Online and Official Price Indexes: Measuring Argentina's Inflation

Alberto Cavallo*

Massachusetts Institute of Technology

Abstract

Prices collected from online retailers can be used to construct daily price indexes that complement official statistics. This paper studies their ability to match official inflation estimates in five Latin American countries, with a focus on Argentina, where official statistics have been heavily criticized in recent years. The data were collected between October 2007 and March 2011 from the largest supermarket in each country. In Brazil, Chile, Colombia, and Venezuela, online price indexes approximate both the level and main dynamics of official inflation. By contrast, Argentina's online inflation rate is nearly three times higher than the official estimate.

*Address: Massachusetts Institute of Technology, Sloan School of Management, 77 Massachusetts Ave E62-512, Cambridge MA 02139 - acavallo@mit.edu - 1-617-715-4837 http://acavallo.mit.edu

An early version of this paper, titled "Argentina: Inflation and Lies", was part of my PhD dissertation at Harvard University . I thank my advisors, Philippe Aghion, Robert Barro, and Roberto Rigobon, for their support and encouragement. I am also grateful to Barry Bosworth, Erwin Diewert, Ernie Berndt, Brian Kovak, and seminar participants at Harvard, MIT, NBER, The World Bank, IMF, and LACEA; to the editor and two referees, for their helpful comments and suggestions.

JEL Code: E31.
Keywords: Argentina, inflation, scraped online data, online price index

1. Introduction

The availability of online prices represents a unique opportunity for the construction of price indexes and the measurement of inflation around the world. An unprecedented amount of micro-level price data can now be collected using special software that finds and aggregates detailed product information available in online retailers across the web. This type of data collection can be collected remotely, at much higher frequencies, and a tiny fraction of the cost of traditional price-collection methods.

Among it many potential uses, price indexes constructed with online data can be used to obtain alternative inflation estimates in countries where official estimates have lost their credibility. In particular, this paper uses online prices to evaluate the widespread claim that the Argentine government has been manipulating official inflation indexes since 2007.[1] Online price indexes are first shown to be able to approximate both the level and dynamic behavior of inflation trends in four Latin American countries: Brazil, Chile, Colombia, and Venezuela. In Argentina, by contrast, there is large unexplained difference in the level of online and official inflation rates. A series of robustness tests show that there is no simple data or methodological explanation that can account for this large discrepancy between online and official data.[2]

The data were collected between October 2007 to March 2011 by the Billion Prices Project (BPP) at MIT. Every day a software scanned the websites of the largest supermarkets in each of these countries, collecting product-level data and storing it in a database. Over time, a panel dataset was constructed with detailed information on each product, including prices, product IDs, and a category indicator. Data from six supermarkets are analyzed in this paper, two for Argentina and one for each of the other countries. On average, there are 20,752 individual products per retailer.

A combination of online prices, standard CPI methodologies, and official category weights is used to build an "online price index" in each country. Each online index is then compared to an equivalent official supermarket index, constructed as a weighted average of the official CPI components of food, beverages, and household products (the same categories available in the online data).

In Brazil, Chile, Colombia, and Venezuela, the online indexes are able to approximate both the average level and main dynamics of official inflation. The matching is best in Chile, with an average annual inflation of 3.00% online and 3.19% offline, and a correlation of 0.97 in the annual inflation series. The match is also close in Colombia, both for the level and dynamics of annual inflation, which dropped in both online and official series from about 8% in late 2008 to about 3% in late 2009. In Brazil and

[1] See [1], [3], [8].

[2] More tests and details are provided in an Appendix available online

Venezuela, the online index is able to match the main inflation trend of the official index, but the annual inflation series are less synchronized over time. Not surprisingly, the degree of matching appears to driven by the representativeness of the data. The supermarket in Chile has a market share of 27%, and Santiago, the city where the prices are collected, represents 55% of the national CPI. By contrast, the supermarket in Brazil has just a 15% market share, and online prices are collected in Rio, which represents only 17.3% of the national CPI.

The results for Argentina are remarkably different. There is a large discrepancy between online and official price indexes that is persistent over time. Over three and a half years, online prices had an annual inflation rate that was consistently two to three times higher than in official statistics, with an average rate of 20.14% for the sample period, compared to just 8.38% in official data. During that time, the online index grew more than 100% while the official index increased just 35%. Surprisingly, although the *level* of inflation is higher, the dynamic behavior of online inflation matches the official data quite well, both at the annual and monthly frequencies. In particular, both the online and official series show a decrease in annual inflation rates during 2009, when the economy was going through a recession. Compared to the other countries, the results in Argentina are puzzling because there are reasons to expect the online data to be even more representative in this case. The Argentine supermarket has a large market share and targets consumers with a broad range of income levels. Furthermore, the online prices come from Buenos Aires which is also the only location where official data collection takes place.

Alternative methodologies and subsets of the data yield similar results. First, an online index is built using data from a second supermarket in Argentina with completely different characteristics: a small, online-only retailer that targets high-income people in Buenos Aires. Second, an index is constructed exclusively with goods that had price controls. Finally, a simple "Subsistence Food" Index, with only 45 goods that are carefully matched to the official data, is compared to the official "Canasta Básica de Alimentos" (CBA) index used to calculate the level of extreme poverty in the country. In all cases there are still large unexplained differences with the equivalent official series.

The best way to approximate the official price index in Argentina is to use one third of the inflation rate observed online. This supports the widespread suspicion that the government has been manipulating the CPI since January 2007, when it intervened the National Statistics Institute (INDEC). The implications for other statistics are significant. For example, using an online adjusted cost for the subsistence-level CBA basket, the share of the population in extreme poverty during the first quarter of 2011 rises from 2.5% in official estimates to 6.69%. Similarly, poverty estimates are 9.9% in official data, but rise to 25.9% with adjusted price series. The implications for real GDP are equally impressive. If the GDP deflator had behaved like the online index since 2007, the real GDP annual growth rate would have been

just 0.5% by March 2011, much lower than the 10% officially reported.

2. The Data

The data were collected by the Billion Prices Project at MIT using a technique called "web scraping" to record the price for all goods sold online, between October 2007 and March 2011, in the largest supermarket in Argentina, Brazil, Chile, Colombia, and Venezuela.

The technology to scrape prices is conceptually simple. Most webpages are built using a structured coding language called HyperText Markup Language (HTML). This code has simple "tags", such as `<center>` and `<bold>`, that determine the style and placement of text in a page. These tags tend to remain constant over time, as they provide a distinctive "look and feel" to each page. By contrast, the information *within* these tags, such as a product's price, changes all the time. The scraping software can be taught to use the HTML tags to locate relevant information about a product and store it in a database. Repeating the process every day produces a panel database with a one record per product per day. In addition, the web address or "URL" of the page where each product is located can used to classify products into standardized categories.

Table 1 describes the six databases included in this paper: two for Argentina, and one for each of the other countries. Argentina's Retailer #1 is the largest supermarket chain in the country. It is used for the main online price index, in section 3, because it has the largest market share and detailed category indicators. Retailer #2, used for robustness in section 5.3.2, is a smaller supermarket that sells exclusively through the web. The supermarkets in Brazil, Chile, and Colombia are market leaders, with websites that target the cities of Rio de Janeiro, Santiago, and Bogot. The supermarket in Venezuela is a smaller retailer that sells in Caracas.

