Scraped Data and Sticky Prices

Alberto Cavallo*†

MIT Sloan

January 27, 2011

Abstract

This paper introduces *Scraped Data*, a new source of micro-price information, to provide stylized facts that challenge some commonly-held views in the price stickiness literature. Scraped data are collected from the public webpages of online retailers and have a unique advantage in terms of sampling frequency, size, and country availability. Using a dataset with daily prices of 80 thousand supermarket products in four countries, between October 2007 to August 2010, I present patterns of price stickiness yielding three main empirical results. First, the distribution of the size of price changes is bimodal in most countries, with few changes close to zero percent. Second, hazard functions are hump-shaped, with the probability of a price change increasing for the first 40 to 90 days after a price adjustment. Third, there is daily synchronization in the timing of price changes among closely competing goods. These results differ considerably from previous findings in the literature that uses CPI and scanner data, implying a more important role for adjustment costs and strategic interactions in price setting decisions. The availability of daily prices is essential to measure these empirical patterns and explain some of the differences with previous papers.

JEL Classifications: E30, E60

*Email: *acavallo@mit.edu*

†I owe special thanks to Philippe Aghion, Robert Barro, and Roberto Rigobon for their help with this project. I also wish to thank Alberto Alesina, Gadi Barlevy, Michael Bordo, Jeffrey Campbell, Eduardo Cavallo, Benjamin Friedman, Gita Gopinath, Oleg Itskhoki, Alejandro Justiniano, Pete Klenow, David Laibson, Greg Mankiw, Robert Pindyck, Julio Rotemberg, Tom Stoker, and seminar participants at Harvard, MIT, UCLA, LACEA, and the Chicago Federal Reserve for their helpful comments and suggestions. Surveys of offline prices were conducted with financial support from the Warburg Fund at Harvard University and research assistance from Monica Almonacid, Enrique Escobar, Pedro Garcia, Andre Giudice de Oliveira, and Andrea Albagli Iruretagoyena.

1 Introduction

Starting with Bils and Klenow (2004), a large number of empirical papers have studied the pricing decisions of firms and their implications for price stickiness using micro-level data from CPI and scanner datasets.[1] The main results of this literature have been recently summarized by Klenow and Malin (2009) in a set of ten stylized facts that consistently appear in price data around the world. From this set, three facts are particularly intriguing: 1) the distribution of the size of price changes is unimodal around zero percent, with a large share of tiny price changes, 2) the likelihood or hazard rate of a price change does not increase with the time since the last price adjustment, and 3) the timing of price changes is largely un-synchronized across sellers. These findings are at odds with some basic predictions of standard sticky-price models. In particular, the existence of a large share of small changes and the lack of increasing hazard rates are largely inconsistent with adjustment costs in standard state-dependent (menu-cost) models, while the lack of synchronization suggests little interaction among firm's pricing decisions.

This paper shows that these stylized facts can change dramatically with the introduction of a new source of micro-price information, called *Scraped Data*, which has a unique advantage in sampling frequency, sampling size, and country availability. Using data from four countries, I present three main empirical findings: 1) the distribution of the size of price changes is *bimodal* in most countries, with few changes close to zero percent, 2) hazard functions are *hump-shaped*, increasing for the first 40 to 90 days , and 3) there is a *daily* synchronization in the timing of price changes among close substitutes. Compared to previous findings, my results imply a more important role for adjustment costs and strategic interactions in price setting decisions, and are consistent with recent models that combine elements of time and state dependent pricing such as Alvarez et al. (2010) and Woodford (2009).

Scraped data are collected using a software that scans the underlying code of a retailer's online store to find relevant price information and record it in a database.[2] The software can be set to run automatically every day, visiting the retailer's public website, following the links to pages with product data, and collecting detailed information for all items on display. The resulting datasets contain *daily* prices for *all* products sold by the retailer over time, which greatly reduces the chances of measurement errors. This information can be collected at a

[1] See Nakamura and Steinsson (2008), Klenow and Kryvtsov (2008), Klenow and Willis (2007), Klenow and Kryvtsov (2008), Dhyne et al. (2005), Boivin et al. (2007), Wulfsberg (2008), Gagnon (2007). and the surveys by Mackowiak and Smets (2008) and Klenow and Malin (2009). Scanner data are collected with barcode scanning machines in supermarkets.

[2] The primary language used to write content on the Web is called HyperText Markup Language, or HTML. It is written using tags, such as < *center* >, which provide instructions for the browser to render a page graphically to the user. These tags can also be can be used by the scraping software to identify relevant product and price information in the code.

relatively low cost in any country with retailers that post their prices online, and is therefore easier to access than micro-level CPI data, which statistical offices seldom release, and scanner datasets, which are expensive and hard to obtain. Furthermore, scraped prices can be used to build datasets with identical products, time periods, and sampling characteristics in many retailers and countries at the same time, making it easier to identify the empirical patterns that are truly robust. Here, I use scraped data from four Latin American countries where price stickiness has not been extensively studied before: Argentina, Brazil, Chile, and Colombia. The dataset contains a total of 34 million price observations from over 80 thousand individual supermarket products, scraped on a daily basis between October 2007 and August 2010.

I first show that the distribution of the size of price changes is bimodal in Argentina, Chile, and Brazil. The bimodality is caused by a drop in the mass of price changes close to zero percent. This effect, which is robust over time, is shown graphically using detailed histograms, and formally using two non-parametric tests of modality, Hartigan's Dip and Silverman's Bandwidth tests. The lack of mass close to zero percent is consistent with state-dependent mechanisms, which predict that small changes are not optimal in the presence of adjustment costs.

I then use survival analysis to find evidence of hump-shaped hazard functions in individual price adjustments. Hazards measure the probability of a price change conditional on the time passed since the previous change. In these data, aggregate hazard functions are upward-sloping in all countries for the first 40 to 90 days, and then become downward-sloping over time. Given heterogeneity in individual hazards, the downward-sloping portion can be driven by "survivor bias": goods with more upward sloping hazards tend to disappear faster from the aggregate sample, causing the probability of price changes to drop significantly because only long-lasting duration spells remain. When I separate goods into three –more homogeneous– levels of rigidity, I find that the shape of the hazard function becomes more upward-sloping over time. This suggests that survivor bias is affecting aggregate hazard estimates, and could help explain both the decrease in hazards after a few months in my data and the overall downward-sloping functions previously found by the literature.

Finally, I show that there is *daily* price change synchronization among closely competing goods. I focus on goods displayed next to each other in a single webpage or "URL", which corresponds to a narrowly-defined category such as "extra-virgin olive oil" or "ground beef". To measure synchronization I develop a new measure of independence based on the binomial distribution. For each day, I count the number of products that adjust their prices at the same time. Under a null of no synchronization, this is a collection of independent Bernoulli random variables that is binomially distributed. With the number of competing products in each URL and the observed frequencies of simultaneous changes, I calculate the implied probability that is consistent with the binomial distribution. If there is no synchronization, then this probability is constant. By contrast, when there are incentives to synchronize changes, the probability increases when another product changes its price. The rate at which the probability increases with each additional price changes is a measure of the degree of price change synchronization (or departure from the null hypothesis of no-synchronization). Taking the mean across URLs, I show that there are positive degrees of synchronization in every country, and that they are stronger for price increases than decreases.

These results differ considerably from previous findings in the literature. Klenow and Kryvtsov (2008) and Midrigan (2005b) found unimodal distributions in the size of changes, with a large share of small price changes. Klenow and Kryvtsov (2008) and Nakamura and Steinsson (2008) found hazard functions that are either downward-sloping or flat in US data. Although there is little evidence for price change synchronization in CPI data, a few studies such as Lach and Tsiddon (1996) and Midrigan (2005b) found synchronization with grocery-store data for selected categories of goods at weekly or monthly frequencies. This paper goes further, focusing exclusively on closely competing brands of products and measuring synchronization on a daily basis.

Some of the differences with the literature can be explained by the special sampling characteristics of scraped data.[3] In CPI and scanner data, prices are often adjusted to reflect changes in the use of coupons and other discounts. In scanner data, prices are often computed as weekly averages and "unit values", dividing the sales volume over the number of units sold in a week. Both of these factor can increase the number of small changes observed in the data. Indeed, I show that using weekly averages in my data makes the distributions appear unimodal in all countries.[4] Daily prices are needed to observe the initial increase in hazard rate, because it lasts only a couple of months. The large number of products sampled also facilitates the estimation of hazard rates by increasing the number of available price changes and duration spells. In addition, noth of these sampling characteristics allow me to measure synchronization in narrowly-defined categories goods on a daily basis. This high-frequency interaction between closely competing firms are particularly relevant for theories of strategic price-setting behaviors.

Finally, I test whether *online* and *offline* data have similar price-change dynamics. First,

[3]Inflation levels do not explain the differences with previous papers. My main findings are still present in Chile, where inflation is relatively low and comparable to that of the US and Europe during periods previously studied in the literature. Furthermore, Cavallo and Rigobon (2010) show that the bimodality in the size of changes is also present in a larger set of countries that includes developed economies with lower levels of inflation.

[4]Bimodality is still present in my data when I use a monthly sampling that replicates some characteristics of CPI data. However, the BLS CPI manual describes adjustments that include corrections for coupons, seasonal items, and hedonics that could explain many of the small changes that appear in the data.

I compare the results from a survey of offline prices collected manually in physical stores of the same retailers included in the scraped dataset. Even though price levels are not always identical, there are no significant differences in the timing and size of price changes between the online and offline samples. Second, I show that scraped online prices can also match country-level trends in inflation. Price indexes constructed with online data can closely track CPI statistics in Brazil, Chile and Colombia.[5]

The paper is organized as follows. In section 2, I describe the collection methodology and characteristics of scraped data. In section 3, I present results for the distribution of the size of changes, the hazard functions, and price synchronization. In section 4, I address concerns about the special dynamics of *online* data. Section 5 concludes.

2 Scraped Data

2.1 The Scraping Methodology

A large and growing share of retail prices are being posted online all over the world. Retailers show these prices either to sell online or to advertise prices for potential offline customers. This source of data provides an important opportunity for economists wanting to study price dynamics, yet it has been largely untapped because the information is widely dispersed among thousands of webpages and retailers. Furthermore, there is no historical record of these prices, so they need to be continually collected over time.

The technology to periodically record online prices on a large scale is only now becoming available. Using a combination of web programming languages, I built an automated procedure that scans the code of publicly available webpages, identifies each relevant piece of information, and stores the data in an electronic file. This technique is commonly called "web scraping", so I will use the term *Scraped Data*.

The scraping methodology works in 3 steps: First, at a fixed time each day, a software downloads all public web-pages where product and price information are shown. These pages are individually retrieved using the same URL or web-address every day. Second, the underlying code is analyzed to locate each piece of relevant information. This is done by using custom characters in the code that identify the start and end of each variable, according to the format of that particular page and supermarket. For example, prices are usually shown

[5]Argentina is the only country where scraped data inflation is significantly different from official estimates, with a 17.1% average annual rate over the whole period versus the official estimate of 7.6%. This does not mean that scraped prices are biased, but quite the opposite: official statistics have become widely discredited in Argentina since 2007, when the government intervened the National Statistical Institute. See Cavallo (2010)

with a dollar sign in front of them and two decimal digits at the end. This set of characters can be used by the scraping software to identify and record the price every day. Third, the scraped variables are stored in a database that contains one record per product per day. Along with the price and other characteristics, retailers show an id for each product which allows me to follow a product over time.

2.2 Comparing Data Sources

The differences between scraped data and the two other sources of price information commonly used in studies of price dynamics, CPI and Scanner Data, are summarized in Table 1.

Scraped data have important advantages that make them a unique source of information. First, these datasets contain *daily* prices, which can greatly reduce measurement error biases, as is later shown in this paper. Second, the data are available for a much larger set of countries. In this paper, I focus on *developing* countries, where scanner data are scarce and product-level CPI prices are seldom disclosed.[6] Third, scraped data contain detailed information on the full array of a retailer's products. In particular, the ability to identify products displayed next to each other plays a key role in measuring price synchronization among close substitutes. Fifth, there are no *forced* item substitution, which occur frequently in official statistics to measure inflation in out-of-stock, seasonal or discontinued products.[7] Sixth, scraped datasets are directly comparable across countries, with prices on the same categories of goods and time periods. This makes it possible to perform simultaneous cross-country analyses. Finally, scraped data are available on a real-time basis, without any delays to access the information. This can be used to provide estimates of stickiness that quickly capture changes in the underlying economic conditions.

