Technical Note

Bread price index: a pilot case

Roberto Rigobon
MIT & NBER

November 2011

Introduction

This note presents the methodology used by PriceStats for the computation and analysis of the bread indices used in its joint research project with the United Nations' Global Pulse Initiative.[1]

The objective of this partnership was to investigate and show how scraping online prices could provide real-time insights on price dynamics, focusing on the case of bread.

We first provide a short description of the data gathering process, before discussing the actual computation of the indices.

Data and Data collection

PriceStats received the data from the Billion Prices Project (BPP) at MIT.

The data is collected using ascraping software that records, on a daily basis, the price of all goods sold or advertised online. This is done following a 3-step methodology:

1. The software identifies and downloads all public web pages where product and price information appear, for each retailer at a given time of the day. These pages are individually retrieved using the same URL or web-address every day;
2. It analyzes the underlying code and locates each piece of information that needs to be collected. This is performed on the basis of custom characters in the code that allow the software to match the format used on a particular page by a specific supermarket in order to identify the location as well as the starting and end points of the price information. For example, prices are often preceded by a dollar sign and followed by two decimals: in such a case this series of characters is the 'marker used by the scraping software to identify and record the price every day;
3. It stores the scraped variables in a database containing one observation per product-day. Along with the price and product characteristics, retailers show an ID for each product in the page's code (typically not visible when the page is displayed to the customer), which allows us to uniquely identify each product over time.

[1] This methods white paper arose from an on-going series of collaborative research projects conducted by the United Nations Global Pulse in 2011. Global Pulse is an innovation initiative of the Executive Office of the UN Secretary-General, which seeks to harness the opportunities in digital data to strengthen evidence-based decision-making. This research was designed to better understand where digital data can add value to existing policy analysis, and to contribute to future applications of digital data to global development. This project was conducted in collaboration with PriceStats and the Billion Prices Project at MIT. For more information on this project or the other projects in this series, please visit: http://www.unglobalpulse.org/research.

The retailers included in this pilot study are the biggest supermarkets in Argentina, Brazil, Chile, Colombia, Uruguay, and Venezuela. Prices were collected on a daily basis between October 2007 and July 2011, with different starting dates for each supermarket. On average, each supermarket offers 12,000 products everyday. The collection process focused exclusively on bread.

Bread indices

The computation of price changes is based on consecutive data points, i.e. it relies on data available for days T and T-1 (or T and T+1) with no interruption in the series.

The collection of high-frequency price information of every single product sold in each supermarket produces a great number of data points. At the same time, it also leads to many gaps in individual price series. In most cases these gaps occur when the scraping software fails or when individual items are temporarily out of stock. Scraping failures are typically resolved in a few days by the PriceStats scraping team, but seasonal products can create missing values that last several months. Some products are temporarily discontinued from stores while being substituted for another item—for example the case of iPad 1 and 2. Others are strictly seasonal (such as Christmas trees); and others may be temporarily out of stock while the store is being revamped or reorganized. It is also possible that the observation is missing because the product is being reclassified and the scraping simply fails. All these reasons will produce a "missing" observation that is not the reflection of a supply disruption.

The standard treatment suggested by the literature is to fill missing prices with the last recorded price available for each product. Fortunately, bread (and for that matter all staple foods) is regularly sold in supermarkets and the gaps tend to be very short or non-existent. Bread is not seasonal; it is never "reclassified", and the scrape jobs are stable and rarely fail. Therefore, a missing observation is very likely indicative of a supply problem, especially if the missing observation lasts for a significant period of time. This is important because our methodology relies on using these gaps and price hikes as the detection mechanism for the supply disruption.

We subsequently calculated a daily inflation rate of bread for each country, as follows:

1. We estimated the average daily inflation rate for each of the bread products in our data, by country;
2. We then constructed an index aggregating those daily changes, by country. We treated every bread product equally and took a simple geometric average of their daily price inflations;
3. All the inflation indices were normalized to 1 on the first day data was available.