In all cases, the online data contains a combination food, beverages, and household products. Categories range from "Eggs" to "Appliances", with about a third of them corresponding to household products (including cleaning materials, health and beauty products, furniture, appliances, and books). These categories account for between 28.44% (Colombia) and 48.51% (Argentina) of CPI weights in these countries, as shown in Table 1.[3]

3. Online Price Indexes

The online price indexes use a combination of online data and official category weights. The methodology follows the way CPI statistics are constructed in these countries as closely as possible, but there

[3]Detailed official category weights, for the products available in each supermarket, are shown in Appendix Table **??**.

are some difference in the treatment of the data.

First, daily data are used to construct the online price indexes. Such high-frequency is useful to observe short-term patterns in the data that help validate the online information, as discussed in Section 5.3.3, although similar results can be obtained with monthly data.

Second, the online indexes are built using prices for all products available for purchase at each retailer. This implies that the basket of goods changes dynamically over time as products appear or disappear from the online stores, and that the number of prices for product varieties tends to be much larger than in official statistics. Section 5.3.3 finds similar results when a fixed basket or alternative sub-samples of goods are used.

Third, there are no forced product substitutions or adjustment for quality changes. All goods are treated independently, so products that are discontinued on a given date stop affecting the index from that day forward. Similarly, new goods only impact the index on their second day in the sample, when their first price change can be observed. Substitutions and quality adjustments are not common in official statistics for the categories of goods analyzed in this paper.

Finally, short gaps in an individual price series -lasting only a few days- are common in this high-frequency data, caused either by failures in the scraping method or because some products go temporarily out of stock. These gaps are filled by carrying forward the last available price for each product. All results are robust to the use of cell-relative adjustments, the method used by the Bureau of Labor Statistics to fill price gaps in the US CPI.[4]

3.1. Index Computation

To build the index, price changes are calculated at the product level, then averaged inside categories using unweighted geometric means, and finally aggregated across categories with a weighted arithmetic mean. In particular, the first step is to obtain the unweighted geometric average of price changes in category j for each day t:

$$R^j_{t,t-1} = \prod_i \left(\frac{p^i_t}{p^i_{t-1}} \right)^{\frac{1}{n_{j,t}}} \tag{1}$$

where p^i_t is the price of good i at time t, $n_{j,t}$ is the number of products in category j that are present in the sample that day.

The second step is to compute the category-level index at t:

[4]See **??** in the Appendix for details.

$$I_t^j = R_{1,0}^j \cdot R_{2,1}^j \cdots R_{t,t-1}^j \tag{2}$$

Finally, the Supermarket Index at time t is the weighted arithmetic average of all category indexes:

$$S_t = \sum_j \frac{w^j}{W} I_t^j \tag{3}$$

where w^j is the official CPI weight for that category and W is the sum of all the weights included in the sample.

The classification of products and weighing of categories is one of the most complex parts of this process. In the original data, each product is linked to a web address (URL) that corresponds to the webpage where the product is located. These URLs group similar items together, into various levels of aggregation chosen by each supermarket. The number of URLs ranges from about 300 to 1000 in these retailers. The advantage of having URLs is that thousands of items can be easily classified into a set of standardized official categories, so that their corresponding official weights w^j could be used to obtain the aggregate index.[5]

Daily estimates for the "monthly" and "annual" inflation rates are also obtained. At any point in time, the monthly inflation rate computes the percentage change in the average index of the last 30 days with respect to the average of the previous 30 days. For example, on November 30th 2010, it is the percentage change between the average of the daily index from November 1st to November 30th 2010, and the average of the daily index from October 2nd to October 31st 2010. Similarly, the annual inflation rate is the percentage change in the average index of the last 30 days with respect to the average of the same period a year ago.[6]

Compared to CPI statistics, these online indexes have an advantage in terms of frequency and the number of items sampled within each category. For example, in Argentina alone, there are prices for 781 different product varieties in the "Milk" category alone (this includes different brands and package sizes). An main disadvantage is that these online prices come from a single retailer in only one city. As discussed below, this can limit the ability of online indexes to match the short-term inflation dynamics. However, this problem is minimized in this case because these Latin American supermarkets have huge market shares, shown in Table1, and are located in cities that concentrate a large percentage of each

[5]Table **??** of the Appendix has a detailed list of categories and weights used. A complete list of official weights in each country is available in [2], [9], [10], and [12].

[6]The annual inflation rate is not equal to the sum of the monthly inflation rates because it is not computed as the difference in logged values of the index across periods. Instead, it is the difference in index levels between periods. This is consistent to the way the Argentine Statistical Office (INDEC) computes both monthly and annual inflation rates every month. See the Appendix for details.

country's population and CPI weights.[7]

4. Four Latin American Countries

This section studies the performance of online price indexes in Brazil, Chile, Colombia, and Venezuela. These countries have similar online markets and CPI methodologies as Argentina, which is studied in detail in the next section.[8] All indexes are weighted using official CPI weights, with the exception of Venezuela, where the online data could not be classified into the standard official categories. In that case, an un-weighted index is constructed with a simple geometric average of all price changes observed each day.[9]

4.1. Online and Official Data

The online and official supermarket indexes are shown in Figures 1 and 2. Table 2 shows the summary statistics in each country. The official supermarket index is a weighted average of the "Food and Beverages" and "Household Products" official indexes. These are the sub-component of the CPI that are directly comparable to the online price indexes built in this paper. Using the CPI as the basis for comparison would not change the results, because the supermarket index is closely related to the CPI in these countries, as shown in Table 2.[10]

The graphs in Figure 1 show a remarkable ability of online indexes to track the main inflation trends over long periods of time. Although there are periods when online inflation is rising faster (or slower) than official estimates, over time both indexes follow a common trend. Indeed, Table 2 shows that the average annual inflation rates for the period are nearly identical for all countries. This happens both in a low-inflation country like Chile, where the annual rate is 3% online and 3.19% in official series, and also in a high-inflation country like Venezuela, where the annual rate is 27.43% online and 29.38% in official data.

In the annual inflation series shown in Figure 2, Chile and Colombia stand out for the close match in both the level and dynamics of annual rates. In Chile, the online and official estimates are nearly

[7] Santiago has 55%, Buenos Aires 32%, Caracas and Bogota 22%, and Rio de Janeiro 8% of the total population in each country).

[8] The 2009 World Development Indicators (World Bank) estimate an internet user penetration of 30.4% in Argentina, 39.2% in Brazil, 33.9% in Chile, 45.5% in Colombia, and 31.2% in Venezuela.

[9] The unweighted index implicitly weighs products by the number of items within each sub-category. For example, milk comes in many forms and product sizes, so when all individual products are included in the index, "milk" will receive a relatively large weight in the supermarket index. Unweighted indexes can provide a good approximation if there is a link between the variety of products and the share of expenditure that a category represents. Consistent with this idea, [14] find a strong correlation between the number of SKUs and purchase volume in a large US retailer. The results of this paper hold if unweighted indexes are used in all countries, as shown in the Appendix.