Still, scraped datasets have some disadvantages. First, they typically cover a much smaller set of retailers and product categories than CPI prices. This limitation will recede over time, as a growing number of firms start posting their prices online. It is not a major issue for this paper because supermarket products represent over 40% of all CPI expenditure weights in

[6]The study of stickiness in developing countries is rare in the literature. A recent exception is Gagnon (2007), who provides a detailed analysis of sticky prices in Mexico using disaggregated CPI data.

[7]Forced substitutions occur in CPI data when the agent surveying prices does not find the item she was looking for, and decides to replace it with another product, which becomes the surveyed item from then on. In practice, if the old item is supplied again and/or the new product was being supplied before, official statistics ignore their prices, effectively censoring the price series. By contrast, in scraped data, prices are recorded from the first moment they enter the sample until the last day they have been offered to consumers, which solves substitutions for items that go temporarily out of stock. I do not attempt to link price series of goods that are discontinued with those of similar goods that may replace them, but such substitutions could be potentially achieved with these data using product descriptions.

these four Latin American countries (versus only 12% in the US). Second, scraped data do not include information on quantities sold, as scanner datasets typically do. In the context of measuring stickiness, quantities can be useful to measure elasticities or determine category weights in frequency statistics, but they are not needed for the stylized facts discussed in this paper.

2.3 The Data in this Paper

I use a dataset with more than 34 million supermarket prices in Argentina, Brazil, Chile, and Colombia. The data come from the online price tables of four different retailers, one in each country, from October 2007 to August 2010.

All the supermarkets included in the dataset are market leaders in their respective countries, with market shares of approximately 28% in Argentina, 15% in Brazil, 27% in Chile, and 30% in Colombia. With hundreds of physical stores, they also sell online in cities such as Buenos Aires, Santiago, Rio de Janeiro, and Bogotá. Every day, for nearly three years, I accessed these websites and recorded all this information for every good on display. Because buyers cannot physically see the products, retailers make an effort to display detailed information on each item, including a price, the product's identification number (id), name, brand, package size, category, and whether it is on sale or under price control.

Table 2 provides details on each country's database. There are roughly 25,000 daily prices for each country in Argentina, Chile, Brazil, and 10,000 in Colombia. The initial date for each database differs by a few days around October 2007, but they all end on August 13th 2010. To compare results for the same product categories across countries, I matched each supermarket's classifications into 95 standardized categories containing a large variety of foods and household items.[8] Products can be further classified into "pages" or URLs. A URL uniquely identifies a group of competing products, corresponding to the narrowest grouping of items in each supermarket.

Tables 3 to 5 present general price change statistics in all countries. First, as a measure of price stickiness in Table 3, I calculate the median frequency and implied durations following the methodology in Bils and Klenow (2004). On one extreme, prices are stickiest in Chile and Argentina, with the daily median frequency of 0.015 and an implied duration of 66 days. They are followed by Colombia, with a median frequency of 0.019 and an implied duration of 52 days. Finally, Brazil is the most flexible country, with a median frequency of 0.027 and an implied median duration of 36 days. Price stickiness is not necessarily correlated with inflation, as seen in Table 4. Argentina is one of the stickiest countries, but also has the

[8]See Table A4 in the Appendix for a complete list of product categories. These are based on the ELI classification used by the US Bureau of Labor Statistics.

highest inflation by far, with an average annual rate of 17.1% over this period, followed by Brazil at 5.1%, Colombia at 4.2%, and Chile at 2.7%.[9]

3 Re-evaluating Three Stylized Facts

3.1 Standard Models and Empirical Literature

Many microeconomic mechanisms have been proposed to explain why prices are sticky, but most of them can be broadly classified into either time-dependent or state-dependent pricing behaviors. These models have different implications for the stylized facts discussed later in the paper.

In time-dependent pricing (TDP) models, the decision to adjust prices is driven by time. In early models, the timing was exogenous and adjustment could occur after a fixed number of periods, as in Taylor, 1980, or randomly every period, as in Calvo, 1983. More recent versions of these models can endogenously generate the timing of adjustment with imperfect information and observations costs. In all cases, the price stickiness (and therefore the real effects of monetary policy) comes from the fact that firms are not constantly monitoring their prices, which is done only at specific times.

In state-dependent pricing (SDP) models, by contrast, the decision to change prices depends directly on how far the current price is from the optimal price. Early examples include the menu cost models of Barro (1972) and Sheshinski and Weiss (1977), and more recently Dotsey et al. (1999) and Golosov and Lucas (2007). In these models, firms are able to change their prices at any time, but must pay an adjustment cost to do so ("menu cost"). [10] Whether a firm changes its price or not depends on whether the benefits from adjustment (given by

[9] I measure inflation in all countries using simple price indexes with online scraped data. Section 4 compares these findings with official statistics and provides more details on the use of daily online indexes. In principle, we would expect price stickiness and inflation to be negatively correlated across countries. That is the case for Chile, Colombia, and Brazil, but Argentina breaks the pattern completely. Although the low frequency of changes in Argentina appears compensated with a large mean size of changes, as seen in Table 5, the size of price increases and decreases, when computed separately, is not larger than in the other countries. The mean size of changes is large in Argentina mostly because there are more increases than decreases, as can be seen in Table 5. The last rows in Table 3 suggest that what is important for inflation is not the overall level of stickiness, but the relative stickiness of price increases over price decreases. Section 3 in the Appendix shows similar results across categories of goods within countries

[10] "Menu costs" may include various kinds of adjustment costs, such as labor costs to stamp new prices, managerial costs to make a decision, or even "customer anger" costs linked to the consumer's reaction after a price adjustment. An example of models that include the latter are "Fair Pricing" models, such as Rotemberg (2005), Rotemberg (2008), and L'Huillier (2009). These are based on the idea that prices are sticky because firms do not want to antagonize customers. Blinder et al. (1998) found in a survey of price setters that this was a major concern for firms setting prices in the US. I explore some evidence for this type of mechanisms in the Appendix, using data from Argentina.

the curvature of the profit function) are greater than the menu costs.

A more recent strand of the literature combines TDP and SDP mechanisms using a mixture of information and adjustment costs. Examples include Alvarez et al. (2010), Woodford (2009), and Bonomo et al. (2010). These are imperfect information models where firms must pay an information-gathering cost to compute their optimal price, and then pay an additional menu cost if they decide to adjust the current price. If the information cost is relatively high, firms will prefer a time-dependent rule. If, by contrast, the menu cost is relatively more important, firms will tend to have a state-dependent pricing rule.

Over time, the empirical literature has produced a set of stylized facts that can be used to test these models. Most of them where found first in the US CPI data and later confirmed with European CPI and US scanner datasets. Klenow and Malin (2009) provide ten stylized facts that are becoming part of the conventional wisdom on price stickiness. In this paper, I will re-evaluate three of them: 1) there are many small price changes, 2) the hazard rate of price changes does not increase with the age of the price, and 3) the timing of price changes is little synchronized across sellers. I focus on these three facts because standard models have sharp predictions about them, so they can greatly influence the way models are built in the future.

3.2 Bimodal Distributions of the Size of Changes

Standard SDP models, such as the menu-cost model of Golosov and Lucas (2007), predict *bimodal* distributions of the size of price changes, with relatively little mass close to zero percent. The intuition is that the benefit from correcting small deviations from the optimum price is not enough to cover the adjustment cost. These costs therefore create a dip in the distribution of changes around zero percent. By contrast, in standard TDP models such as Calvo (1983), any size of price change is possible. When the time comes to adjust, firms will make all price changes, regardless of their size, so that the predicted distribution of sizes tends to be unimodal around zero percent.[11]

Several papers in the literature have looked at the distribution of price changes using CPI and Scanner data. Examples include Kashyap (1995), Midrigan (2005b), Klenow and Willis (2007), Baudry et al. (2007), and Kackmeister (2007). In most cases, the distributions where found to be unimodal and centered at zero percent, with a large number of small changes. This has shaped the theoretical literature considerably. For example, Midrigan (2005b) builds

[11]In fact, the distribution of the size of changes will inherit the properties of the distribution of shock to marginal costs. If shocks are approximately normal, the distribution of changes will be *unimodal* with a mode that depends on the overall level of inflation. If inflation is low, the mode is close to zero percent, and there are many small price changes. If inflation is high, the shape would still be unimodal but the mode would shift to some positive value given by the average level of inflation.

a SDP model with economies of scope in menu costs that can generate small price changes, while unimodality contributes to the conclusion by Woodford (2009) that the predictions in Calvo's TDP model are more reliable than what has often been suggested. Looking forward, the relative importance of small price changes has key implications for the parametrization of recent models like Woodford (2009) and Alvarez et al. (2010).

Scraped data can be used to examine the distribution of the size of changes in much greater detail. The sampling characteristics of the data means that there are between 108K and 366K observed price changes in each retailer, which allows me to focus on what happens with the mass of price changes close to zero percent by using histograms with very narrow bins (0.1%) in Figures 1 and 2. These distributions are conditional on a price change (i.e. no mass at zero), and are truncated at an absolute value of 50% to facilitate the graphical analysis.

The most striking feature for most distributions is their bimodality, with a sharp dip of mass close to zero percent. This happens in Argentina, Brazil, and Chile. The effect can be seen in the smoothed kernel densities shown for each graph, while Table 5 presents some statistics that emphasize this lack of small changes. The share of changes below 1% in absolute value is only 4.2% in Argentina, 4.3% in Brazil, and 3.6% in Chile, significantly lower than the 11.3% reported by Klenow and Kryvtsov (2008) using US CPI data. The shape of these distribution robust over time in all countries, as can be seen in Figure 3, where I plot one distribution for each year of data. Furthermore, the *dip* in these distribution does not disappear once sales are excluded, as shown in Figure 6

Overall, this bimodality provides evidence of the existence of adjustment costs in price changes, and is consistent with the predictions of SDP models like Golosov and Lucas (2007), and also mixed imperfect information models like Alvarez et al. (2010). The only case where the distribution appears unimodal is Colombia, where 7% of price changes are smaller than 1% in absolute value. The shape of the distribution closely resembles the findings in previous papers in the literature, not only with a large share of small changes, but also relatively fat tails. The unimodality could be an indication of TDP, but is also consistent with menu costs under certain conditions. For example, there could be different menu costs for different goods (as in Dotsey et al., 1999), one menu cost for a large number of goods (as in Midrigan, 2005b), or simply smaller menu costs for online pricing.

It is important to formally test the departure from unimodality and show that, not only graphically but also statistically, the distributions in Figures 1 and 2 do not have a single mode at zero percent. In the statistics literature, two non-parametric tests are commonly used to determine modality: Hartigan's Dip and Siverman's Bandswidth tests.[12]

[12]Non-parametric methods are important to avoid making ex-ante assumptions on the number of modes.

Hartigan's Dip is a simple test of unimodality. It relies on the fact that the cumulative distribution function of a density function f with a single mode at m_f is convex on the interval $(-\infty, m_f)$ and concave on the interval (m_f, ∞).[13] In other words, at the left side of the mode, the density is non decreasing, while the opposite occurs at the left of the mode. With this insight, one can find the unimodal distibution that minimizes the difference with the observed empirical distribution. This difference is measured by the dip statistic, which can be used as a sort of "score" to measure the departure from unimodality. Positive dip values provide evidence to reject the null hypothesis of unimodality. [14]

Silverman's Bandwidth method can be used to test for multiple modes. It uses the non-parametric smoothed kernel density to evaluate the number of modes in an empirical distribution. The basic insight in Silverman (1981) is that the larger the smoothing applied, the fewer the number of modes in the estimated density. So for the null hypothesis of unimodality, he proposed using as a test statistic the minimum smoothing required for the density to have a single mode. Large values of this statistic (the "critical bandwidth") are evidence against the null hypothesis of unimodality, because they mean that larger degrees of smoothing are needed to eliminate additional modes in the density estimate. The statistical significance of the score can be evaluated using a smoothed bootstrap method.[15]

Table 6 reports the results for both tests in all countries. The dip score, shown in column one, is consistent with the graphical analysis: Argentina has the largest departure from unimodality, while Colombia the smallest. Statistically, we can reject the null hypothesis of unimodality in all countries. Silverman's test provides similar results: the critical bandwidth is largest in Argentina and smallest in Colombia. We can also reject the null of unimodality in every country (for Colombia at a higher statistical significance of 3%).[16]

---

[13]See Hartigan and Hartigan (1985)

[14]To determine the statistical significance of a positive dip, Hartigan and Hartigan (1985) sets the null hypothesis equal to the uniform distribution, for which, asymptotically, the dip value is stochastically largest among all unimodal distributions. Hartigan and Hartigan (1985) also show that this is not always the case with small samples. To address this concern, I use a calibration of the dip test proposed by Cheng and Hall (1998).

