Scraped Data and Sticky Prices

Alberto Cavallo[*‡]

MIT & NBER

This version: August 23, 2016

Abstract

I use daily prices collected from online retailers in five countries to study the impact of measurement bias on three common price stickiness statistics. Relative to previous results, I find that online prices have longer durations, with fewer price changes close to zero, and hazard functions that initially increase over time. I show that time-averaging and imputed prices in scanner and CPI data can fully explain the differences with the literature. I then report summary statistics for the duration and size of price changes using scraped data collected from 181 retailers in 31 countries.

JEL Classifications: E30, E60

1   Introduction

Sticky prices are a fundamental element of the monetary transmission mechanism in many macroeconomic models. Over the past decade, a large empirical literature has tried to measure stickiness and understand its micro-foundations.[1] These studies produced a set of stylized facts, summarized by Klenow and Malin (2010), which have been used to motivate many theoretical papers.[2]

The increase in empirical work has been possible due to unprecedented access to micro-level Consumer Price Index (CPI) data and scanner datasets in several countries. While valuable, these datasets are not collected for research purposes and their sampling characteristics can introduce measurement errors and biases that affect some of the stylized facts in the literature.[3]

In this paper, I use a new type of micro-level data based on online prices called "Scraped Data" to explicitly document the impact of measurement biases on some key stickiness statistics. In particular, I argue that two main sampling characteristics, time-averaging and imputations of missing prices, can greatly affect the observed duration and size of price changes in traditional data sources.

Time averages are intrinsic in scanner datasets, such as Nielsen's "Retail Scanner Data", which reports weekly averages of individual product prices. Imputed prices for substitutions and temporarily missing products are a common characteristic of CPI datasets. In the U.S., the Bureau of Labor Statistics (BLS) imputes many of these missing prices with a method called "cell-relative imputation" that uses the average price change within related categories of goods.[4] These two sampling characteristics, while reasonable for the purposes of the original data collection efforts, can greatly increase the number of price changes observed in the data, reduce the perceived size of these changes, and affect key statistics such as the distribution of the size and the hazard rate of price changes.

Scraped data are not affected by these sources of measurement bias. Online prices are collected using specialized software that scans the websites of retailers that show prices online, finds relevant information, and stores it in a database.[5] Once it is set up, the software can

---

[1] Cecchetti (1986), Kashyap (1995), and Lach and Tsiddon (1996) provided pioneering contributions to the literature using samples of goods such as magazines and groceries. Bils and Klenow (2004) made a seminal contribution with micro-level U.S. CPI data. They were followed by papers such as Nakamura and Steinsson (2008); Klenow and Kryvtsov (2008); Klenow and Willis (2007); Dhyne et al. (2006); Boivin et al. (2009); Wulfsberg (2009); and Gagnon (2009), just to name a few. For a recent survey of the literature, see Nakamura and Steinsson (2013).

[2] Some examples are Midrigan (2011); Gorodnichenko (2008); Woodford (2009); Bonomo et al. (2011); Costain and Nakov (2011); and Alvarez and Lippi (2014).

[3] For previous discussions of measurement error in the literature, see Campbell and Eden (2014); Cavallo and Rigobon (2011); and Eichenbaum et al. (2014).

[4] See Bureau of Labor Statistics (2015a), page 20. Before January 2015, the BLS imputed prices using relatively broad item strata and geographic index areas. The latest methodology uses narrower elementary level items (ELIs) and metropolitan areas. This change is explained in Bureau of Labor Statistics (2015b). My results suggest that this is likely to reduce the magnitude of the imputation bias in the U.S. CPI data in the future.

[5] The primary language used to write content on the Web is called Hyper Text Markup Language, or HTML. It is written using tags, such as <div class=price>, which provide instructions for the browser. These tags can be used by the scraping software to locate relevant price and product information on the

run automatically every day, providing high-frequency information for all goods sold by the sampled retailers in a set of selected countries. The scraped dataset used in this paper was collected by the Billion Prices Project at MIT every day between October 2007 and August 2010 for over 250 thousand individual products in five countries: Argentina, Brazil, Chile, Colombia, and the United States. I also used a larger dataset collected by PriceStats, a private company, to report summary statistics on price stickiness from 181 retailers in 31 countries.[6]

The first contribution of this paper is the use of online data in five countries to document the impact of measurement bias on three common statistics in the literature: the duration of price changes, the distribution of the size of price changes, and the shape of their hazard function over time.