[10] See the Appendix for more comparisons between the all-items CPI and the official supermarket index

identical over time, with online inflation peaking in April 2009 at 8.7%, falling to -1.3% in February 2010, and climbing back to 3.5% in March 2011. In Colombia, online and official annual inflation series also have very similar patterns over time, with inflation dropping from about 8% in late 2008 to about 3% in late 2009, and remaining relatively stable ever since. In Brazil and Venezuela, the online data can also approximate the average level of annual inflation, as seen in Table 2, but the dynamics of annual inflation are less synchronized. This is evident in both the graphs and the low correlations of annual series shown in Table 2. For Brazil, online and official inflation alternate periods of higher inflation (the correlation improves after June 2010). In Venezuela, the online index puts annual inflation in the 25% - 35% range during this period, consistent with official numbers, but it peaked in December 2009, six months before the official index. This may be reflecting a delay of price adjustments in "public" supermarkets owned by the government, which were 25% of the CPI sample in 2008.

4.2. Explaining the Differences

The differences in the online-official matching across countries appears to be linked to the representativeness of the retailer used in online data, shown in Table 1. In Chile, for example, the supermarket sampled has a relatively large market share of 27%, compared to only 15% for the Brazilian retailer. At the same time, the city of Santiago (where the products priced online are delivered) represents 55% of the national CPI (it was 100% until 2009), while Rio de Janeiro is only 17.3% of the CPI sample in Brazil.

In all countries, as we move from annual to monthly rates, the matching between online and official series becomes weaker. The correlation between series shown in Table 2 tends to be lowest for the monthly rates. This is caused by the use of data from a single retailer in each country, which in the short run can adjust its prices slower or faster than the economy as a whole. Indeed, many of the temporary deviations from the official index in Figure 1 appear to be driven by idiosyncratic characteristics of these retailers. For example, the retailer in Brazil tends to increase a large number of prices few months rather than continually over time, while the supermarket in Colombia has store-wide sales every quarter, causing the online index to drop temporarily for about a week each time. Still, these deviations with official data tend to be corrected after a few months, improving the matching at the annual frequency.

In short, these results show that online price indexes –even when constructed with data from a single retailer– are able to match both the average level of annual inflation and the long-run inflation trends in these countries. Their ability to match high-frequency inflation dynamics appears to depend greatly on the representativeness of the retailer (market share) and the importance of the city where the online products are collected.

5. Argentina

This section focuses on Argentina, where the official estimates of inflation have become widely discredited in recent years. It first documents large differences between official inflation estimates and those obtained independently with online data. It then provides several alternative indexes to evaluate the robustness of these findings, and concludes that the best approximation to the official numbers is simply to use one-third of the actual inflation rate observed online.

5.1. Government Intervention in the Statistical Office

Since 2003, Argentina's inflation grew steadily as a result of an expansionary monetary policy designed to stimulate consumption and avoid an appreciation of the currency . Inflation became a politically-sensitive issue in 2006, when the annual inflation rate increased over 12%. A combination of subsidies and price controls failed to contain prices, so in January 2007 the government took a drastic decision: to take direct control of the National Statistics and Census Institute (INDEC) and fire the people responsible for computing and publishing the CPI. Since then, official statistics have become widely discredited in the media and academic circles.[11]

During 2007, official estimates of annual inflation remained below 9.7%, while surveys of inflation expectations reached 30% by the end of the year. Gradually, economists and private institutions started to monitor the prices for small baskets of products and reporting inflation rates that were significantly higher than official estimates.[12] Provincial governments also computed regional inflation estimates inconsistent with INDEC's estimates. The government has repeatedly claimed that these metrics are flawed because the samples are small and not representative. Over time, it has increased its pressure on economists and institutions publishing alternative inflation estimates. In February 2011, several economists received official letters imposing large fines and threatening with jail-time if they continued to publish their own inflation results.

In the current context, online data provides a unique opportunity to measure alternative inflation rates in Argentina. The government cannot interfere with the data collection, which is done remotely, the sample size is several order of magnitude larger than in other private estimates, and more importantly, the previous section shows that the same online data and methodologies can provide close approximations to official inflation rates in other countries.

[11] See [3], [1], [8], and [13].

[12] In March 2008 I created a website that published daily inflation statistics (www.inflacionverdadera.com). The CBA Index shown in Figure 5 was part of that effort.

5.2. Large Differences with Official Data

The online supermarket index for Argentina, built using online data and official category weights, is compared to an equivalent official index in Figure 3.

In Argentina's case, the online price index follows a completely different trend. The accumulated difference with official data has continued to grown steadily over time. Between October 2007 and March 2011, the online index increased over 100%, while the official index grew only 35%.

The annual inflation series in Figure 2(b) presents two main results. First, online inflation has been consistently between 2 to 3 times higher than official inflation. Second, the online and official estimates share a surprisingly similar pattern over time: inflation fell in early 2009, increased in early 2010 and fell slightly in early 2011. In fact, as seen in Table 2, the correlation in annual inflation rates is high in Argentina, just like in Chile and Colombia.

The monthly inflation rates tell a similar story. The online rate has been consistently above the official rate for the whole period. A simple OLS Regression between monthly online and official rates, shown in Table 2, suggests that the official index is missing an average of 84 basis points every month compared to the online inflation. In fact, Argentina is the only country where this regression yield a statistically significant constant. Table 2 also shows how the official monthly inflation rate is far more stable in Argentina than in any of the other country (relative to the observed online volatility).

These results are puzzling because they are exactly the opposite to those in the other countries. Online data cannot match the official *level* of inflation in Argentina, which on average is 20.14% vs the official 8.38%, but it does an excellent job in matching even the short-run dynamic behavior of inflation rates. Indeed, the contemporaneous correlation in the monthly inflation series is higher in Argentina than in any other country. If the data from this single retailer were not representative, then we would expect the short-term dynamics to be affected first, as it appears to happen in Brazil.

5.3. Robustness

This section considers several robustness tests to find alternative explanations for the differences in Argentina's online and official inflation estimates. I consider three alternatives: using data from a different online supermarket, using only the prices of goods that were under price controls, and using a fixed-basket methodology to construct a basic food index.

5.3.1. Alternative Supermarket

The fact that online data can match official inflation rates so well in other countries, but so badly in Argentina, suggests that the differences are not caused by general characteristics of online data. Still,

there may be concerns that the online data used in Argentina are not as good or representative as the data used in other countries. One possibility is that online prices behave differently than offline prices in this particular retailer. However, in another paper, [5], I directly compared the online and offline prices of a small sample of goods in this supermarket, and found that online and offline prices had similar levels of inflation in January 2009. Even if that changed later on, it seems highly unlikely that online and offline prices can behave so differently for such a long period of time. For example, assuming that online and offline prices were the same in October 2010, by March 2011 online consumers would have been paying over 60% more than offline buyers in the same supermarket.

Another possibility is that this particular retailer is not representative of the country as a whole. To test this, I constructed an online price index using data from Retailer #2. As described in Section 2, this is a supermarket with widely different characteristics: a small, online-only supermarket that targets high-income people. There is no category information for products in this retailer, so the index includes only a simple geometric average of daily price changes among all products in the retailer each day. This is the same methodology used for Venezuela, where it does a good job at matching official inflation.

The online indexes for both retailers are shown in Figure 4. No matter what retailer is used, the online inflation rate is significantly higher than the one reported in official data.