[15]See Henderson et al. (2008) for more details for both statistical tests.

[16]Note that with Silverman's test we stop rejecting in Colombia when there are two or less modes, and in the other countries when there are three or less modes. In other words, the test suggests there is bimodality in Colombia and three or more modes in the other countries. The reason is that this test is sensitive to tiny bumps in the distribution, a problem that is derived from the use of a single bandwidth in the kernel smoothing estimates. This leads to frequent rejections of the null hypothesis in large samples. Another issue, which applies to both tests, is that they do not tell us if the rejection of unimodality is caused by what happens around zero percent (the focus of the sticky price literature). Cavallo and Rigobon (2010) propose an alternative modality test that solves both issues by ignoring tiny bumps in the distribution and focuses on the relative mass close to zero percent.

3.2.1 Differences with the Literature

As mentioned before, the bimodality in the size of price changes is at odds with previous findings in the literature.

The differences with papers that use scanner data appear to be caused by the use of daily data. In scanner datasets, prices are typically constructed as "unit values", taking the ratio between revenues and quantities sold for a product during a period of time (usually a week). This means that prices are being averaged along two dimensions. First, at the same point in time there may be different prices for different units sold of the same product, because consumers sometimes purchase products with coupons, loyalty cards, or in bundles. This can create additional price changes in the data. Second, because scanner data are reported on a weekly basis, prices are also averaged over time. As Campbell and Eden (2005) pointed out, this can make one large price change appear like two smaller ones. For example, consider a three-week period with a single price change in the middle of the second week. Computing weekly averages would yield three different prices, one for each week, and two small price changes instead of a single larger change. I can test this by looking at the effects of weekly averages in my own data, as shown in Figure 4. Using weekly averages greatly increases the number of small price changes and the bimodality disappears in every country.

The differences with CPI results are harder to reconcile. Monthly-sampled prices could potentially change the distribution of sizes by aggregating price changes over time. Yet the effect can have opposite implications depending on the nature and persistence of price changes. In a low inflation context with many temporary shocks, several price changes that go in opposite directions could end up looking like a single small change in monthly data. In a high inflation context, the opposite would be true: persistent increases in prices could be accumulated over a month and look like a single larger change. I simulated the effects of monthly sampling with scraped data by randomly selecting a price for each product every month and re-calculating the distributions, but Figure 5 shows that monthly sampling has no impact at all in the number of modes in my data.

A more likely explanation for the differences with CPI data is directly related to the way official prices are recorded. In the US, for example, the BLS Handbook of Methods[17] describes several adjustments in individual prices that can affect the distribution of the size of changes. First, changes in a price spell can occur because of *forced item substitutions* that happen when an item can no longer be found in a store. In these cases, the BLS estimates a price change using the average price change for that category of products or using hedonic quality adjustments. Second, even when no product substitutions occur, the

[17]See , Chapter 17, pages 30 to 33.

BLS sometimes imputes prices that are considered to be temporarily missing, like seasonal items. Third, individual prices can also be adjusted for coupons, rebates, loyalty cards, bonus merchandize, and quantity discounts, depending on the share of sales volume that had these discounts during the collection period. Finally, some food items that are sold on a unit basis –like apples– are sometimes weighted in pairs to calculate an average-weight price. Unfortunately, at this stage I do not have access to CPI prices to know how frequent these changes really are in the data.[18]

3.3 Hump-Shaped and Upward-Sloping Hazards

A second stylized fact that can be used to distinguish between models is the shape of the hazard function. The hazard is the instantaneous probability of price change at time t, conditional on the price not changing until that point in time. In Calvo (1983)'s TDP model, the hazard function is flat because the probability of price change is fixed and exogenously determined. In Taylor (1980)'s model, the hazard is equal to one at the time when all price changes take place (eg. a month). With heterogeneity across goods, this can be generalized to have hazards with "spikes" at given frequencies. By contrast, in SDP models hazard functions tend to be upward-sloping. The intuition is that inflation (or deflation) increases deviations from the optimal price over time, so as the price gets "older", the conditional probability of a price change also rises. Upward-sloping hazards are intuitively appealing, but there is no evidence for them in the current empirical literature. Nakamura and Steinsson (2008) found evidence of downward sloping hazards in US CPI prices, while Klenow and Kryvtsov (2008) found mostly flat hazard functions in similar data.

Scraped data has two major advantages for the study of hazards rates. First, we can look at hazards in countries with higher inflation that the US or Europe. In contexts where aggregate shocks are strong and persistent, it should be easier to find evidence of upward-sloping hazards. Second, we can see how the probability of change varies on a daily basis, which is important when most goods adjust within a few months.

I measure hazard rates using standard Survival Analysis, which studies the time elapsed from the "onset of risk" until the occurrence of a "failure" event. In a price-setting context, we are interested in the time between the firm's optimal price adjustments. The set of

[18]There are other possible explanations for the differences with both CPI and scanner data. First, the countries studied here have higher inflation levels. Inflation could be moving the mass of price changes away from zero, simply because the marginal cost shocks experienced by the firms are larger. However, this does not appear to be an important explanation because I find bimodality in both high-inflation Argentina and low-inflation Chile (see the critical bandwidth scores in Table 6). In addition, the average size of price increases and decreases does not vary much across countries, as can be seen in Table 5, so it is hard to argue that inflation moves prices further away from zero. Second, this paper uses online prices that could behave differently from offline prices. In section 4, I use a survey of offline prices to show that this is not the case.

constant prices between these two dates is called a "price spell" and the duration (measured in days) is the length of the spell. I use a non-parametric approach due to Nelson (1972) and Aalen (1978), which does not require any distributional assumptions.[19] It provides a simple estimate of the cumulative hazard function $H(t)$, given by:

$$\widehat{H}(t) = \sum_{j|t_j \leqslant t} \frac{c_j}{n_j} \tag{1}$$

where c_j is the number of price changes at time t_j and n_j is the number of price spells that can still change at time t_j. The incremental steps c_j/n_j are an estimate for the probability of price change at t_j, taking into account only those price spells that have survived until that point in time.

To obtain the smoothed hazard function $\widehat{h}(t)$, I take the discrete changes in $\widehat{H}(t)$ and weight them using a kernel function:

$$\widehat{h}(t) = \frac{1}{b} \sum_{j \in D} K\left(\frac{t - t_j}{b}\right) \Delta \widehat{H}(t_j) \tag{2}$$

where K is a symmetric kernel density, b is the smoothing bandwidth, and D is the set of times with price changes. Following the literature, I implicitly assume that each price change restores the optimum price and treat all duration spells independently. I include right-censored spells, because we know for certain how old they are at each point in time, affecting n_j in equation 1. However, I exclude left-censored spells, for which the time since the last adjustment is unknown. I further exclude sale prices in Argentina, Brazil, and Colombia. Sale events are likely to have special dynamics and have short durations that can increase survival biases, as explained below.[20] Finally, spells of all durations are used to construct hazard estimates, but since only a small fraction of spells lasts more than six months as shown in Figure 7, I focus the discussion on the shape of the hazard functions during the first six months.[21]

Figure 8 plots the estimated hazard with 99% confidence intervals, with a y-axis scale matched to facilitate comparisons. In all countries, the aggregate hazard function has a hump-shaped pattern. For a period of time lasting from 1 to 3 months, the probability of

[19] I choose this method because I want to study the *shape* of the hazard function $h(t)$, not the effects of any covariates. My results are robust to the use of a semi-parametric Cox model that can incorporate covariates and account for unobserved heterogeneity at the category level.

[20] The median duration of sale prices is extremely short, with 6, 7, and 17 days in Argentina, Brazil, and Colombia respectively. Hazards including sales are shown in the Appendix. Sales significantly increase the hazard rate at durations with one and two weeks, although hazard functions continue to be hump-shaped.

[21] Hazards are either flat of downward sloping after six months, but there are few spells on which to base the analysis. See the Appendix for the full three-year hazard functions.

price change increases with the age of the price spell.[22] In Argentina, the hazard reaches its maximum at approximately 90 days. In Chile, this occurs sooner, at 40 days (likely affected by the inclusion of sales, which tend to have very short durations). Both Brazil and Colombia have peaks at 60 days.

These hump-shaped patterns do not fit standard TDP or SDP models. However, they could potentially be explained with TDP models if there are multiple firms adjusting at different times, with a majority of goods doing so at 40, 60, or 90 days. They could also be explained with SDP models when temporary shocks are relatively important, as Nakamura and Steinsson (2008) point out, because these shocks would cause a reversal of the adjustment within a short period of time.

The evidence for SDP is reinforced by the fact that Argentina, with its high-inflation rates, has an upward-sloping hazard for a longer period of time. Even in standard SDP models, the higher the inflation rate, the more upward-sloping hazards become, because the deviation from the optimal price increases over time.

The methodology makes it hard to find upward-sloping functions because I am not able to correct for the "survival" bias caused by heterogeneity across products in the shape of individual hazards. This bias is illustrated in Figure 9 with a hypothetical example. Consider two types of goods with upward sloping hazards. One type changes prices more frequently, so it has higher hazard rates and will disappear from the sample faster. If we estimate the aggregate hazard for both goods, initially we would be using spells from both of them, but at some point in time we would start using only spells from goods with the lower hazard rates. This"survival" bias would tend to flatten the estimate, creating hump-shaped results. This is a well-know problem in survival analysis, for which there are not simple solutions.

I find evidence of the existence of survival bias in Figure 10, where I separate goods in terms of their average durations and re-estimated their hazard functions. The dotted line represents goods that have average durations of less than 50 days, the dashed line is for goods with average durations of 50 to 100 days, and the solid line represent stickier goods with average durations over 100 days. The patterns are indeed very similar to the example in Figure 9. The more flexible goods still have hump-shaped patterns, but the confidence interval widens as the hazard becomes flat, reflecting the fact that there are fewer spells with which to obtain an estimate. Furthermore, as we separate goods into different categories, each one of these hazards became more consistently upward sloping.[23]. The hump-

[22] Also, the overall level of each hazard function is consistent with the fact that prices more flexible in Brazil and Colombia, and stickier in Chile and Argentina.

[23] This is not caused by the peak moving further to the left, which naturally occurs with stickier goods if the hazards are really hump-shaped. The relevant finding is that hazards for all categories become closer to straight upward- sloping lines as we separate them into different categories

shaped patterns does not disappear completely because each one of these three hazards is itself constructed by aggregating across many goods, and therefore they are still affected by survivor biases. Completely controlling for heterogeneity at the individual-good level requires far more price changes per product that I currently have, but these results are an indication that the underlying hazard rates tend to be far more upward sloping than what the aggregate estimate is showing. As such, my results should be taken as a lower-bound indication for the existence of upward-sloping hazards.

Once again, my results differ considerably from previous findings in the literature, where hazards are typically flat or downward sloping hazards. One possible reason for this difference is that these scraped prices come from higher inflation economies, were SDP models predict more upward-sloping hazards. In principle, if there is some sort of state-dependent pricing behavior, it should be easier to find upward-sloping hazards in a country like Argentina. However, a major reason for the difference is related to the use of daily data. At least for this sample of products, where the mean duration is just a few months, having daily data is essential to obtain the upward sloping portion of the aggregate hazards.

