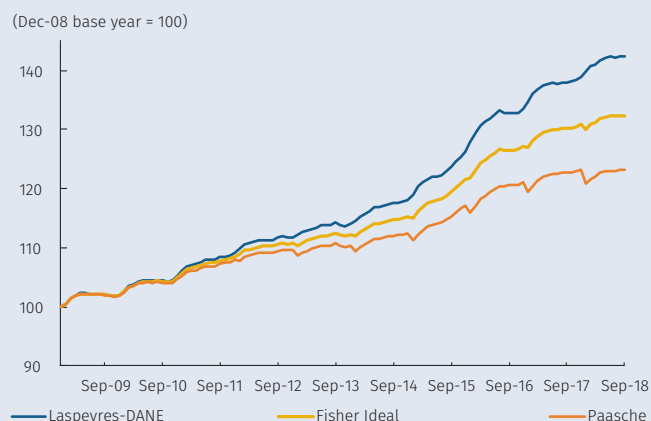
Description	2009 <sup>a/</sup>	2010	2011	2012	2013	2014	2015	2016	2017	2018 <sup>b/</sup>
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Food	28.2	28.0	27.9	27.7	27.5	27.3	27.1	27.0	26.8	26.6
Non-food	71.8	72.0	72.1	72.3	72.5	72.7	72.9	73.0	73.2	73.4

a / The weights for 2009 pertain to the official CPI, Base December 2008 = 100.

b / The weights for 2018 are from the 2016-2017 National Household Budget Survey. They were indexed with the respective CPI up to September 2018.

Source: DANE; authors' calculations.

**Graph B1.1**  
Total Consumer Price Indexes  
(CPI – survey weighted)



Source: DANE; authors' calculations.

Last year, the official figure for annual inflation (with the Laspeyres index) was 4.09%, while the cost of living index (approximated by Fisher’s ideal index) closed out the year at 2.93%. Apparently, the Laspeyres index overestimated 2017 inflation by 1.16 pp, a level slightly below the long-term target for inflation (3.0%).

On the other hand, the fact that the path of the substitution bias follows a positive slope suggests that calculating the CPI with household consumption patterns from years back tends to result in overestimates of inflation. So, it would be ideal if DANE were to update the weight structure of the CPI more frequently. In an effort to prevent an accumulation of upward biases in the calculation of the official CPI, several countries have been updating their CPI weights every year: Italy, Spain, England, Japan and the Netherlands are examples. The United States, on the other hand, does so twice a year.

**Table B1.2**  
Total Inflation and Substitution Bias  
(Percentage points)

Total	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018 <sup>a/</sup>	Average
Laspeyres-DANE	2.00	3.17	3.73	2.44	1.94	3.66	6.77	5.75	4.09	3.07	3.66
Fisher Ideal	1.91	3.06	3.38	1.97	1.35	2.74	5.46	4.57	2.93	1.72	2.91
Paasche	1.81	2.94	3.03	1.50	0.77	1.83	4.17	3.40	1.78	0.39	2.16
Bias	0.10	0.11	0.35	0.47	0.59	0.92	1.31	1.18	1.16	1.35	0.75

a / Corresponds to year-to-date inflation at September 2018.

Source: DANE; authors' calculations.

It is important to improve the production of the CPI in Colombia on a regular basis, since this indicator has broad implications for the country's economy. For example, letting biases in the CPI measurement accumulate outsize inflation; these higher levels affect the growth of many real economic variables and lead to indexing for salaries, leases, public utilities, taxes and balance sheets, among other items. Likewise, oversized inflation anchors expectations to high values, reduces the credibility of monetary policy management, and distorts decision-making on the part of economic agents.

## References

- Boldesen, C. (2011). Session 4: Calculation of Elementary Price Indices. UNECE Workshop on Consumer Price Indices. Istanbul, Turkey.
- International Monetary Fund -IMF- (2006). Manual del Índice de Precios al Consumidor: Teoría y Práctica. Washington, International Monetary Fund.
- Kalisch, D. (2017). Introduction of the 17th Series Australian Consumer Price Index. Australian Bureau of Statistics.
- Karsaulidze, L. (2018). Possible Reasons of Bias in Estimating the Cost of Living Index by the CPI. National Statistics Office of Georgia. Geneva, Switzerland.
- Lizardi, C. (2008). Sesgos de Medición del Índice Nacional de Precios al Consumidor, 2002-2007. *Investigación Económica*, 67(266), 37-65.