5.3.2. Alternative Data: Price Controls

Another explanation for the differences with official data may be that INDEC uses prices for goods with price controls. These goods can be identified online because retailers place images next to the products that read things like "Government Agreed Price". The scraping software can automatically record whether a good was under a price control or not each day with a binary indicator.

During this period, the government periodically imposed price controls on a set of goods after reaching "agreements" with the major supermarket chains in the country. Although the details were never made public, the scraped data reveals that 597 products where under a price control at some point in time in Retailer #1, with restrictions lasting a few weeks each time.

Figure 4 shows a price index that includes only price-controlled goods. The inflation rate is far more volatile than when we use all goods in the supermarkets. This volatility rises naturally from the fact that each price control lasted only a few weeks, and retailers tended to increase prices much faster when the controls ended, probably to compensate for any delayed price-adjustments. Therefore, the price control policy was clearly ineffective at containing the inflation rate of these goods over the long run. Indeed, the average inflation rate is similar to the other online price indexes, and significantly higher than the official estimates.

5.3.3. Alternative Method: The Basic Food (CBA) Index

Beyond the data, it can be argued that the methodology of the supermarket index, which works so well in other countries, is somehow incompatible with official CPI statistics in Argentina. So, instead of constructing a supermarket index, I now focus my comparisons on a much simpler indicator: the "Canasta Básica de Alimentos" (CBA). This is a Basic Food Index used to measure the level of extreme poverty in the country. Its methodology is well documented, with details of the exact weighing and characteristics of the products that underlie its construction, so I can replicate the exact same basket of goods. This basket has a much smaller number of products within each category, instead of every item in the store.[13]

Figure 4(a) compares the online and official CBA indexes. Once again, the differences are huge and persistent over time: by March 2011, the online CBA index had an accumulated inflation of 91%, while the official INDEC CBA had increased by only 31%. Interestingly, there is a spike in online prices during March 2008 which perfectly coincides with the timing of the massive "Farm Strike" that took place in the country. This was a strike of farmers in response to the attempt to introduce a variable-scale export tax regime on several key crops, which in practice meant an increasing the tax rate to 45% for soybeans (Argentina's main agricultural export). Farmers organized road blocks from March to June and caused severe shortages of basic food products for several weeks. The most dramatic price increases occurred during the first strike, which started on March 13th, 2008 and lasted until April 3rd, 2008. The effects were mostly temporary, with the online index going back to its original trend by June 2008. By contrast, the official CBA index was surprisingly unaffected by these events.[14]

5.4. Best Approximation: 1/3 rd of the observed inflation rate

The results so far strongly suggest that the INDEC is manipulating the official inflation estimates. How this is done is an open question, but the main results in Figure 3 suggest a simple answer: the government is reporting just a fraction of the real inflation rate, usually between a third or one-half of the actual numbers. In fact, a simple OLS regression of the official and the online annual inflation series yields a coefficient of 0.36. To test this hypothesis, we can run a simple simulation: an index built using only a third of the inflation rate observed each day.

This simple approach yields an index with a remarkable resemblance to the official data, as seen in Figure 6. This suggests that the way the data is being altered is far simpler than commonly assumed. After all, the INDEC is a large organization, with many employees involved with the data collection and

[13] Details for the basket and weights is provided in the Appendix.

[14] The online supermarket index also has a sharp increase in march 2008, but the effects appear to be permanent in this index. It is possible that supermarkets used this shock as an opportunity to justify increases in the aggregate index that had been delayed for some time. In particular, the year before the strikes the government was very aggressive with the implementation of price controls and boycotts against supermarkets that raised prices.

construction of the price indexes. Instead of changing the prices at the item level, it is probably easier for the government to change the aggregate numbers, which are seen by just a handful of people at the end of the CPI calculation process. [15]

5.5. Implications for Other Statistics

The bias in inflation estimates also affects other statistics, such as the poverty and real GDP estimates. For example, the differences with the CBA index shown in Section 5.3.3 has a direct implication for poverty estimates. Every quarter, INDEC uses the cost of the CBA basket to see how many individuals are in extreme poverty conditions. Taking the cost of the official basket in the first quarter of 2008, and adjusting it with the CBA inflation rate observed online, the basket in July 2011 would have a cost of $259.5 Argentine pesos. Using INDEC's income survey, this implies that 6.69% of the population was under extreme poverty at the time, compared to only 2.5% reported in official statistics. Similarly, after adjusting the CBA to obtain the broader "Canasta Basica Total" (CBT), which adds non-food items to the basket, the level of poverty becomes 25.9% compared to the 9.9% officially reported.[16]

To estimate the impact on real GDP, we start by looking at how the CPI and the GDP deflator have behaved in the past decade. Figure 7 plots both series from 1994 to 2011. The data from 1994 and 2006 shows that both series were closely correlated, which is the expected behavior under normal conditions.[17] However, since 2007, the GDP deflator has increased significantly faster than the CPI. This means that government has recognized higher inflation in the GDP deflator. This is not surprising. If the deflator had increased as little as the CPI, the growth rate of real GDP would have been over 10% for several years, which would be impossible for the government to justify. On the other hand, allowing for a higher GDP Deflator has little political cost, as it is rarely followed by the media or the general public as a metric for inflation.

Still, the GDP Deflator has increased considerably less than the online index since 2007. To get a lower bound for the real growth rate, an "Adjusted Real GDP" can be calculated using the online index to deflate the official nominal GDP. The results, shown in Figure 7, provide GDP growth rates of -4% in June 2009 (right before the government lost the parliamentary elections), 5% in June 2010 , and only 0.5%

[15] The Appendix contains several other alternative approaches that try to replicate the official numbers. Some are methodological, such as using the cell-relative imputation of missing prices or the use of unweighted indexes, and others use special subsets of the data, such as including only the lowest inflation goods in each sub-category. None of them is able to match the low inflation estimates in the official data.

[16] The CBA cost is multiplied by 2.16 to obtain the CBT cost. This coefficient, reported by the INDEC, is the inverse of the share of food over all basic household expenditures.

[17] There are reasons to expect small and temporary differences between the GDP deflator and the CPI. The CPI includes only consumer prices, while the GDP deflator includes the prices of goods purchased by the government and firms. In addition, the CPI includes prices of both imports and exports, while the GDP deflator focuses exclusively on domestically produced goods. These reasons, however, do not seem able to justify the persistent differences observed since 2006.

in March 2011. Although these numbers should be taken as an approximation, because an alternative GDP cannot be estimated directly with online data, they do provide a rough estimate for the true growth performance of the country.

6. Conclusions

Online price indexes, constructed using a combination of online data and official methods and weights, are capable of matching both the level and main dynamics of official inflation in Brazil, Chile, Colombia, and Venezuela. The matching is best at annual frequencies and improves when the data comes from supermarkets with large market shares and cities that are more representative of the country as a whole.

The results for Argentina, by contrast, confirm the suspicion that the government is manipulating the official inflation series. It is the only country where online inflation deviates significantly from official estimates over time. Two things are surprising: the magnitude of the difference, and its persistence. Indeed, online inflation has been consistently between two and three times larger than in official estimates for over three years. On average, the online index had an annual inflation rate of 20.14%, compared to just 8.38% in official data. The accumulated difference was 65% by March 2011. Surprisingly, the difference lies exclusively in the *level* of inflation reported every period, not the dynamic behavior of inflation rates over time.