3.4 Synchronization of Price Changes

A third stylized fact that has received attention in the literature is the degree of synchronization in the timing of price changes. In TDP models, higher synchronization reduces the persistence of the real effects of monetary policy. In SDP models, synchronization is closely linked to *strategic complementarities*. Firms selling strategic complements will imitate each other's actions by attempting to synchronize the timing of their price changes.[24]

Synchronization with product-level price data has been studied before in the macro literature, but at lower sampling frequencies, with smaller samples and broader categories of goods. For example, Lach and Tsiddon (1996) found evidence of monthly within-store synchronization for wines and meat products in Israel, while Midrigan (2005b) found weekly synchronization with scanner data in the US. To the best of my knowledge, no paper has been able to study synchronization on a *daily* basis and focusing only on goods that are close competitors.

[24] See Cooper and Haltiwanger (1996) for a general discussion of how agents have an incentive to synchronize discrete decisions under strategic complementarities. Strategic complementarities have been introduced in state-dependent models as a form of *real rigidity*, in the spirit of Ball and Romer (1990), to increase the real effects of nominal frictions. The intuition is that while some firms are free to adjust their prices, they may decide to wait until competitors react to the shock. The fact that some firms have not yet adjusted (due to a nominal friction like menu costs), could make other firms delay their own price changes (a real rigidity). See Klenow and Willis, 2006 and Burstein and Hellwig, 2007. See Midrigan (2005b) for results on synchronization with scanner data and Neiman (2008) for evidence of synchronization in international trade data.

Scraped data are especially well suited to find close competitors because a URL indicator is available to identify products displayed next to each other. In addition, the daily nature of the data is key for price interactions in high-inflation countries because, as Lach and Tsiddon (1996) noted, a sufficiently long sampling interval would ensure that all prices appear to change simultaneously regardless of the degree of synchronization.

To focus on synchronization between competing firms, I consider only one product *per brand* in each URL. This eliminates simultaneous price changes caused by the same good with different package sizes and flavors, or different goods sold by the same firm under a single brand.[25]

3.4.1 A Non-Parametric Test of Synchronization

To measure the degree of synchronization in each URL, I propose a simple method based on the binomial distribution.[26] I start by looking at Y_{jt}, the number of products that change their price in URL j on day t:

$$Y_{jt} = \sum_{i} X_{ijt} \tag{3}$$

X_{ijt} is a binary indicator equal to one if good i changed its price at time t. Let $P_{ijt} = Pr(X_{ijt} = 1)$ be the probability that the price of that product changes that day. Then X_{ijt} is a Bernoulli random variable, with success probability p_{ijt}. Assuming all products in an URL are identically distributed with a constant probability of price change, then $p_{ijt} = p_j, \forall i, t$.

If there is no synchronization in price changes, then X_{ijt} is independent across products, and Y_{jt} is distributed as a $Binomial(N_j, p_j)$, where N_j is the number of products in the URL. Therefore, to determine whether prices are synchronized or not, we can observe the distribution of Y_{jt} in each URL and compare it to the binomial distribution. This is done by computing the *implied probabilities* under the assumption of a binomial distribution. That is, given the number of products N_j in a particular URL, we can find the individual probability p_j that would generate the observed frequency of simultaneous changes under the assumption of a binomial distribution. If price changes were really independent, then these implied probabilities would be constant; however, if there are incentives to synchronize changes, the implied probabilities would increase with the number of items adjusting at the

[25] For each brand, I keep the product with the largest number of price observations available. The results are qualitatively robust to a random selection criteria per brand, or the inclusion of all products in an URL. This sample still includes products from the same manufacturer that are sold under different brands (within the same URL). Unfortunately, there is no manufacturer information for individual products or simple ways to link brands to manufacturers.

[26] Similar results can be with the Fisher-Konieczny index (Fischer and Konieczny, 2000).

same time.

To illustrate this methodology, I use the "Rice" URL in each country as an example. First, I compute the distribution Y_{jt} in Figure 11, by plotting the fraction of days with a given number of synchronized price changes. For example, the value at two (e.g. 0.046 for Argentina) indicates the fraction of days where only two products in that URL changed their price, or $Y_{jt} = 2$.

Second, under the hypothesis of a binomial distribution, I calculate the implied probabilities. For example, since there are 25 products in the "Rice" URL for Argentina, when $Y_{jt} = 2$ the implied probability p solves the equation:

$$\Pr[Y_{jt} = 2] = 0.046 = \begin{pmatrix} 25 \\ 2 \end{pmatrix} p^2(1-p)^{25-2} \tag{4}$$

In this case, p is equal to 0.0145. The same calculation is repeated for all values of Y_{jt}, up to $Y_{jt} = 10$.[27]

Figure 12 plots the implied probabilities for the "Rice" URL in all countries. In all cases, the probabilities increase with the number of simultaneous price changes, consistent with synchronization.

For a single URL, we can measure the degree of synchronization by fitting a linear trend and obtaining the slope of implied probabilities. The higher the slope, the larger the deviation from the binomial distribution and, therefore, the stronger the synchronization. We can further generalize the analysis and average all URL slope coefficients to get a country-level measure of synchronization.[28]

Table 7 shows positive levels of daily synchronization in all countries. The average slope of implied probabilities is 0.008 in Argentina, 0.007 in Brazil, 0.007 in Chile, and 0.012 in Colombia. Compared to the median frequencies (the *unconditional* probabilities of daily price change reported in Table 3), these coefficients imply that the probability of price change increases by 63% in Argentina, 25% in Brazil, 46% in Chile, and 63% in Colombia every time an additional price change occurs at the same time.

Table 7 also shows that synchronization is not affected by the exclusion of sales. There is, however, a large difference between price increases and decreases, when considered separately. In Argentina, Brazil, and Colombia, price increases are more synchronized than price

[27]To obtain a unique solution, I solve for p in:

$$\frac{\Pr[Y_{jt} = k]}{\Pr[Y_{jt} = 0]} = \begin{pmatrix} N_j \\ k \end{pmatrix} \left(\frac{p}{1-p}\right)^k \ , \ \forall k \in [1, 10]$$

[28]Only urls with at least 3 products are considered. In addition, urls with slope coefficients that are not statistically significant in a 95% confidence interval are assumed to have no synchronization.

decreases.

As mentioned before, if there are strategic complementarities across products, firms will try to match the timing of each other's price changes. Still, there are other possible causes for price change synchronization within URLs. Prices could be driven by a common sectoral shock affecting the category, or supermarkets may choose to change the prices of many similar products at the same time to save on adjustment costs.[29] It is likely that these reasons are playing a complementary role in explaining the daily synchronization observed in the data.

4 Online vs Offline Prices

Online purchases are still a small share of transactions in most countries, so it is natural to question the representativeness of scraped data. In this section, I address this concern in two parts. I first consider whether online and offline prices behave similarly for each retailer. I then examine whether these supermarkets are representative of each country's aggregate inflation trends.

4.0.2 Matching Offline Price Behaviors

Between December 2008 and February 2009, I conducted simultaneous surveys of *offline* and *online* prices in all the retailers where I collect scraped data. These surveys took place in Buenos Aires, Santiago, Rio de Janeiro, and Bogotá, with the help of four local volunteers. They were asked to select any branch of the supermarket and randomly buy 100 products, divided in 10 pre-defined categories. These categories were chosen to ensure some variety in the type of goods purchased: Dairy, Bakery, Beverages, Cereal and Flours, Fats and Oils, Meats, Pasta and Rice, Fruits and Vegetables, Cleaning Products, and Bath Products. After the first purchase, I used the receipts to get unique product ids and check whether the same items where sold online or not.[30] Those items that could not be matched to the online database were removed from the product list for subsequent purchases. In total, four purchases took place in each supermarket, at 15-day intervals, always in the same branch. The same items were bought every time, with identical flavors and package sizes. If a product was out of stock, no price was recorded for that day, but we attempted to buy the product again in subsequent purchases.

[29] For example, when there is a fixed cost to walk to an URL and manually change prices (or connect to a database and input the new values) but low marginal costs to change the price of additional items within the URL. This is the case of increasing returns to scale in adjustment costs, studied by Midrigan (2005b).

[30] In Argentina, Brazil, and Colombia, the matching was based exclusively on product ids. In Chile the matching was based on the item's name, description, and package size.

Table 8 shows the results from this validation exercise. The percentage of offline products that were also available online ranges from 74% in Colombia to 100% in Argentina. Most of the products that could not be matched are raw-food items, which tend to be re-packaged for online sales and have different id numbers and descriptions.

I compare prices both in terms of their *levels* and the timing and size of *changes*. Even though price levels are not always the same across samples, online and offline price *changes* behave similarly in terms of timing and size of adjustments in all countries.

In Chile, the matching of price levels is extremely close. 361 out of 388 comparable prices were exactly the same. The 27 price discrepancies, which averaged 2% in size, were concentrated in only 12 goods (mostly raw-food products), so that 89% of products have identical price levels across samples. In Argentina, price levels are typically higher online: 252 out of 323 comparable prices were higher in the scraped database. Yet in nearly every case, there was a difference of 5% across samples. A constant markup means that price *changes* are highly correlated: 93% of products have identical price change series (conditional on a change, with 1 for price increases and -1 for price decreases). Furthermore, the ratio of all price changes over total observations is 0.215 in both samples, and the mean size of these changes is close, at 1.6% offline and 1.4% online.

The cases of Brazil and Colombia are more complex, but the samples still show similar price change behaviors. The evidence suggests these supermarkets treat their online stores as independent branches, with similar strategies in terms of price adjustments. In Brazil, price levels are identical only 42% of the time. Still, in terms of price changes, the matching is much better because most of the differences are concentrated in a small share of products: 75% of all goods have identical price change series across samples. For all products, the ratio of changes over total observations is 0.356 offline and 0.411 online, while the mean size of changes is 4.9% offline and 5.3% online. In Colombia, the matching of price levels, at 29%, is even lower than in Brazil, but price differences are smaller. The matching of price changes is still relatively high, with 67% of identical price changes series. In terms of the frequency of price changes, both samples have identical ratios of changes over total observations, at 0.433. Finally, the mean size of changes is also very close, with 8.1% offline and 8.2% online.

The offline survey can also be used to compare patterns in the distribution of the size of changes. Although the sample is very small, Figure 13 shows that the country with the highest share of changes close to zero percent is Colombia, just like we conclude with the use of online scraped data.[31]

[31]Unfortunately the data is not enough to compute any hazard functions or measure synchronization.

4.0.3 Tracking Official Statistics

Another way to test the validity of scraped data is to see if they have inflation dynamics that resemble those obtained from CPI statistics, which are constructed using surveys from a large number of offline retailers. In Figure 14, I plot an online supermarket index together with the official CPI in each country. The supermarket index is constructed using daily scraped data with a simple methodology described in the Appendix.[32]

Figure 14 shows that daily online indexes closely track the official CPI series in Brazil, Chile, and Colombia. Although these scraped indexes are meant to be rough approximations to the official statistics, they can still capture major trends in inflation extremely well. This can also be seen in Figure 15, where I plot a daily estimate of annual inflation in every country.

Argentina is the only case where scraped indexes are not consistent with official statistics. The scraped data show a mean annual inflation rate of 17.1%, but the mean CPI inflation was only 7.6% per year during this time period. However, the difference is not surprising because official data have become widely discredited since January 2007, when the government started interfering with the construction and publication of price indexes at the National Statistics Institute (INDEC).[33]

5 Conclusions

This paper introduces a new way of collecting price data and applies it to re-evaluate some important stylized facts in the price stickiness literature. Scraped data, obtained directly from online sources, are a unique source of price information. Scraped prices are easier to collect than CPI and scanner data, and can provide information at daily frequencies for all products sold by hundreds of retailers in many countries around the world. The data can be collected without any delays and the collection methodology can be customized to match the specific needs of the researcher. Furthermore, *online* prices behave similarly to offline prices in terms of timing and size of changes, and price indexes created with scraped data can capture the main inflation patterns in official statistics.

Using this unique dataset, the paper provides three new stylized facts for the sticky-price literature. First, the distribution of the size of price changes tends to be bimodal, with few changes close to zero percent. Second, aggregate hazard functions are hump-shaped, with the conditional probability of a price change increasing for a period between 30 and 90 days.

[32]here I focus the comparison on CPI to emphasize aggregate levels of inflation, while in Section A.3 I provide similar results with a subset of food indexes.

[33]For more on Argentina's inflation, see Cavallo (2010).