The results are shown below:

Figure 1:

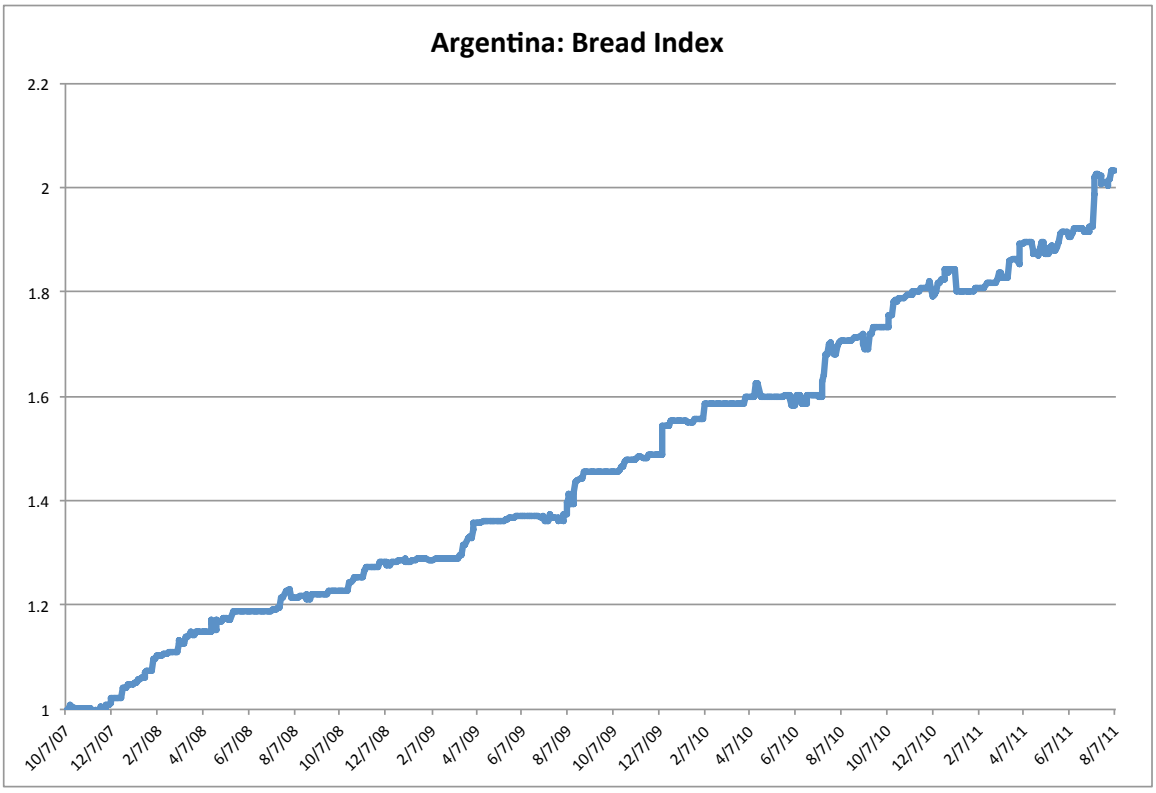


Figure 2:

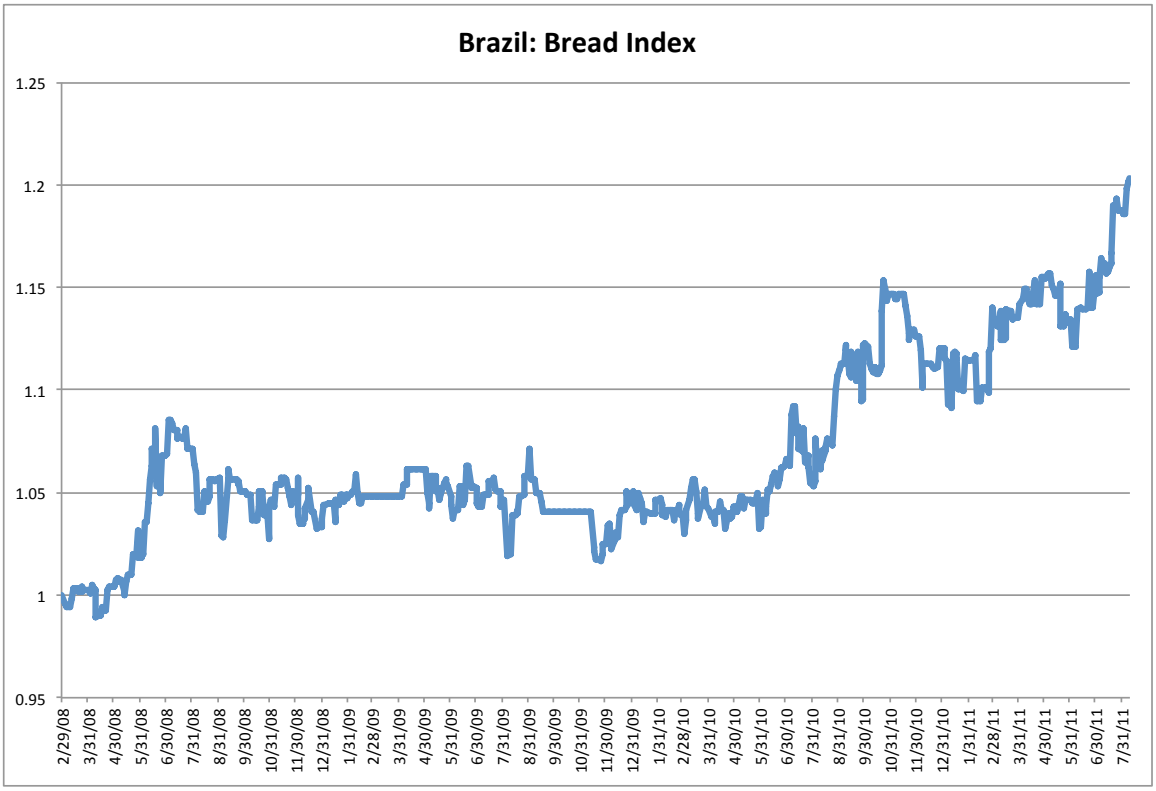


Figure 3:

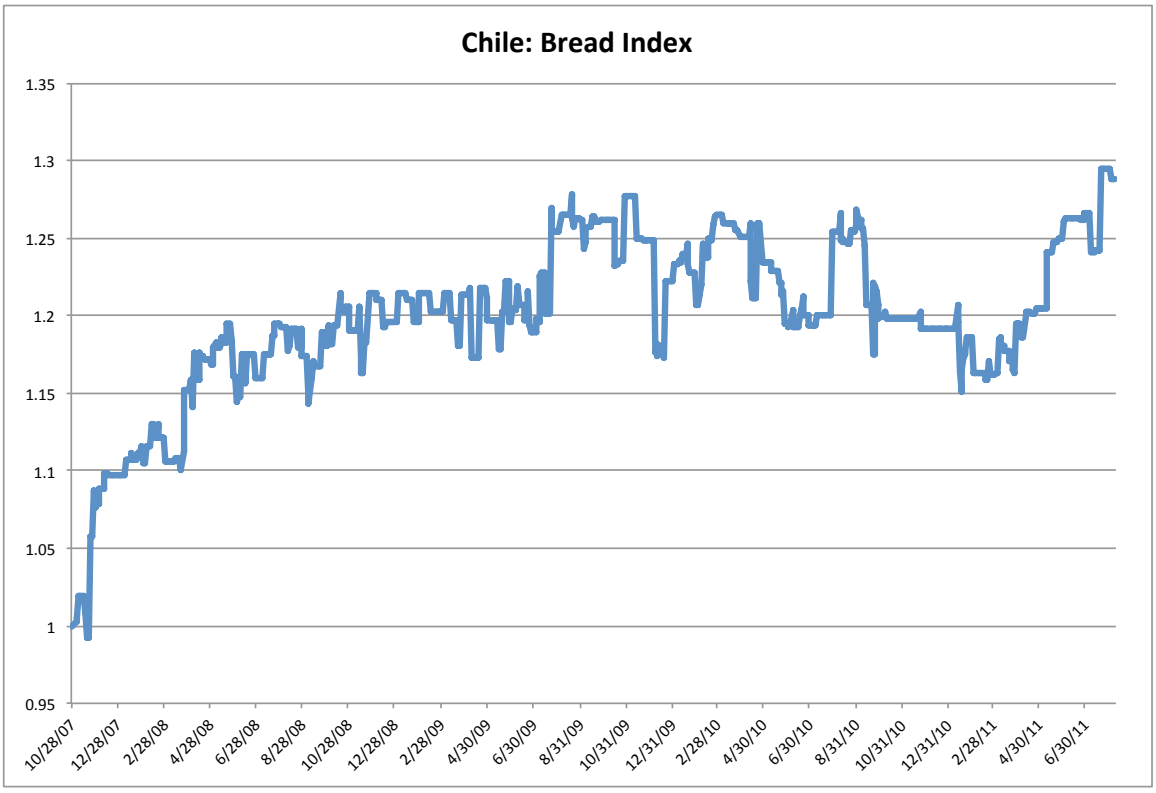


Figure 4:

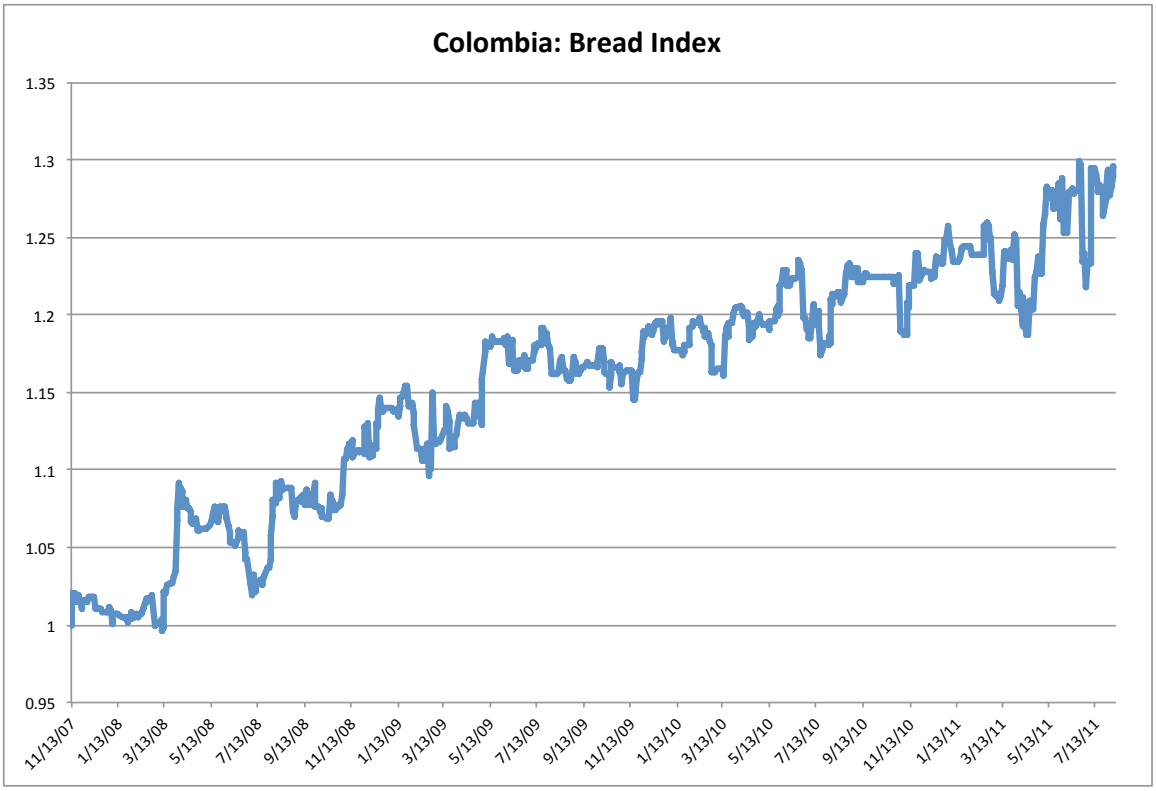


Figure 5:

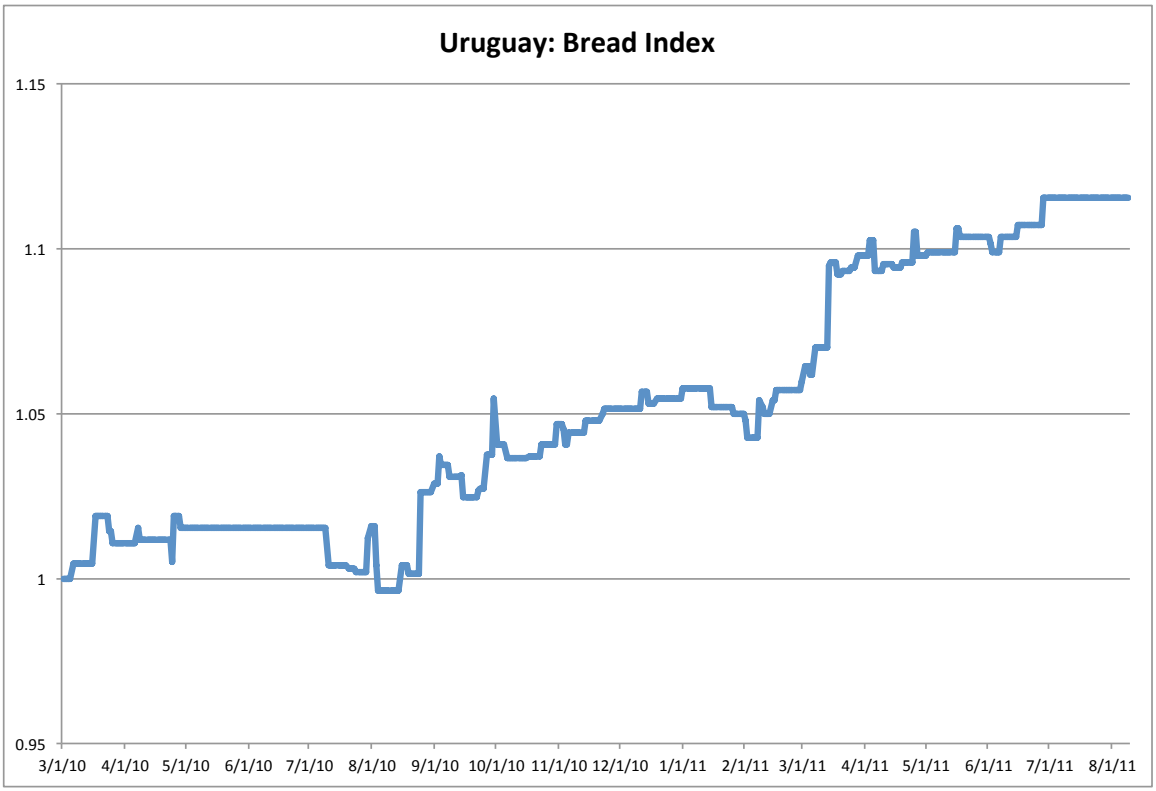
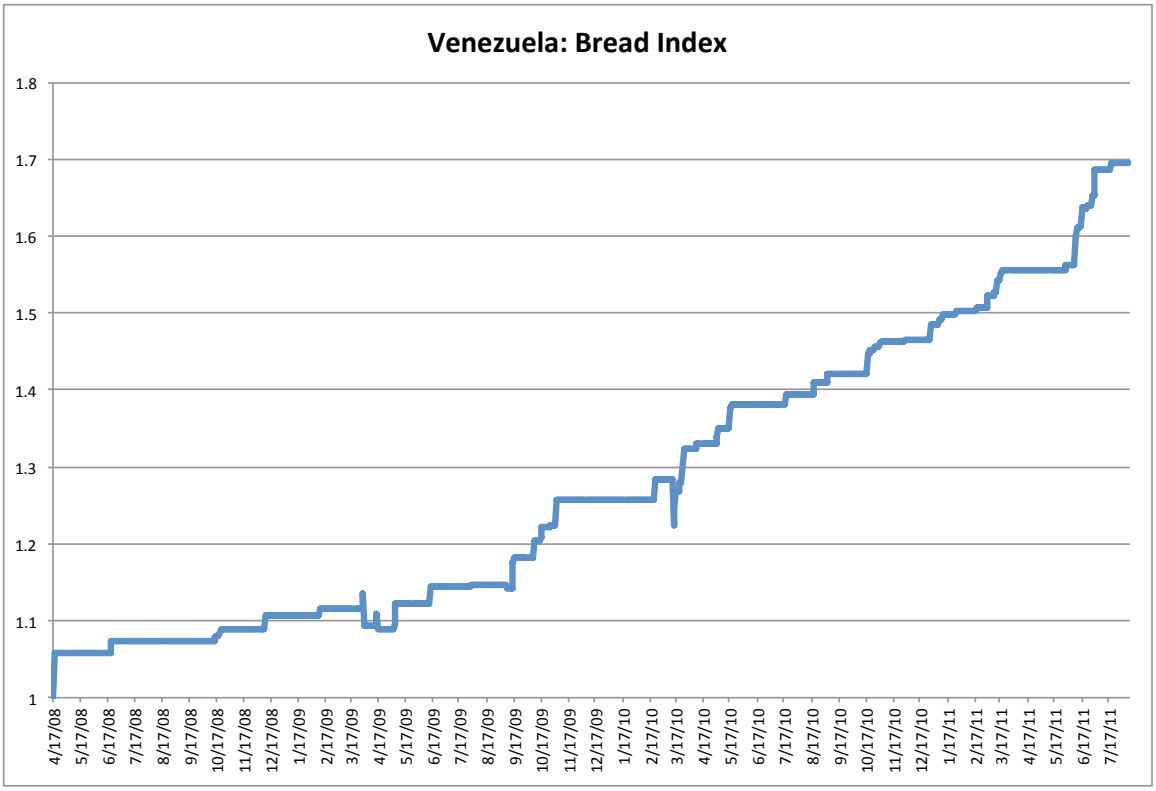