Several robustness exercises were considered, but none seems able to account for these large discrepancies with official data. The best approximation to the official series is to simply assume that the inflation rate is one-third of the one observed in the online data.

There is no clear reason for why the government continues to manipulate the official price indexes. Some economists point to lower interest payments for inflation-linked bonds, while others highlight the fact that, by using artificially low inflation estimates in the budget, the government can avoid distributing any excess tax income to the provinces. However, these short-term resources are negligible next to the negative effects and uncertainty the manipulation has introduced in the economy. More likely, in 2007 the government was simply trying to hide what it though was a temporary rise in inflation, and as time went by it became increasingly harder to recognize it was lying. In any case, the government is not backing down, but quite the contrary. A collaboration with the IMF announced in 2010 to construct a new price index continues to be delayed and appears to have been just a way to avoid sanctions. Independent economists have continued to be threatened with fines or jail time if they publish their own estimates. Most of them have been forced to comply. Provincial governments, which depend greatly on tax resources sent by the federal government, are also under increasing pressure. Seven provinces have

recently announced that they are no longer going to publish their own estimates. Meanwhile, in both 2011 and 2012 the government increased the minimum-wage by approximately 25%, consistent with the inflation estimates in this paper. Sooner or later, the official inflation series will have to become accurate. In the meantime, online price indexes can provide a good approximation to the real inflation rate in the country.

[1] Barrionuevo, A., 2011. Inflation, an old scourge, plagues argentina again. New York Times , February 5th 2011.
[2] BCC, 2010. Indice de Precios al Consumidor. Banco Central de Chile.
[3] Berumen, E., Beker, V.A., 2011. Recent developments in price and related statistics in argentina. Statistical Journal of the IAOS 27, 7–11.
[4] BLS, 2009. Handbook of methods, in: BLS Handbook of Methods.
[5] Cavallo, A., 2010. Scraped data and prices in macroeconomics. Harvard University - Phd Dissertation .
[6] Cavallo, A., 2012. Scraped data and sticky prices. MIT Sloan Working Paper .
[7] DANE, 2009. Metodologia del Indice de Precios al Consumidor. Coleccion Documentos, DANE - Colombia.
[8] Economist, T., 2011. Lies and argentine statistics. The Economist .
[9] IBGE, 2007. Sistema Nacional de ndices de Preos ao Consumidor. volume 14 of *5*. Instituto Brasileiro de Geografia e Estatistica.
[10] INDEC, 2001. Indice de Precios al Consumidor Gran Buenos Aires. Documento Metodologico Nro 13, Instituto Nacional de Estadisticas y Censo.
[11] INE, 2008. Indice Nacional de Precios al Consumidor. Venezuela.
[12] INE, 2009. Manual Metodologico del Indice de Precios al Consumidor. IPC Nacional - Base 2009, Instituto Nacional de Estadistica de Chile.
[13] Jueguen, F., 2009. Entrevista con graciela bevacqua: "Cristina avalo la manipulacion de las estadisticas del INDEC". La Nacion Newspaper , May 7th 2009.
[14] Simester, D., Anderson, E., 2011. Menu cost and price rigidities: Micro evidence. Working Paper .

Figures

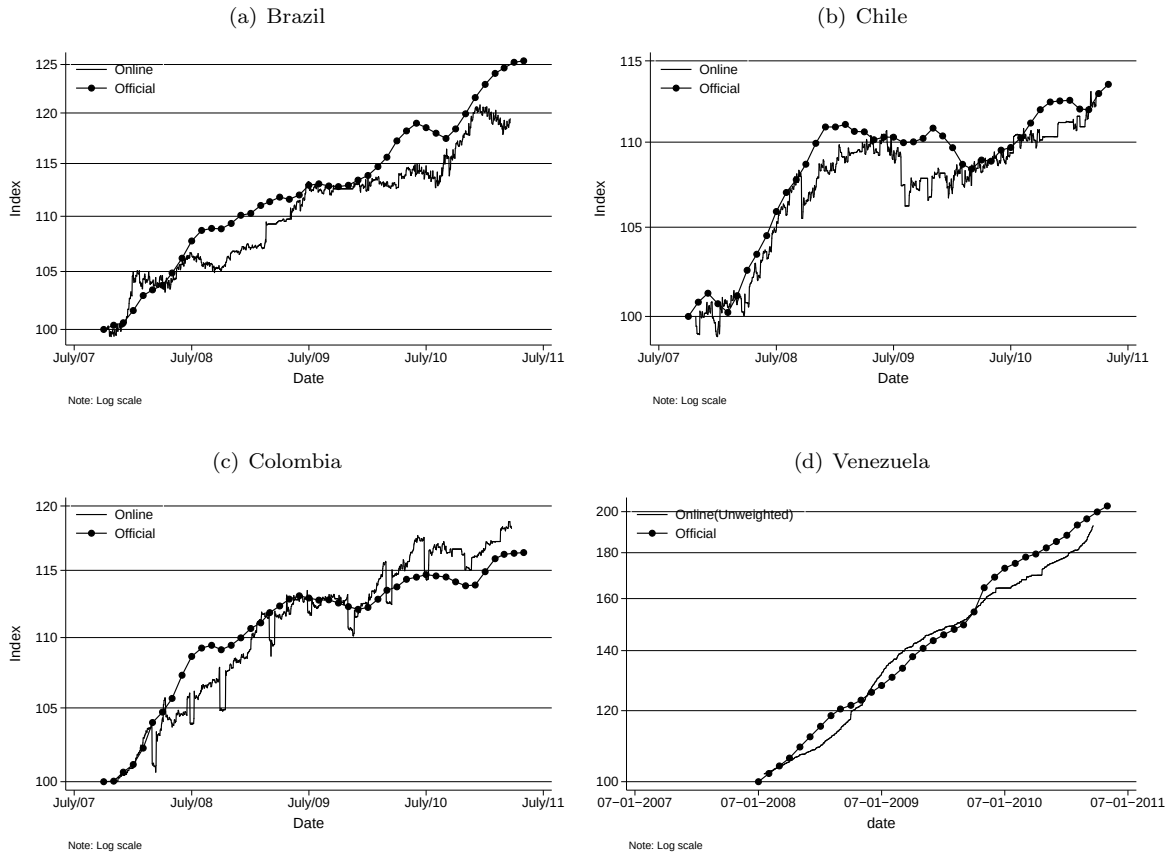


Figure 1: Online and Official Indexes in Four Latin American Countries

Notes: The daily online supermarket index is constructed with a online prices and official CPI category weights. In Venezuela, the online data has no category information and therefore the online index is built as a geometric average of all price changes observed each day. The official supermarket index is an equivalent indicator constructed as a weighted average of the "Food and Beverages" and "Household Products" official price indexes in each country.