Third, there is daily price synchronization among closely competing goods. Although these results do not offer conclusive evidence in favor of any standard sticky-price model, they are mostly consistent with models that combine elements of both TDP and SDP, such as Alvarez et al. (2010).

Still, a great deal of research with scraped data is needed. We need to understand how do these stylized facts change over time, with different levels of inflation, market structures, and other country and sector-level characteristics. At the same time, many other stickiness patterns in the data deserve further attention. For example, in these results the distribution of the size of changes becomes more asymmetric with higher levels of inflation, but the mean size of price increases and decreases stays relatively constant. This is at odds with standard menu-cost models, but is consistent with theories of customer anger or "fair pricing" concerns.[34] More puzzlingly, countries with higher inflation can also have *stickier* prices. Although this observation is based on a small cross-section of four countries, it suggests that inflation is not correlated to the overall level of stickiness, but rather to the relative rigidity of price increases over price decreases.[35] A greater understanding of details like these, both for the frequency and size of price changes, will be key to construct better theoretical models in the future.

More generally, the potential use of scraped data in macroeconomics goes far beyond the one explored in this paper. Scraped prices can be used to create daily price indices that complement official statistics, compare and test theories of international prices, exchange rate and commodity pass-through, study the pricing effect of new product introductions, and provide real-time estimates of sectoral stickiness. These are all topics to explore in future research.

[34] See Rotemberg (2005) and Rotemberg (2008).

[35] See Section 3 in the Appendix for similar results *within* countries.

References

Aalen, Odd, Ornulf Borgan, and Hakon Gjessing, *Survival and Event History Analysis: A Process Point of View.* Statistics for Biology and Health, Springer, 2008.

Alvarez, Fernando E., Francesco Lippi, and Luigi Paciello, "Optimal price setting with observation and menu costs," *National Bureau of Economic Research Working Paper Series*, 2010, *No. 15852.*

Alvarez, Luis J., "What Do Micro Price Data Tell Us on the Validity of the New Keynesian Phillips Curve?," *Economics E-Journal*, June 2008, *2.*

Anderson, Eric and Duncan Simester, "Price Stickiness and Customer Antagonism," *Quarterly Journal of Economics.*

Ball, Laurence and David Romer, "Real Rigidities and the Nonneutrality of Money," *Review of Economic Studies*, 1990, *57*, 183–203.

_ and _ , "Sticky Prices as Coordination Failure," *The American Economic Review*, 1991, *81*, 539–52.

_ , Gregory N. Mankiw, and David Romer, "The New Keynesian Economics and the Output-Inflation Tradeoff," *Brookings Papers on Economic Activity*, 1988, *No 1*, 1–65.

Barro, Robert, "A Theory of Monopolistic Price Adjustment," *Review of Economic Studies*, 1972, *III*, 17–26.

Barros, Rebecca, Marco Bonomo, Carlos Carvalho, and Silvia Matos, "Price Setting in a Variable Macroeconomic Environment: Evidence from Brazilian CPI," *mimeo.*

Baudry, Le Bihan, Sevestre, and Tarrieu, "What do Thirteen Million Price Records have to Say about Consumer Price Rigidity?," *Oxford Bulletin of Economics and Statistics*, 2007, *69* (2), 139–183.

Bertola, Giuseppe and Ricardo Caballero, "Kinked Adjustment Costs and Aggregate Dynamics," *NBER Macroeconomic Annual*, 1990.

Bils, Mark and Peter J. Klenow, "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 2004, *112*, 947–985.

_ , _ , and Benjamin A. Malin, "Reset Price Inflation and the Impact of Monetary Policy Shocks," *NBER Working Papers*, 2009, *14787.*

_ , _ , and Oleksiy Kryvtsov, "Sticky Prices and Monetary Policy Shocks," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2003, *27*, 2–9.

Blanchard, Olivier, "Price Desynchronization and Price Level Inertia," *NBER Working Paper*, 1982, *No. 900.*

_ and Stanley Fischer, *Lectures on Macroeconomics*, MIT Press, 1989.

Blinder, Alan, "Why Are Prices Sticky? Preliminary Results from an Interview Study," *American Economic Review*, May 1991, *81* (2), 89–96.

_ , Eli Canetti, David Lebow, and Jeremy Rudd, "Asking about prices: a new approach to understanding price stickiness," *Russels Sage Foundation, New York.*, 1998.

Boivin, Jean, Marc Giannoni, and Ilian Mihov, "Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data," *National Bureau of Economic Research Working Paper Series*, 2007, *No. 12824.*

Bonomo, Marco, Carlos Carvalho, and Ren Garcia, "State-Dependent Pricing Under Infrequent Information: A Unified Framework," *SSRN eLibrary*, June 2010.

Burstein, Ariel and Christian Hellwig, "Prices and Market Shares in a Menu Cost Model," *NBER Working Paper Series*, September 2007, *No. 13455.*

_ , Martin Eichenbaum, and Sergio Rebelo, "Large Devaluations and the Real Exchange Rate," *Journal of Political Economy*, 2005, *113*, 742–784.

Calvo, Guillermo, "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, 1983, *12*, 383–398.

Campbell, Jeffrey R. and Benjamin Eden, "Rigid prices: evidence from US scanner data," *Federal Reserve Bank of Chicago*, 2005. Published: Federal Reserve Bank of Chicago.

Carlton, Dennis W, "The Rigidity of Prices," *American Economic Review*, September 1986, *76* (4), 637–58.

Carvalho, Carlos, "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks," *Frontiers of Macroeconomics*, 2006, *2(1).*

Cavallo, Alberto, "Argentina : Inflation and Lies," *Unpublished paper*, 2010.

_ and Roberto Rigobon, "The Distribution of the Size of Price Changes," *NBER Working Paper*, 2010, *w16760.*

Cecchetti, Stephen G, "The Frequency of Price Adjustment: A Study of Newsstand Prices of Magazines," *Journal of Econometrics*, 1986, *31*, 255–274.

Cheng, Ming-Yen and Peter Hall, "Calibrating the Excess Mass and Dip Tests of Modality," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 1998, *60* (3), 579–589. ArticleType: primary_article / Full publication date: 1998 / Copyright 1998 Royal Statistical Society.

Cooper, Russell and John Haltiwanger, "Evidence of Macroeconomic Complementarities," *The Review of Economics and Statistics*, 1996, *78, No1*, 78–93.

Dhyne, Emmanuel, Luis Alvarez, Herve Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffman, Nicole Jonker, Patrick Lunnenmann, Fabio Rumler, and Jouko Vilmunen, "Price Setting in the Euro Area: Some Stylized Facts from Individual Price Data," *Working Paper Series, European Central Bank*, 2005, *No. 524.*

Dotsey, Michael, Robert King, and Alexander Wolman, "State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output," *The Quarterly Journal of Economics*, 1999, *114(2).*

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo, "Reference Prices and Nominal Rigidities," *National Bureau of Economic Research Working Paper Series*, March 2008, *No. 13829.*

Ellison, Glenn and Alexander Wolitzky, "A Search Cost Model of Obfuscation," *NBER Working Paper*, 2009, *No. 15237.*

Fischer, Timothy and Jerzy Konieczny, "Synchronization of Price Changes by Multiprocut Firms: Evidence from Canadian Newspaper Prices," *Economic Letters*, 2000, *68*, 271–277.

Fisher, Jerzy D. Konieczny Timothy C. G., "Synchronization of Price Changes by Multiproduct Firms: Evidence from Canadian Newspaper Prices," *Economics Letters*, 2000, *68.*

Gagnon, Etienne, "Price Setting during Low and High Inflation: Evidence from Mexico," *International Finance Discussion Papers, Board of Governors of the Federal Reserve System*, 2007, *896.*

Golosov, Mikhail and Robert Lucas, "Menu Costs and Phillips Curves," *Journal of Political Economy*, 2007, *115(2).*

Goolsbee, Austan and Judith Chevalier, "Measuring Prices and Price Competition Online: Amazon and Barnes and Noble," *National Bureau of Economic Research Working Paper Series*, July 2002, *No. 9085.* published as Goolsbee, Austan and Judith Chevalier. Measuring prices and price competition online: Amazon vs. Barnes and Noble." Quantitative Marketing and Economics I, 2 (June 2003).

Gopinath, Gita and Oleg Itskhoki, "Frequency of Price Adjustment and Pass-trought," *Quarterly Journal of Economics*, 2010.

_ and Roberto Rigobon, "Sticky Borders," *Quarterly Journal of Economics*, 2008, *23(2).*

_ , Oleg Itskhoki, and Roberto Rigobon, "Currency Choice and Exchange Rate Passtrough," *American Economic Review*, 2010.

Guimaraes, Bernardo and Kevin Sheedy, "Sales and Monetary Policy," *CEP Discussion Paper No 887*, 2008.

Hartigan, J. A. and P. M. Hartigan, "The Dip Test of Unimodality," *The Annals of Statistics*, 1985, *13* (1), 70–84.

Henderson, Daniel J., Christopher F. Parmeter, and R. Robert Russell, "Modes, weighted modes, and calibrated modes: evidence of clustering using modality tests," *Journal of Applied Econometrics*, 2008, *23* (5), 607–638.

Horowitz, Joel L., "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica*, September 1999, *67* (5), 1001–1028.

Kackmeister, Alan, "Yesterday's Bad Times Are Today's Good Old Times: Retail Price Changes Are More Frequent Today Than in the 1890s," *Journal of Money, Credit and Banking*, December 2007, *39* (8), 1987–2020.

Kashyap, Anil K, "Sticky Prices: New Evidence from Retail Catalogues.," *Quarterly Journal of Economics*, 1995, *110*, 245–274.

Kehoe, Patrick and Virgiliu Midrigan, "Sales and the Real Effects of Monetary Policy," *Federal Reserve Bank of Minneaplis Working Paper*, 2007, *No. 413.*

_ and _ , "Sticky Prices and Sectoral Real Exchange Rates," *Federal Reserve Bank of Minneaplis Working Paper*, 2007, *No. 656.*

Klenow, Peter J. and Benjamin A. Malin, "Microeconomic Evidence on Price-Setting," *Handbook of Monetary Economics*, 2009, *forthcoming.*

_ and Jonathan L. Willis, "Real rigidities and nominal price changes," *Working Paper*, 2006.

_ and _ , "Sticky information and sticky prices," *Journal of Monetary Economics*, 2007, *54*, 79–99.

_ and Oleksiy Kryvtsov, "State-Dependent or Time-Dependent Pricing: Does it Matter for Recent US Inflation?," *The Quarterly Journal of Economics*, 2008, *73(3)*, 863–903.

Lach, Saul and Daniel Tsiddon, "Staggering and Synzhronization in Price-Setting: Evidence from Multiproduct Firms," *American Economic Review*, 1996, *86(5)*, 1175–1196.

_ and _ , "Small price changes and menu costs," *Managerial and Decision Economics*, 2007, *28* (7), 649–656.

Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica*, July 1979, *47* (4), 939–56.

Lavn, Felipe Morand and Mauricio Tejada, "Price Stickiness in Emerging Economies: Empirical Evidence for Four Latin-American Countries," Working Papers wp286, University of Chile, Department of Economics May 2008.

Levy, Daniel, Mark Bergen, Shantanu Dutta, and Robert Venable, "The Magnitude of Menu Costs: Direct Evidence from Large U.S. Supermarket Chains," *The Quarterly Journal of Economics*, August 1997, *112* (3), 791–825.

L'Huillier, Jean Paul, "Heterogeneous Information and Nominal Rigidity," *Job Market Paper, MIT*, 2009.

Mackowiak, Bartosz and Frank Smets, "On Implications of Micro Price Data for Macro Models," *Conference Series, Federal Reserve Bank of Boston*, 2008.

Mankiw, N. Gregory, "Small Menu Costs and Large Business Cycles: A Macroeconomic Model," *Quarterly Journal of Economics*, 1985, *100 (2)*, 529–38.

Midrigan, Virgiliu, "Is Firm Pricing State or Time-Dependent? Evidence from US Manufacturing," *Review of Economic and Statistics*, 2005.

_ , "Menu Costs, Multi-Product Firms and Aggregate Fluctuations," *Working Paper*, 2005.

_ , "International Price Dispersion in State-Dependent Pricing Models," *Journal of Monetary Economics*, 2007, *54 (8)*, 2231–2250.

