# Measuring inflation in Argentina: IPC Online Project.

Gonzalo Román Ramírez Muñoz de Toro[1,2], Uriarte Juan Ignacio[1], Larrosa Juan MC[3,4] and Gutiérrez Emiliano[3,4].

[1] Hyperia Big Data
[2] Departamento de Derecho, Universidad Nacional del Sur
[3] Departamento de Economía, Universidad Nacional del Sur
[4] Instituto de Investigaciones Económicas y Sociales del Sur (IIESS- CONICET)

**Extended Abstract**

Since 2007 Argentine political administration intervened the national statistical office named Instituto Nacional de Estadísticas y Censos (INDEC), a highly respected institution at that time [1]. The purpose was not clear at the beginning but soon after removing senior professional staff, inflation figures publicly released were suspiciously low. People began to distrust official information on this and other national statistic measurements as perceived variations were higher than reported ones.

By September 2014 is officially launched *IPC Online* (Consumer Price Index Online) and began to make public all the information on local inflation. The quality of the estimation could be directly tested given that a private institution was applying long before a local CPI based on the traditional method of personal survey.

The procedure for registering the CPI has customarily included a cautious outline of price lists, certainly guided by the information of the technical reports from INDEC and adapting them to local environment (for instance, subway tickets cannot be considered in a city that lacks that service) to gather price information. Web-scraped data indexes of prices have a tendency to be significantly bigger and considerably messier than their hand-picked counterparts, up to the degree that human preparing and categorizing information are unreasonable. The mechanization of these assignments from prices to categories concordant with those of the national CPI is subsequently a noteworthy segment of the present project. The categorization was done by using artificial intelligence algorithms that recognize common words and presentations using a "bag-of-words" machine learning algorithm to classify products into categories. Even doing so, most data collected could not be processed entirely. Approximately 200 thousand prices are collected each month but *only* 50 thousand are processed in the CPI.

For the other hand since 1996, Bahia Blanca has a private institution that estimates local inflation by using traditional survey of basket of goods and services. The Centro Regional de Estudios Económicos Bahía Blanca Argentina (CREEBBA) [Bahia Blanca Argentina Economic Studies Regional Center] performs the task of measuring inflation in a monthly basis in a period of highly stable prices in Argentina. The methodology implemented is a scaled-down variant of INDEC by using no more than

four surveyors that gather prices in main supermarkets and middle retail business and also recollect information on services prices. An average basket comprehends 800 items and prices are collected once a month in a hand-picking fashion.

As observed in figure 1, inflation measured by INDEC show lower rates in the period previous to normalization compared to CREEBBA and IPC Online, and figures were catching up closely in the posterior period after April 2016.
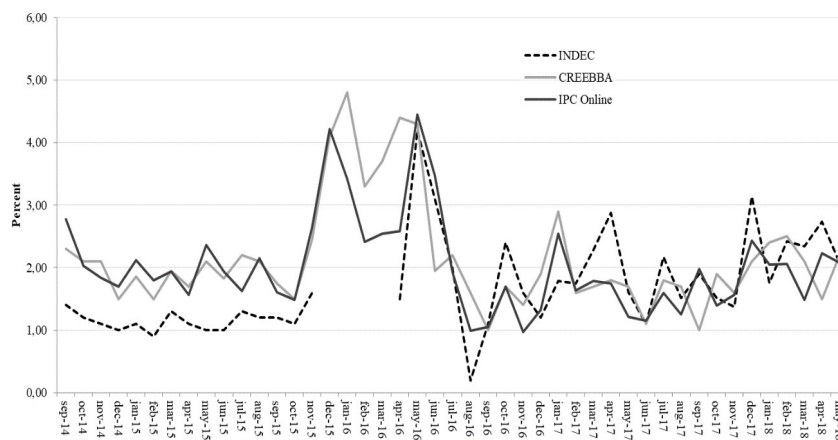


**Fig. 1.** Inflation during September-2014 to May-2018

Another milestone was achieved in November 2015 when webscraping frequency was augmented to four times per month. Each week (average weighted price) is compared against the full CPI of the previous month. As each week advances during the month, *IMWAV* adds the new one (average weighted price set) and, again, compares it against the full CPI of the previous month. This way, we can observe how monthly inflation is constructed up to the 4th week were the average of the *IMWAV* is equivalent to a monthly average compared to previous month. So W4 of the *IMWAV* represents monthly inflation.

The *IMWAV* ($P_{t,i}$) of week $i$ in month $t$ is defined by (1):

$$IMWAV\,(P_{t,i}) = \left[\left(\prod_{k=1}^{i} P_k\right)^{1/i} \frac{1}{P_{t-1}}\right] \qquad (1)$$

Where $P_{t-1}$ is CPI general level (full CPI) of month $t$-1, and $IMWAV\,(P_{t,4}) = P_t$ (i.e., the *IMWAV* accumulated to week 4th is the inflation of the month). Figure 2 shows the weekly accumulated inflation (IMWAV) for June 2018. It

is interesting to note that by June 14th peso depreciated by 6.5% against US dollar. Week 3 and 4 report how the exchange rate pass-through (ERPT) effect on the many chapters of the IPC Online. By slicing the month in weeks it is possible to observe when it took place, what category was more affected and how that effect lasted.
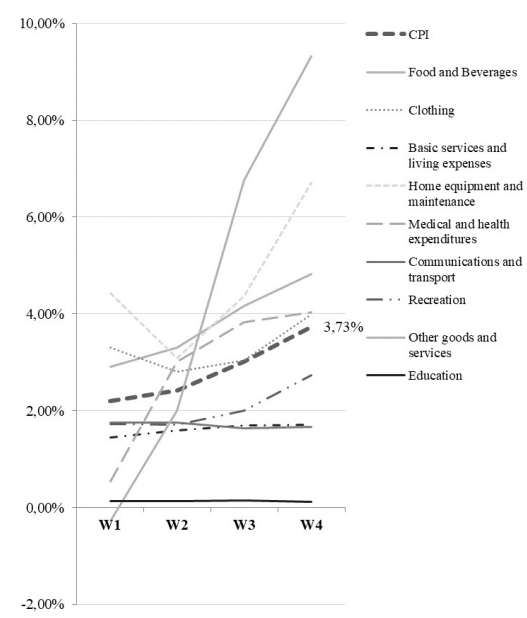


**Fig. 2.** IMWAV in June 2018

The motivation of the IPC Online, was to act and make accountable a policy that mis-informed economic agents. It allows for cross checking to highly important variable such as prices with different geographical scopes. It has been shown how webscrap-ing-data-based urban indices were estimated and contrasted to traditional-based indi-ces granting highly efficient measurements at one thousandth of cost. Combined with many other variables that can be captured or downloaded from the Internet [2] paints a future of combined indicators for tracking the socioeconomic life, even remotely.

**Referencies**

1. Cavallo, A. (2009), "Scraped Data and Sticky Prices: Frequency, Hazards, and Synchronization", Harvard University working paper; retrieved on 06/06/2018; http://www.people.fas.harvard.edu/~acavallo/papers/Cavallo-Scraped_Data-Paper.pdf
2. Hackl, P. (2016), "Big Data: What can official statistics expect?" Statistical Journal of the IAOS 32 (2016) 43–52 43. DOI 10.3233/SJI-160965