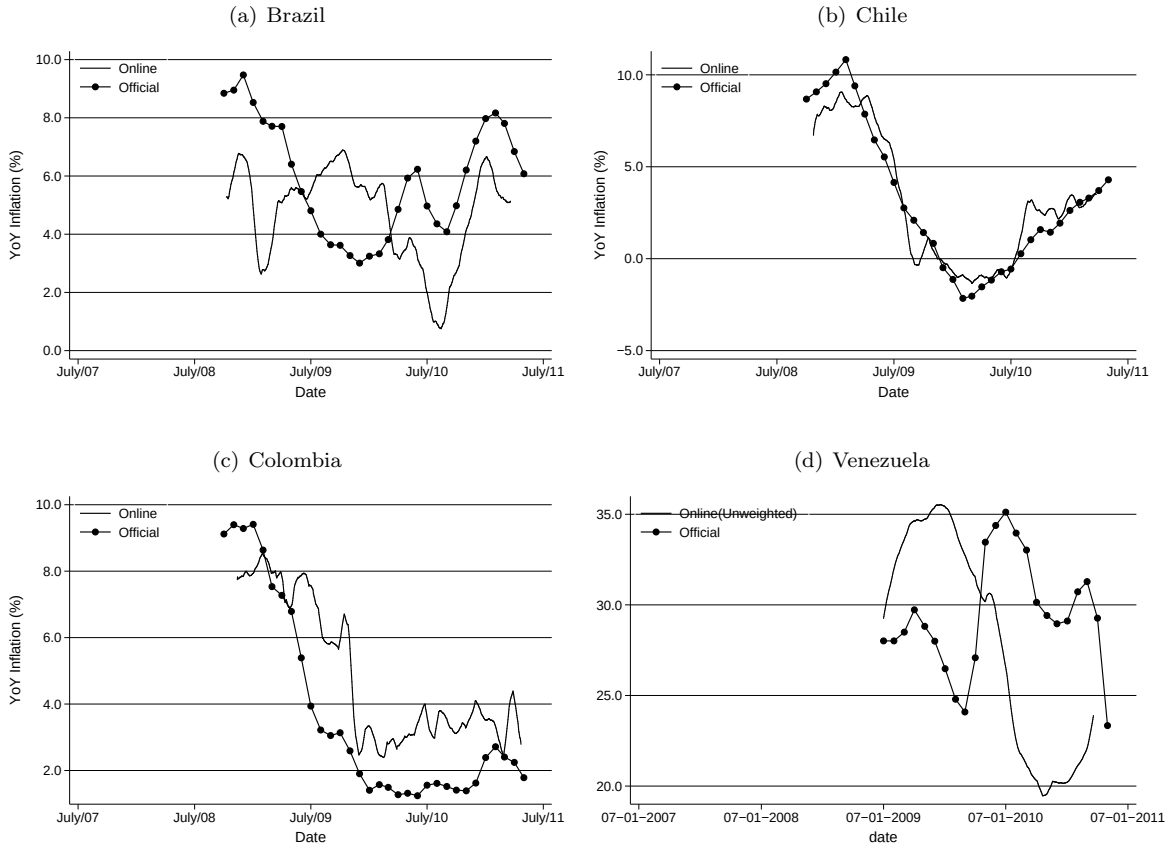
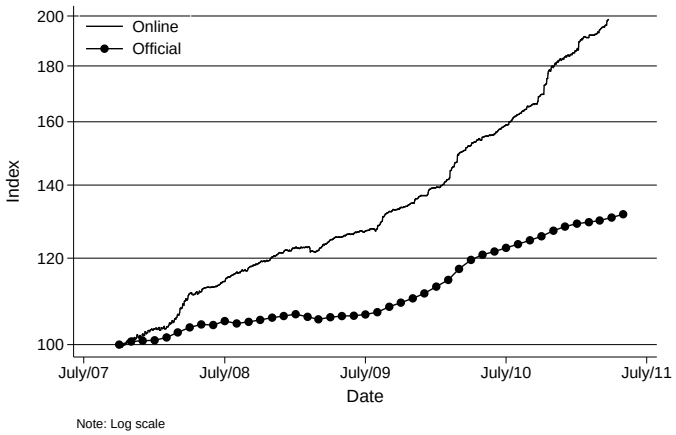


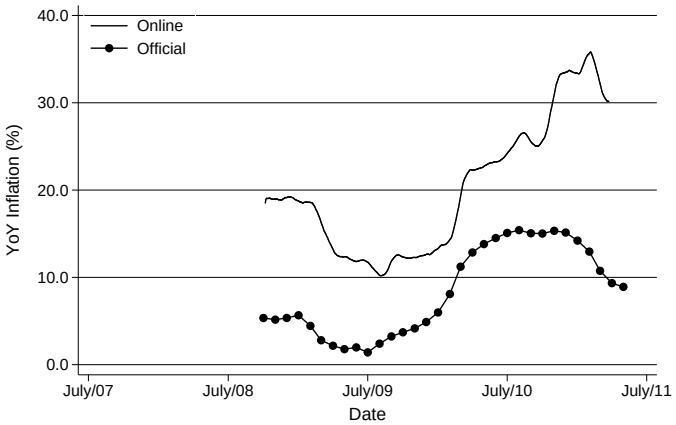
Figure 2: Online and Official Indexes - Annual Inflation Rate

Notes: The annual online inflation rate is a daily time series computed as the percentage change in the average of the index during the previous 30 days with respect to the average of the index in the same period a year before. The annual official inflation rate is a monthly time series computed as the percentage change in the index in the previous 12 months.

(a) Daily Index



(b) Annual Inflation



(c) Monthly Inflation

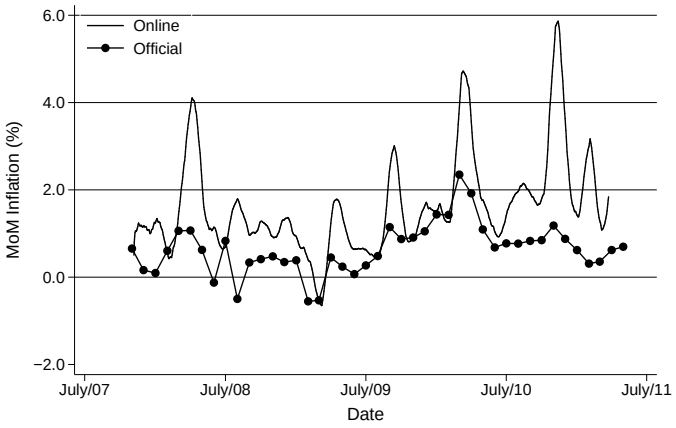
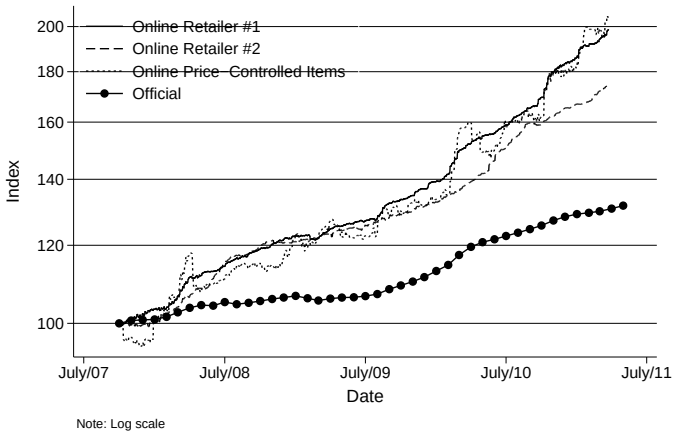