Nakamura, Emi, "Pass-Through in Retail and Wholesale," *American Economic Review*, 2008, *98*, forthcoming.

_ **and Jon Steinsson**, "Five Facts About Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 2008, *123(4)*, 1415–1464.

Neiman, Brent, "Multinationals, Intrafirm Trades and International Macro Dynamics," *Job Market Paper - Harvard University*, 2008.

Okun, Arthur M., *Prices and Quantity: A Macroeconomic Analysis*, Brookings Institution Press, 1981.

Papke, Leslie and Jeffrey Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 1996, *11(6)*, 619–632.

Rauch, James, "Networks Versus Markets in International Trade.," *Journal of International Economics*, 1999, *48(1)*, 7–35.

Rotemberg, Julio, "Monopolistic Price Adjustment and Aggregate Output," *Review of Economic Studies*, October 1982, *49* (4), 517–31.

_ , "Sticky Prices in the United States," *The Journal of Political Economy*, 1982, *90*, 1187–1211.

_ , "The New Keynesian Microfoundations," *Macroeconomics Annual*, 1987, *2*, 69–104.

_ , "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy," *Journal of Monetary Economics*, May 2005, *52*, 829–852.

— , "Behavioural Aspects of Price Setting, and their Policy Implications," *HBS Working Paper*, 2007.

— , "Fair Pricing," *HBS Working Paper*, 2008.

— , "Altruistic Dynamic Pricing with Curstomer Regret," *HBS Working Paper*, 2009.

Sheshinski, Eytan and Yoram Weiss, "Inflation and Costs of Price Adjustment.," *Review of Economic Studies*, June 1977, *44*, 287.

— **and** — , "Staggered and Synchronized Price Policies Under Inflation: The Multiproduct Monopoly Case," *Review of Economic Studies*, 1992, *59*, 331–359.

Silverman, B. W., "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1981, *43* (1), 97–99. ArticleType: primary_article / Full publication date: 1981 / Copyright 1981 Royal Statistical Society.

Sorensen, Alan T., "An Empirical Model of Heterogeneous Consumer Search for Retail Prescription Drugs," *National Bureau of Economic Research Working Paper Series*, October 2001, *No. 8548.*

Taylor, John, "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, 1980, *88*, 1–23.

Tommasi, Mariano, "Inflation and Relative Prices: Evidence from Argentina," *UCLA Working Paper Series*, 1992, (661).

Wang, J.L., "Smoothing Hazard Rates," *Encyclopedia of Biostatistics*, 2005, *2nd Edition, Vol 7*, 4986–4997.

Woodford, Michael, "Information-constrained State-dependent Pricing," *Journal of Monetary Economics*, 2009, *56*, S100 – S124.

Wulfsberg, Fredik, "Price Adjustments and Inflation - Evidence from Consumer Price Data in Norway 1975-2004," *Norges Bank Working Paper*, 2008.

Tables

Table 1: Alternative Data Sources

	Scraped Data	CPI Data	Scanner Data
Data Frequency	Daily	Monthly - Bi-Monthly	Weekly
Countries Available for Research	~50*	10-15	<5
All Products in Retailer (Census)	Yes	No	No
Comparable data across countries	Yes	Limited	Limited
Details: sale, price control, other	Yes	Limited	Yes
No *Forced* Substitutions	Yes	No	Yes
Real-Time data availability	Yes	No	No
Product Categories Covered	Few	Many	Few
Retailers Covered	Few	Many	Few
Quantities Sold	No	No	Yes

Notes: *Data from over 70 countries are currently being collected by the Billion Prices Project (bpp.mit.edu). ** Only goods purchased are scanned.

Table 2: Database Description

	Argentina	Brazil	Chile	Colombia
Total observations	10.8M	9.8M	9.7M	3.9M
Total Products	28813	23115	24336	9526
Initial date	10/7/2007	10/10/2007	10/24/2007	11/13/2007
Final date	08/13/2010	08/13/2010	08/13/2010	08/13/2010
Days	1041	1038	1024	1004
Categories	74	72	72	59
Urls	993	319	292	122
Product Description	Yes	Yes	Yes	Yes
Sale indicator	Yes	Yes	-	Yes
Price Controls	Yes	-	-	-
Brand, Size, Bulk Price	Yes	Yes	Yes	Yes
Missing obs. within spells	32%	25%	33%	22%
Obs with sales	2.99%	4.38%	-	7.55%
Products with sales	39%	22%	-	25%
Products with price controls	1.5%	-	-	-
Life of goods (in days, Mean/Median)	549/540	558/502	590/634	523/525
Obs per good (Mean/Median)	375/304	423/376	398/380	410/349

Notes: The missing values are caused by items that go out of stock or failures in the scraping software that tend to last for only a few days. Gor the analysis in this paper, I replaced missing values *within* price series with the previous price available for that particular product. For those results that exclude sales, I created a regular price series by replacing all sale prices with the previous non-sale price available for that product. I also removed all price changes exceeding 500%. These represent a negligible number of observations but can bias statistics related to the magnitude of price changes. See Section A.1 in the Appendix for more details on data treatments.

Table 3: Median Frequencies by Country - Increases and Decreases

	Including Sales				Excluding Sales		
	Arg	Brazil	Chile	Col	Arg	Brazil	Col
Daily Frequency	0.015	0.027	0.015	0.019	0.012	0.023	0.016
Implied Durations (days)	66	36	66	52	83	43	62
Implied Durations (months)	2.2	1.2	2.2	1.73	2.8	1.4	2.1
Frequency of Increases (Freq+)	0.010	0.016	0.008	0.011	0.009	0.014	0.009
Frequency of Decreases (Freq-)	0.004	0.0011	0.007	0.008	0.003	0.009	0.007
Freq+/Freq-	2.5	1.5	1.1	1.4	3	1.6	1.3

Notes: *Bils and Klenow (2004) methodology taking the mean within categories and then the median across categories.

Table 4: Price Changes by Country and Sale Treatment

	Including Sales				Excluding Sales*		
	Arg.	Brazil	Chile	Col.	Arg.	Brazil	Col.
Price Changes	244K	366K	221K	108K	195K	289K	
Products with no price changes	19%	10%	21%	14%	22%	10%	21%
Price changes per good (Mean/Median)	8/5	15/8	9/4	11/7	2.6/2	13/7	4/3
Price increases (% of price changes)	68%	57%	54%	56%	84%	59%	57%
Price decreases (% of price changes)	32%	43%	46%	43%	16%	41%	43%
Inflation (%, average annual rate)	17.1%	5.1%	2.7%	4.2%			

Notes: *No sales information is available for Chile.

Table 5: Size of Price Changes by Country and Sale Treatment

	Including Sales				Excluding Sales		
	Arg.	Brazil	Chile	Col.	Arg.	Brazil	Col.
Size of changes (Mean*)	5.0%	1.7%	3.6%	2.0%	5.5%	1.5%	1.8%
Size of price increases (Mean*)	13%	11.7%	17.6%	11.8%	11.8%	9.3%	8.3%
Size of price decreases (Mean*)	-11%	-11.9%	-14.2%	-10.5%	-10.2%	-10.1%	-7.4%
Share of price changes under |1%|	4.3%	%	3.6 %	%	%	%	%
Share of price changes under |5%|	27%	33%	25%	37%	%	%	%

Notes: *Mean size of changes per individual good, then the mean per category, and finally the mean across all categories.

Table 6: Statistical Tests of Modality

	DIP Test (calibrated)		Silverman's Test			
	Dip Stat.	Null 1 mode p-value	Critical Band.	Null 1 mode p-value	Null 2 modes p-value	Null 3 modes p-value
Argentina	0.07	0.00	1.92	0.00	0.00	0.06
Brazil	0.02	0.00	1.12	0.00	0.00	0.07
Chile	0.03	0.00	1.74	0.00	0.00	0.18
Colombia	0.01	0.00	0.66	0.03	0.97	0.99

Table 7: Mean Synchronization within Urls

	Argentina	Brazil	Chile	Colombia
All Price Changes	0.008	0.007	0.007	0.012
Excluding Sales	0.008	0.006	-	0.011
Price Increases	0.007	0.006	0.002	0.006
Price Decreases	0.003	0.005	0.002	0.004

Notes: Results based on 824 urls in Argentina, 281 in Brazil, 256 in Chile and 103 in Colombia.

Table 8: Online vs. Offline Prices

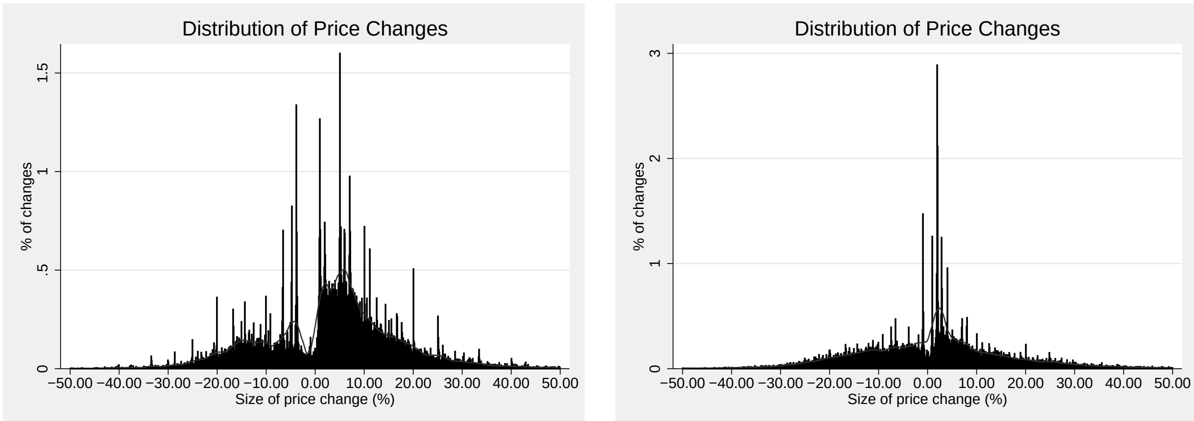
	Argentina	Brazil	Chile	Colombia
Matching ids	Yes	Yes	No	Yes
% Available Online	100%	80%	90%	74%
PRICE LEVELS				
online=offline	18%	42%	93%	29%
online>offline	78%	34%	4%	32%
Price Difference (Mean %)	5	9	2	0
PRICE CHANGES				
Products with Identical Change Series*	93%	75%	94%	67%
Ratio of Changes over Observations				
Offline	0.215	0.356	0.274	0.433
Online	0.215	0.411	0.249	0.433
Mean Size of Changes (%)				
Offline	1.6	4.9	1.4	8.1
Online	1.4	5.3	1.3	8.3

Notes: *Indicator variable conditional on change: 1 if the price increased, -1 if it decreased.

Table 9: Average Annual Inflation (% per year)

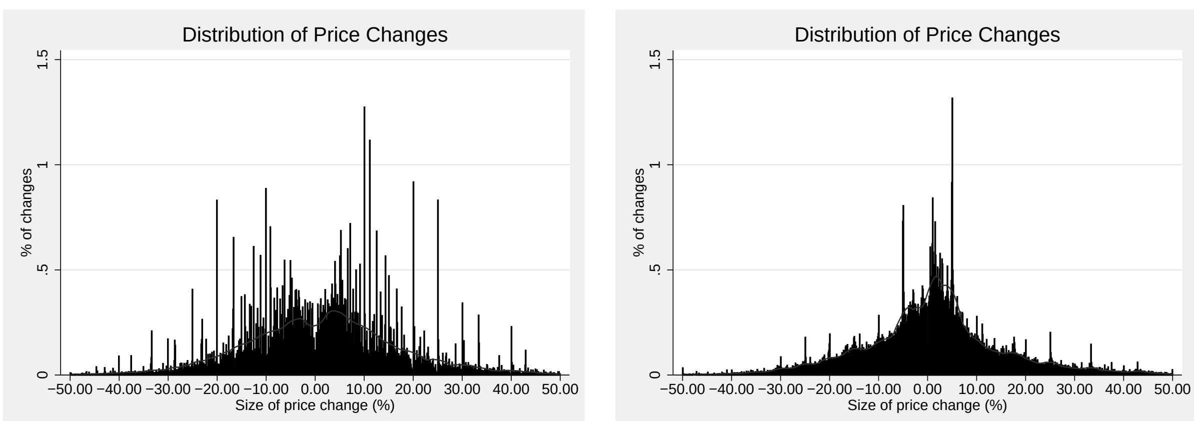
	Scraped Supermarket Index	Official Consumer Prices (CPI)
Argentina	17.1	7.6
Brazil	5.1	4.6
Chile	2.7	3.7
Colombia	4.2	6.1

Figures



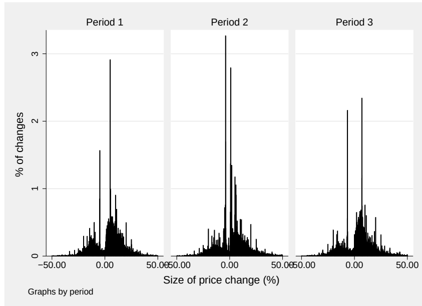
(a) Argentina (b) Brazil

Figure 1: Distribution of the Size of Price Changes
Notes: Bin size is 0.1%. Smoothed kernel density shown.