Figure 6:



Preliminary observations

As Figures 1—6 show, bread inflation follows very different trajectories in these 6 countries. In Argentina (Figure 1), the price index rose from 1 in October 2007 to more than 2 in August 2011. This represents an overall inflation of more than 100 percent over the period—way above that of international wheat prices.

Venezuela (Figure 6) is the second country in terms of bread inflation in our sample, with a 70 percent increase between April 2008 and August 2011. Although the overall inflation is close to Argentina's, it is interesting to notice that prices in Venezuela tend to rise in steps or jumps, rather than more gradually as in the Argentinean case. In other words, changes in bread price in Venezuela seem to happen concomitantly across bread types, creating these sharp discontinuities. These differences may reflect more effective price controls in Venezuela than in Argentina—although any greater effectiveness would clearly be short lived. Beyond these differences, the rate of bread inflation in both places is extremely large. The Argentinean annualized rate is 20.31 percent, while in Venezuela it is 17.26 percent.

The other four countries (Brazil, Figure 2; Chile, Figure 3; Colombia, Figure 4, Uruguay, Figure 5) in the sample exhibit a very different behavior. In contrast to the close to monotonous increase in prices observed in Argentina and Venezuela, the price of bread in these 4 countries tended to fluctuate, reflecting somewhat the behavior of wheat prices. The patterns of bread inflation in Argentina and Venezuela—way above that of wheat—suggest that larger inflation pressures are at play, while in the other four countries the factors affecting the price of bread seem to be more directly related to wheat prices. We were not yet able to further test this hypothesis formally because of the length of the data series.

In summary, these differences in price behavior are interesting because bread is mostly wheat plus energy. Still, the stochastic processes in these countries seem very different.

Some reasons could be explored, for which longer data series are needed. Two possible explanations could be: (i) some of these countries are exporters of wheat, while others are importers; (ii) and some of these countries face very high inflationary pressures—mostly from excessive monetary and fiscal expansions—that has led them to resort to price controls to temporarily mitigate the inflationary pressure on staple products, thereby masking the underlying bread inflation. Further research is needed to identify which of these effects may be at play.

What this tool shows is the potential offered by online marketplaces to scrape prices in real time and construct highly reactive inflation measures that complement other existing sources of data.