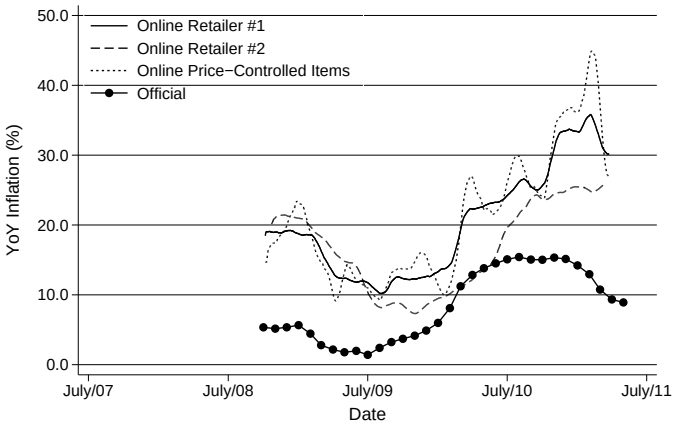
Figure 3: Online Supermarket Index in Argentina

Notes: The monthly online inflation rate is a daily time series computed as the percentage change in the average of the index in the last 30 days with respect to the average of the index in the same period a month before. The monthly official inflation rate is a monthly time series computed as the percentage change in the index over the previous month.

(a) Daily Index



(b) Annual Inflation



(c) Monthly Inflation

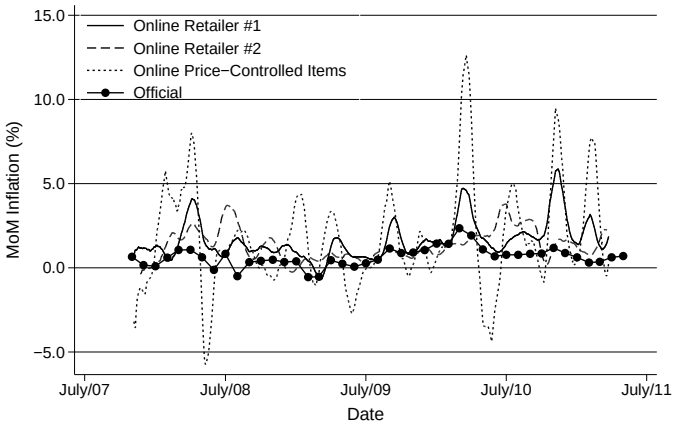
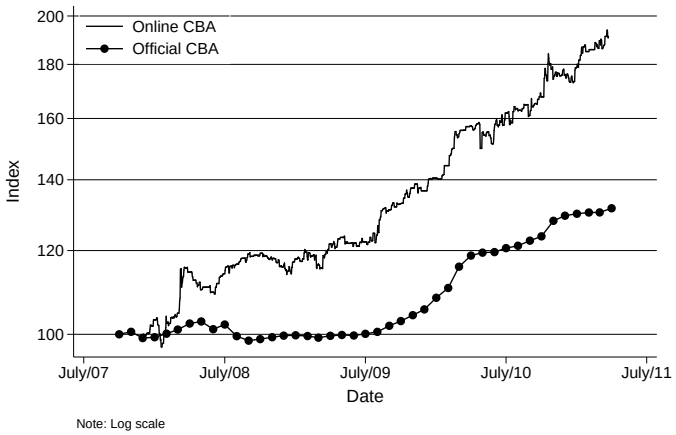


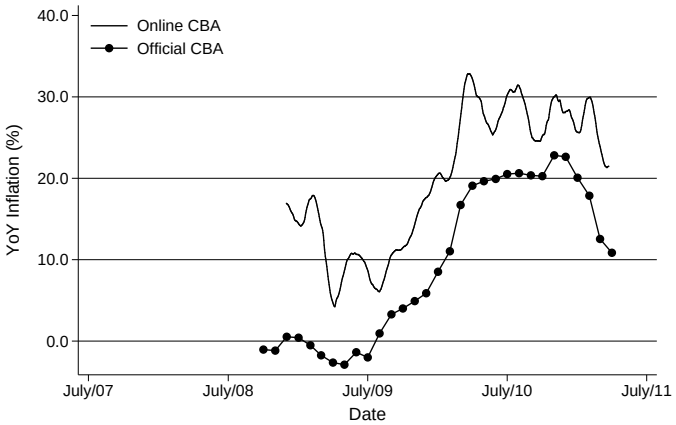
Figure 4: Different Supermarkets and Price Controls

Notes: Argentina's Retailer #1 is the largest supermarket chain in the country. Retailer #2 is a smaller supermarket that sells exclusively online to a higher-income population. The index for Retailer #2 is an un-weighted index because the data lacks information on product categories. The Price-Controlled Index uses data from 597 products that were under price controls imposed by the government at some point in time during the sample period.

(a) Daily Index



(b) Annual Inflation



(c) Monthly Inflation

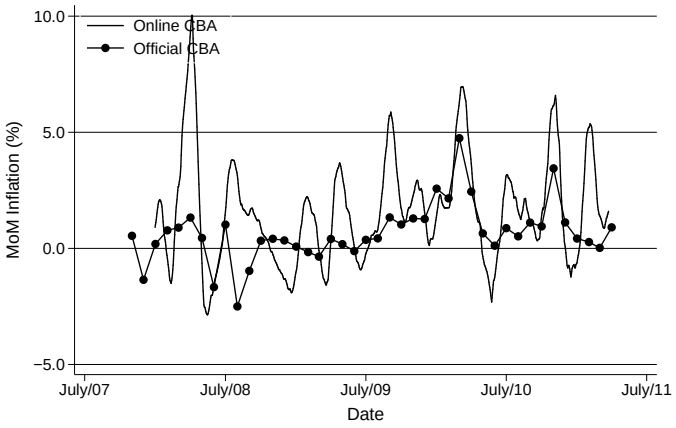
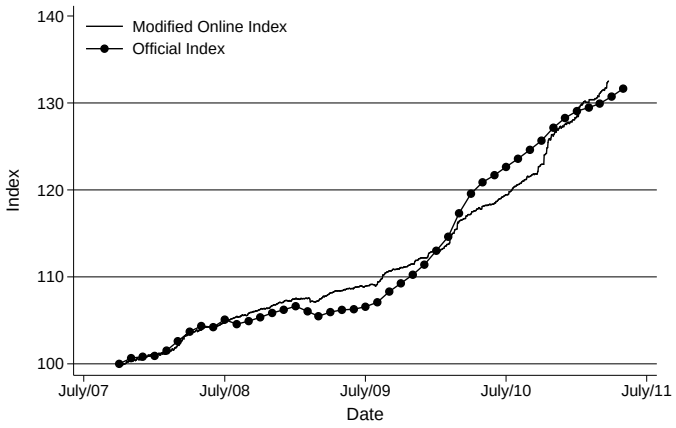


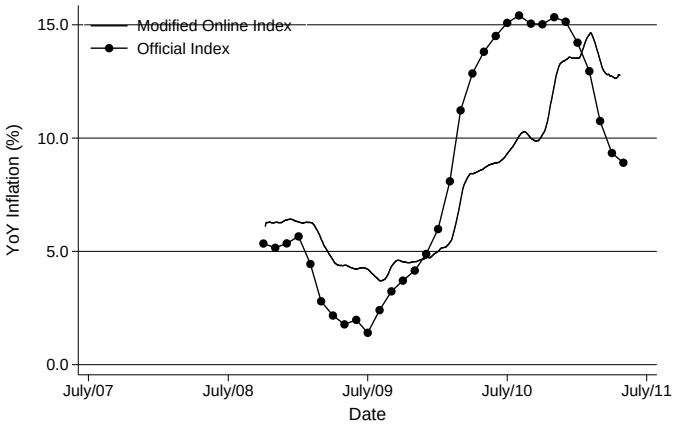
Figure 5: Subsistence Food (CBA) Index

Notes: The "Canasta Basica de Alimentos" (CBA) index is a subsistence food indicator used to measure the level of extreme poverty in the country. It uses 45 carefully selected goods, with official weights and one price from each one of the two retailers available in the country's dataset. Construction details are provided in the Appendix.