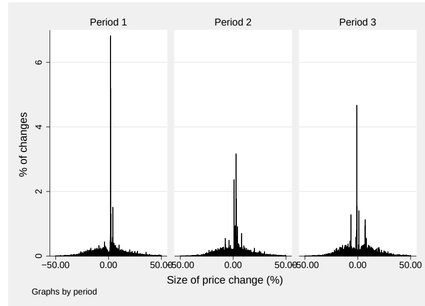


(a) Chile (b) Colombia

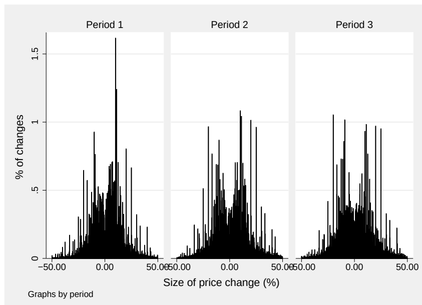
Figure 2: Distribution of the Size of Price Changes
Notes: Bin size is 0.1%. Smoothed kernel density shown.



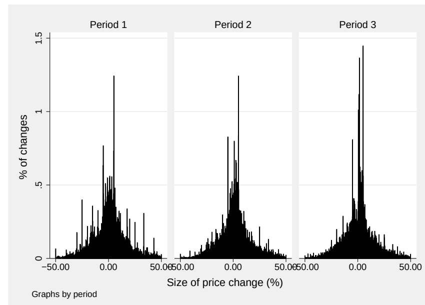
(a) Argentina



(b) Brazil

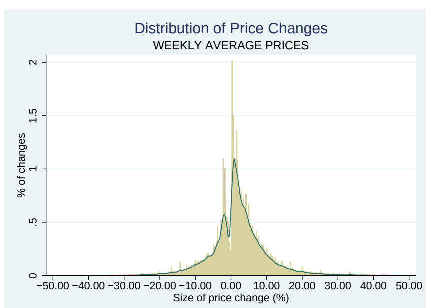


(c) Chile

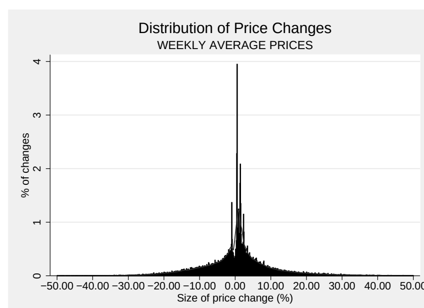


(d) Colombia

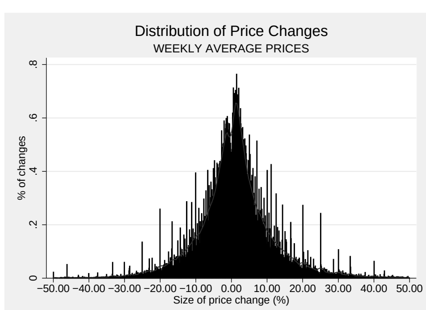
Figure 3: Distribution of the Size of Price Changes - Over Time
Notes: Bin size is 0.1%. Smoothed kernel density shown.



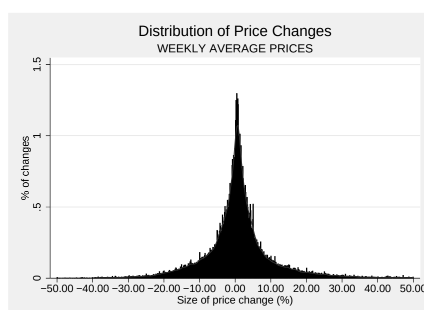
(a) Argentina



(b) Brazil

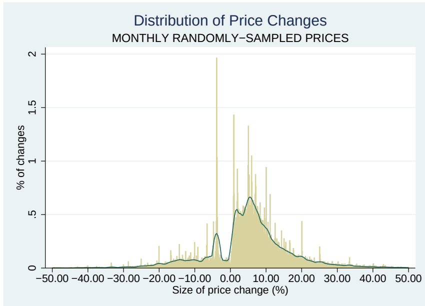


(c) Chile

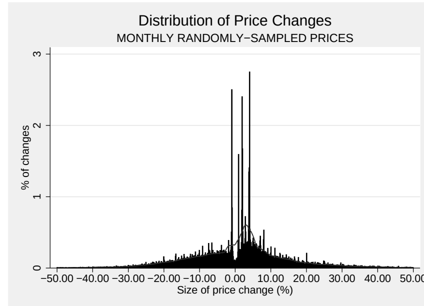


(d) Colombia

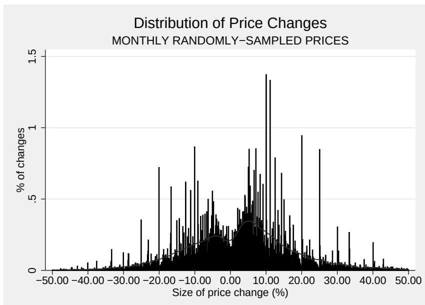
Figure 4: Distribution of the Size of Price Changes - Weekly Averages
Notes: Bin size is 0.1%. Smoothed kernel density shown.



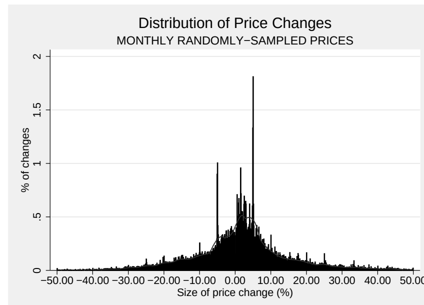
(a) Argentina



(b) Brazil

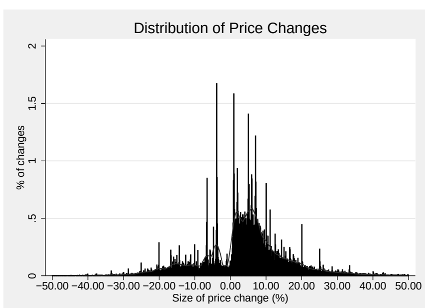


(c) Chile

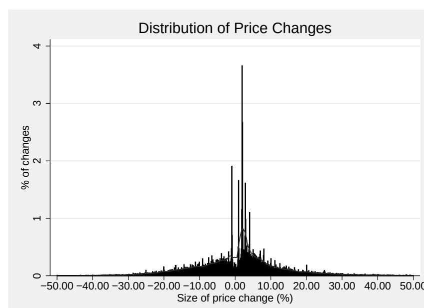


(d) Colombia

Figure 5: Distribution of the Size of Price Changes - Monthly Sampling
Notes: Bin size is 0.1%. Smoothed kernel density shown.

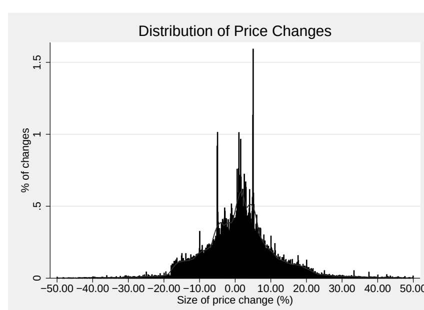


(a) Argentina



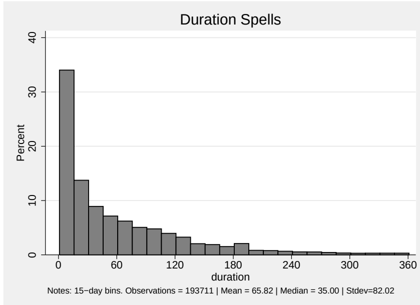
(b) Brazil

(c) Chile - Not Available

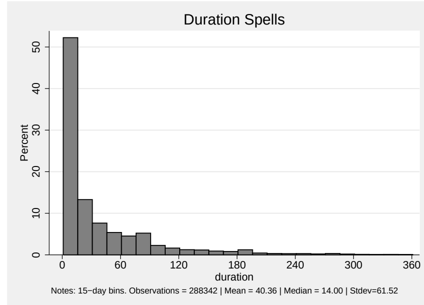


(d) Colombia

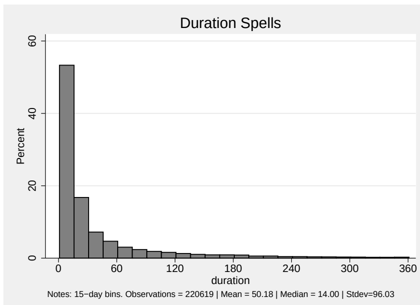
Figure 6: Magnitude of Price Changes - Excluding Sales
Notes: Bin size is 0.1%. Estimated kernel density shown. Brazil shown without changes on 15/12/07 and 29/12/07.



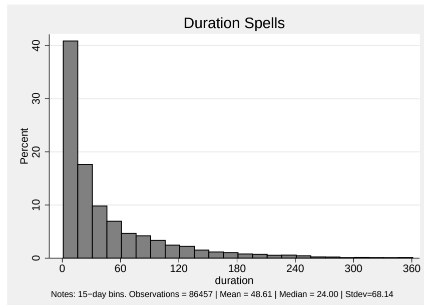
(a) Argentina



(b) Brazil

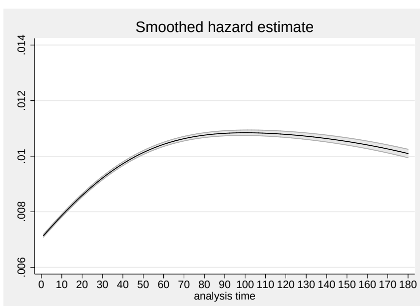


(c) Chile

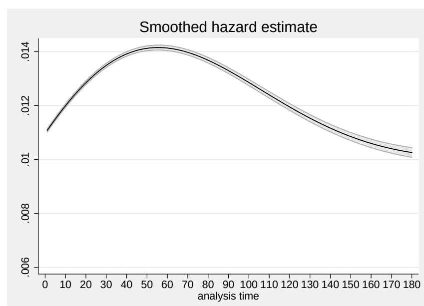


(d) Colombia

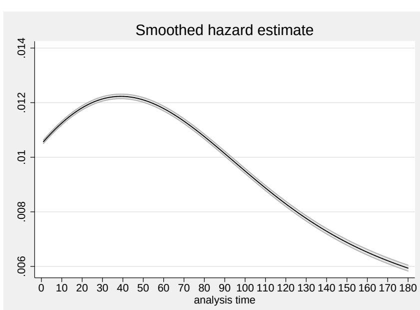
Figure 7: Histogram of Duration Spells



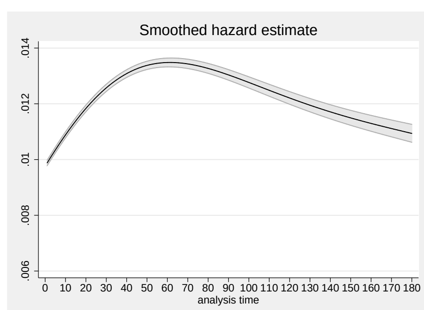
(a) Argentina



(b) Brazil



(c) Chile



(d) Colombia

Figure 8: Smoothed Hazard Functions

Notes: Left-censored spells are excluded. Sales are excluded in Argentina, Brazil, and Colombia. Initial 180 days shown.

