Prices or more specifically inflation is a very complex and broad topic in the world of economics. Throughout history we have seen a small sea of methods and new (alternative) forms of calculating inflation for which compliments and criticism have followed. In light of this we have been witnesses that within governments many debates have occurred on the topic and what is better? A few examples on this part include the views of the Bureau of Labor Statistics, John Williams and David Ranson, naturally many more exist. However just from these three examples we can observe significant differences in the “final results”.

* How the data came to exist?

Argentina is one of the few countries in the world is undergoing “chronic” inflation. They have experienced high inflation for at least five decades. However, the last two decades have brought significant attention to the matter, which was a consequence that emerged from the suspicions and multiple accusations of governmental manipulation. Since 2007 several academic and private analysts have estimated that the actual CPI has been considerably higher than the one reported by the official series. Hence the following sources appeared:

General comment for all sources: given the definition of Big Data and its 3V conditions, all of the below qualify as Big Data sources – as they incorporate all three aspects.

* Verdadera – the first step in the “revolution”

*Description/Overview:* created in 2007 to provide alternative price indexes to the official ones in Argentina, since then has shifted focus to Venezuela

*Collection method and processing:* Argentina – online price collection/ web-scraping (HTML form – semi-structured Big Data); Venezuela – by volunteers in stores through an app specially designed to harmonize the data format; methodologies are explained.

*PRO’s* – provided the existence of an alternative (“a second opinion”); openness to a different type of collection method.

*CON’s* – pure dependency on sheer will of individuals, no control on the frequency of data collected or the amount of data.

* The Billion Prices Project – (by-product of Verdadera)

*Description/Overview:* established in 2008 as an academic initiative, their scope monitored daily price fluctuations of around 15 mil products, sold by 1000+ retailers in more than 50 countries.

*Collection method:* web-scraping from hundreds of online retailers, daily

*PRO’s* – publicly available, MIT and Harvard Business School backing it up, methodologies are explained in great detail and transparently communicated (and quite complex).

*CON’s* – no global overview, relatively “old news”; one possible reason for no new data is possibly the shift in interest of Verdadera to Venezuela.

* PriceStat – its existence also owed to Verdadera

*Description/Overview:* the first private company to use online prices to estimate daily inflation, as of 2012 Verdera is using their indexes, currently covering 23 countries.

*Collection method and processing:* web-scraping

*PRO’s* – published daily, they do it as a private company (you pay to get their analysis – gives us reassurance that their motives go beyond volunteer work – adds legitimacy in a capitalistic view).

*CON’s* – daily? (do we really need it?), no free access only preview of sections of analysis.

* Numbeo

*Description/Overview:* Created in 2009, the goal/motivation was to offer individuals a free database for personal use to enable them to compare costs of living around the world.

*Collection method and processing:* reliant on user manual inputs and authoritative sources (twice per year). Implemented automatic and semi-automatic filters through which they identify irregularities.

PRO’s: free access, mixed model (indicating potentially greater added value)

CON’s: Methodology only partially presented and lacking all additional explanations to be considered transparent, no official re-checking of the data; especially inputs from people can be misleading (or purposefully incorrect); the methodology excludes some “extreme data” (which in my opinion should be revised in detail before exclusion).

* Data accuracy?

From the point of data accuracy it is very difficult to argue this case.

Firstly, because within the official government institutions there are disagreements on how to calculate prices/inflation and it is only natural that some deviations and variability will occur when the data is web-scrapped and processed using different methods.

Second, price measurement is directly linked to supermarkets and their offers, which today presents an almost infinite number of items making it impossible to fixate the quality and quantity of a basket of goods.

One might argue this case by saying: all the data is accurate, but hot it was used is an issue. We could say that the accuracy of the methodology and processes is at question? However, in the case of Argentina this would be hard to defend as the best approximation to the official reports was to divide the online data by 3. Surprisingly, the differences (according to Alberto Cavallo), were only in the level of inflation, not it’s dynamic.

Also, a statement of mild concern is worth noting regarding the accuracy of data being collected for Venezuela’s inflation as they are being collected by non-experts.

In conclusion, I would defend online price indexes (compilation of online and offline data). They show more promising value in accuracy rather than the government official ones, beyond doubt in the case of Argentina. Taking everything into consideration, my overall preferred source of choice would definitely be PriceStat.

* “a-ha” moment
* Verdadera switching its strategy to volunteers gathering data
* “wow” moment
* The magnitude of the difference and its persistence (Argentina)

Final remark: I find it amazing and inspirational, that information, knowledge and time proved to be no obstacle for these sources (organizations) to do the work that they have done.

I am personally very fond of “questioning the status-quo”, and as a third party that is only writing a homework about this magnificent work, their achievement has proven to be an invaluable alternative against the so far gold standard - government data.