(a) Daily Index



(b) Annual Inflation



(c) Monthly Inflation

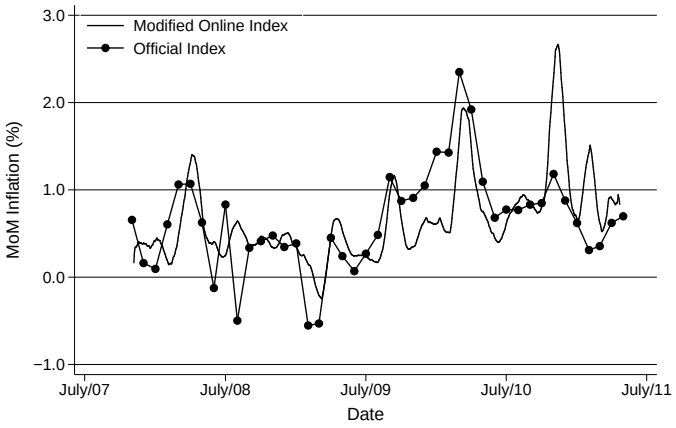
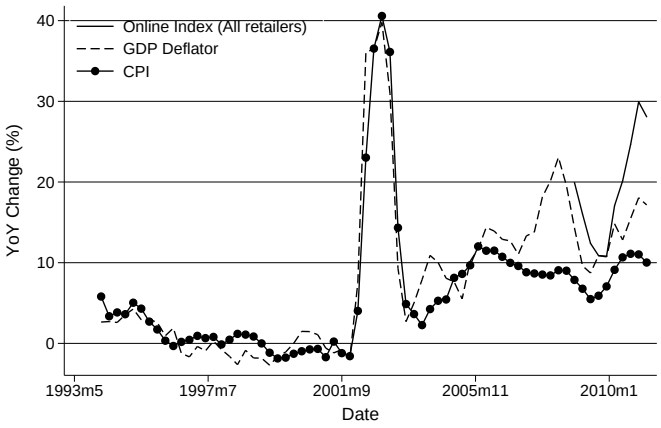


Figure 6: Best approximation: 1/3rd of the observed inflation rate

Notes: The *modified* online index is a simulation that uses just a third of the daily inflation observed with online data. It provides a surprisingly good approximation to the official price index.

(a) GDP Deflator, CPI, and Online Index - Annual Change



(b) Real GDP - Annual Growth Rate

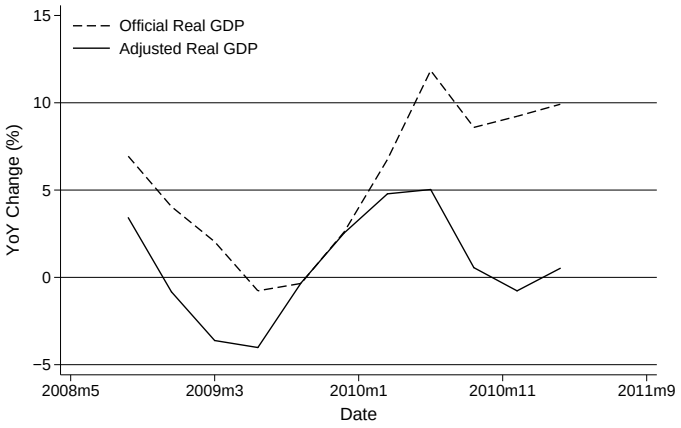


Figure 7: Implications for Real GDP Growth

Notes: The GDP deflator and CPI series in figure (a) co-moved closely together from 1994 to 2006, but started to deviate from 2007 onwards. Although higher than the CPI, the GDP deflator still has less inflation than the online index in the past three years. Assuming the deflator had increased at the same rate as the online index, then we can compute an "Adjusted Real GDP" with a growth rate that is significantly lower than in official estimates.

Tables

Table 1: Online Data Description

	Argentina Retailer #1	Argentina Retailer #2*	Brazil	Chile	Colombia	Venezuela
Starts	10/7/2007	23/7/2007	10/10/2007	10/24/2007	11/13/2007	04/16/2008
Ends	3/24/2011	03/20/2011	03/01/2010	03/20/2011	03/24/2011	03/01/2010
Prices P/day (mean)	11,560	4,790	11,000	12,000	5,000	9,256
Total Products	26,333	10,929	21,804	3,5432	9,166	20,847
Price Changes	204,449	136,781	25,9875	12,0112	76,979	94,808
Category Indicator	Yes	No	Yes	Yes	Yes	No
CPI Weights Covered	48.51%	-	27.93%	31.00%	28.44%	-
Retailer Market Share**	28%	n/a	15%	27%	30%	n/a

Note: *Argentina's Retailer #2 is used only in the robustness results discussed in Section 5.3.2 and Figure 4.
**Market shares are based on the information posted on the corporate webpages of each supermarket.

Table 2: Online vs Official Series

	Argentina	Brazil	Chile	Colombia	Venezuela
Mean Annual Inflation (%)					
Official CPI Index	8.53	5.28	2.44	3.79	27.37
Official Supermarket Index	8.38	5.91	3.19	3.73	29.38
Online Supermarket Index	20.14	4.72	3	4.88	27.43
Correlations between Online and Official Supermarket Series					
Price Index	0.98	0.96	0.97	0.95	0.92
Annual Inflation	0.84	0.09	0.97	0.89	-0.08
Monthly Inflation	0.6	0.5	0.38	0.43	0.18
Regression - Official Supermarket on Online Monthly Inflation Rates (12 lags)					
Constant	0.84	-0.54	0.14	0.03	-1.96
Constant p-value	0.000	0.19	0.17	0.86	0.23
R2	0.9	0.55	0.66	0.59	0.66
Monthly Inflation Rate Volatility (standard deviation)					
Official Supermarket Index	0.57	0.51	0.58	0.48	0.98
Online Supermarket Index	1.11	0.73	0.62	0.76	0.86

Note: The top panel shows that Argentina is the only country where online data does not approximate the average official annual inflation rates. However, the second panel shows that the correlation between the monthly inflation rates is higher in Argentina than in any of the other countries. The discrepancy is therefore in the level of inflation reported, not its dynamic behavior over time. The third panel reinforces this idea with a simple OLS regression of the official monthly rate and 12 lags of the online monthly rate. The R2 is highest in Argentina, which is also the only country where the constant is statistically significant. The fourth panel shows that the official monthly inflation rate is indeed surprisingly stable in Argentina compared to both the online index and the volatility observed in a high-inflation country like Venezuela.