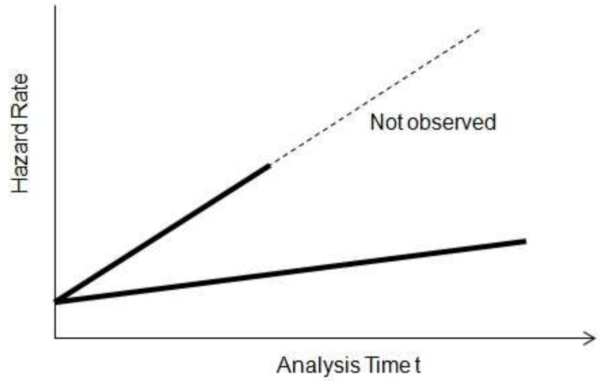
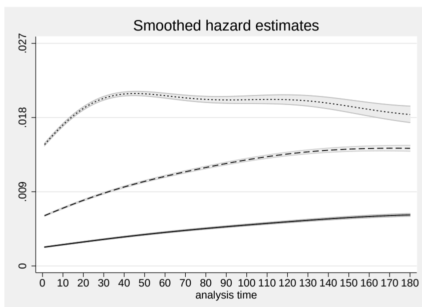
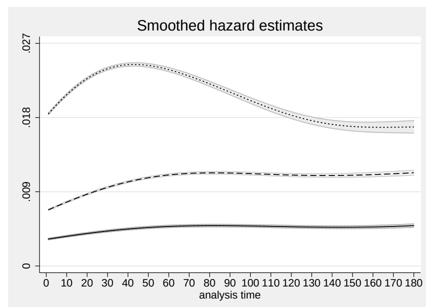


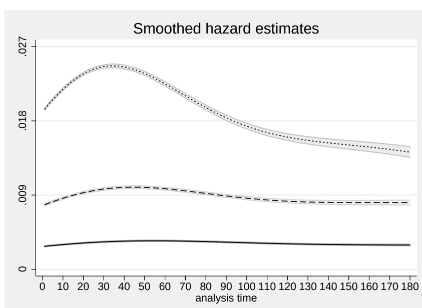
Figure 9: Heterogeneity and Survival Bias



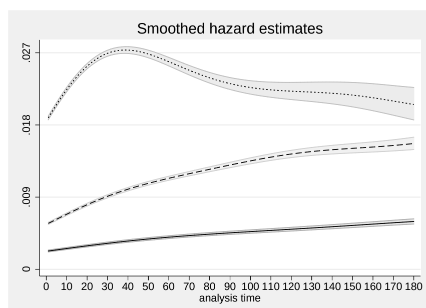
(a) Argentina



(b) Brazil



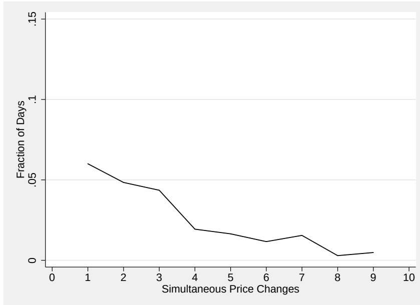
(c) Chile



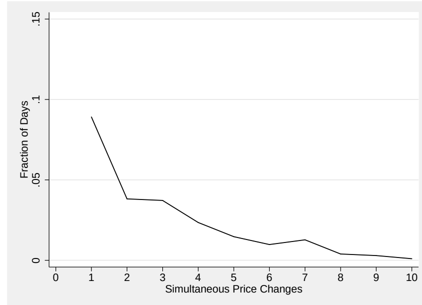
(d) Colombia

Figure 10: Hazards for Different Duration Groups

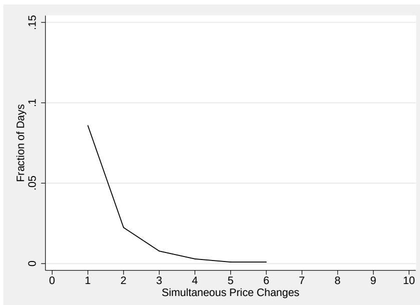
Notes: Left-censored spells are excluded. Sales are excluded in Argentina, Brazil, and Colombia. Initial 180 days shown.



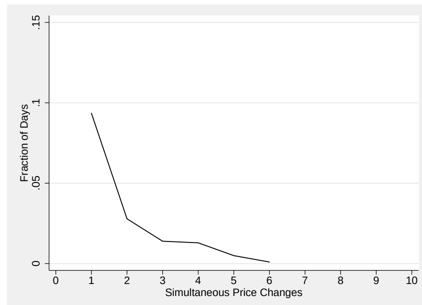
(a) Argentina



(b) Brazil

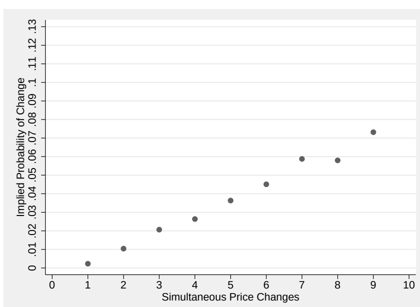


(c) Chile

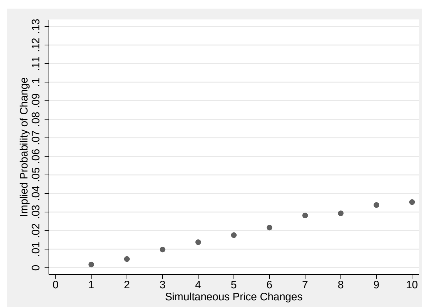


(d) Colombia

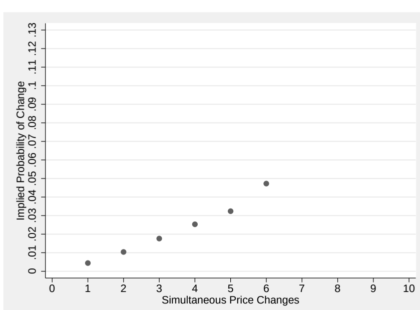
Figure 11: Distribution of Synchronized Changes - Example with Bottled Water Urls



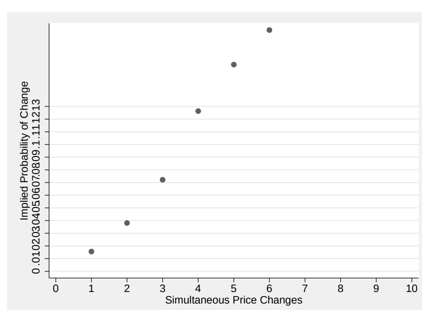
(a) Argentina



(b) Brazil

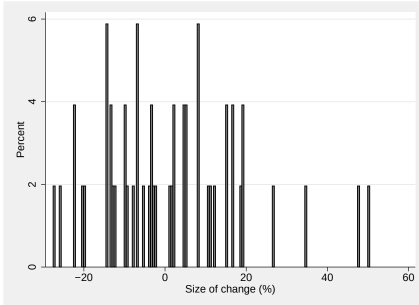


(c) Chile

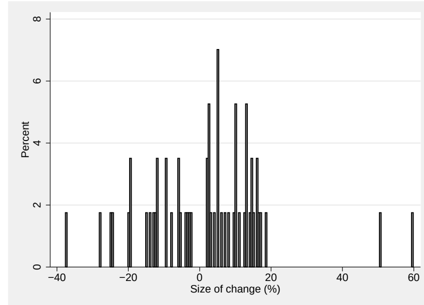


(d) Colombia

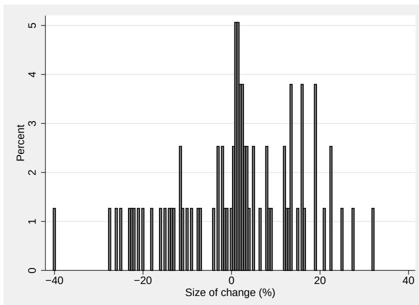
Figure 12: Implied Probabilities - Example with Bottled Water Urls



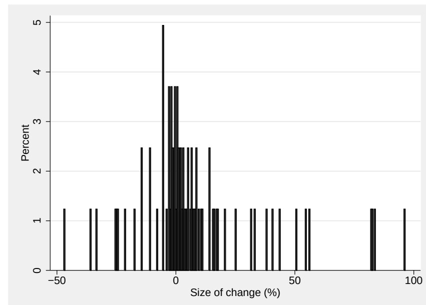
(a) Argentina



(b) Brazil

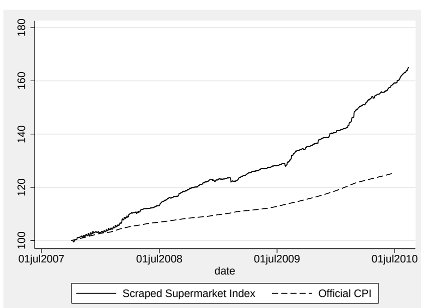


(c) Chile

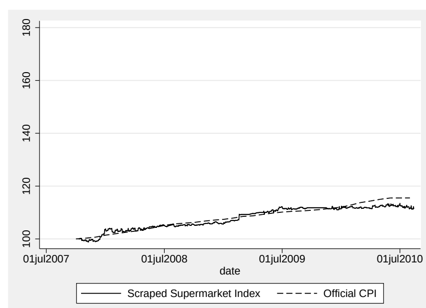


(d) Colombia

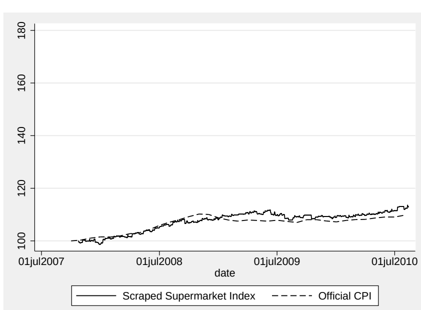
Figure 13: Magnitude of Offline Price Changes
Notes: Bin size is 0.5%



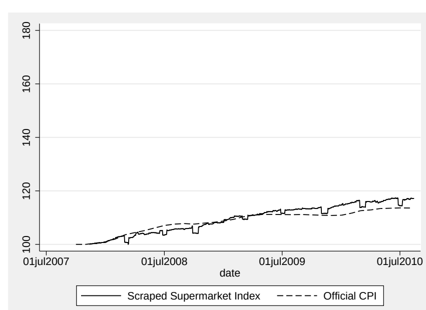
(a) Argentina



(b) Brazil

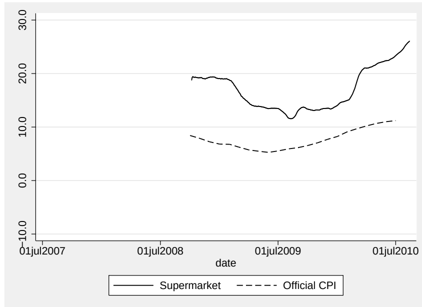


(c) Chile

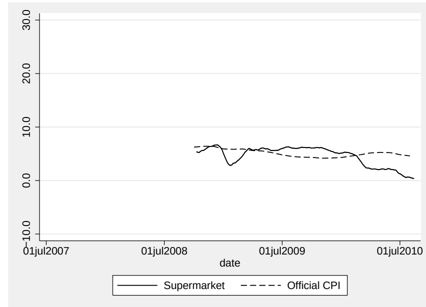


(d) Colombia

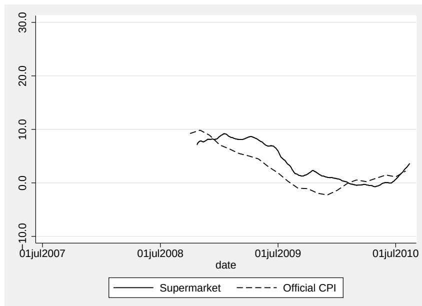
Figure 14: Scraped Index vs. the Official CPI



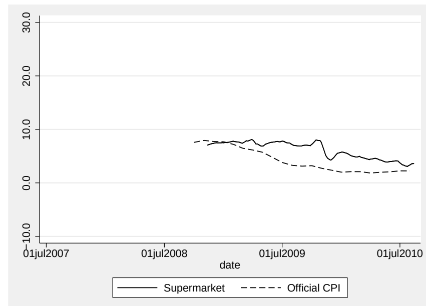
(a) Argentina



(b) Brazil



(c) Chile



(d) Colombia

Figure 15: Annual Inflation: Scraped vs CPI