To show the impact of time-averages in scanner data, I directly compared my findings to those using data provided by Nielsen for the same retailer, location, and time period. I also simulated the weekly time-averaging characteristic in my data, which provided a near perfect match to the scanner data results. As Campbell and Eden (2014) argued, weekly averaging makes a single price change look like two consecutive smaller changes. This creates more frequent and smaller price changes, completely altering the shape of their size distribution. Furthermore, it causes the hazard rate to appear highest on the first week after a change, producing fully downward-sloping hazard functions over time. Overall, time-averaging the data produces similar results to those in papers that use scanner data such as Eichenbaum et al. (2011) and Midrigan (2011).

To determine the effects of imputations in CPI data, I simulated the cell-relative imputation for temporarily missing prices in my online data. I show that imputing missing prices with average changes in the same category also increases the frequency of price changes and greatly reduces their size, making the size distribution completely unimodal, as in Klenow and Kryvtsov (2008). This effect is separate from the one caused by forced item substitutions, previously discussed in the literature.[7] The bias is strongest when broader categories of goods are used as the reference for imputation, as BLS did until January 2015. I also show that daily prices are needed to detect the initial increase in hazard rates during the first few months. Instead, if cell-relative imputation is applied to monthly data, the hazard function resembles those in papers with CPI data, such as Nakamura and Steinsson (2008).

The second contribution of the paper is the use of Scraped Data to compute a set of summary statistics for durations and sizes of price changes in 31 countries. These results can be used to study the robustness of stylized facts across countries, parametrize models, and make cross-country comparisons. In particular, I show that prices are stickier than comparable results reported by various papers from 15 countries summarized by Klenow and Malin (2010), and that the share of small price changes is low in most countries. In the Appendix, I show that inflation is correlated with the relative frequency and size of price increases over decreases in the cross-section of countries. The fact that the same type of

data is used in every country ensures that these findings are not driven by differences in sampling characteristics or the way the statistics are computed, which complicated previous cross-country comparisons in the literature.

My findings have several applications and implications. First, they show that some stylized facts in the literature, such as the prevalence of very small price changes, are driven by the sampling characteristics in traditional data sources. Documenting and adjusting to these biases are critical for papers that rely on these statistics to study the real effects of monetary policy, as done in Alvarez et al. (2016). Second, I provide statistics on the frequency and size of price changes that are free from time averages or cell-relative imputations and can be used to evaluate or parametrize alternative models in the literature, such as those in Woodford (2009) and Alvarez et al. (2011). Third, these findings illustrate how new data collection techniques allow macro-economists to build customized datasets, designed to minimize measurement biases and address specific research needs. As pointed out by Einav and Levin (2014), the emergence of "Big Data" requires economists to develop new data management capabilities, and new data collection skills may become an essential part of that process. The role of online data in this context is discussed in detail in Cavallo and Rigobon (2016).

My paper directly relates to others in the price stickiness literature that discuss potential sources of measurement bias. Campbell and Eden (2014) identified prices which could not be expressed in whole cents in an Nielsen scanner dataset, noting that technical errors and time aggregation likely caused them. My results with scanner data confirm their argument. Eichenbaum et al. (2014) use CPI and scanner data from multiple stores to show how unit-value prices, reported as the ratio of sales revenue of a product to the quantity sold, affect the prevalence of small price changes. While they also use daily data, their focus is on the effects of unit-values and averaged prices across stores. Instead, I compare the weekly-averaged prices to show that even scanner datasets not affected by unit-values, such as Nielsen's Retail Scanner Data, can still produce biased results for frequency, size, and hazard rates of price changes. Eichenbaum et al. (2014) also study the effect of unit-values and bundled goods in CPI categories such as "Electricity" and "Cellular Phone Services". Instead, I focus on price imputations for missing prices, which affect nearly all CPI categories.

My work is also related to papers that use online prices, such as Brynjolfsson et al. (2003); Ellison and Ellison (2009); Lunnemann and Wintr (2011); Gorodnichenko et al. (2014); Ellison et al. (2015); and Gorodnichenko and Talavera (2016). These papers find that online prices tend to be more flexible and have smaller price changes than offline prices. The difference with my results likely comes from the fact that they focus on retailers that participate in price-comparison websites. As Ellison and Ellison (2009) showed, this type of retailer faces a different competitive environment that tends to increase the frequency and reduce the size of their price changes. Instead, I use data from large multi-channel retailers that have an online presence, but sell mostly offline. Lunnemann and Wintr (2011) note that multi-channel retailers represent only 9% of all price quotes in their sample.

The paper is organized as follows. Section 2, describes the collection methodology and characteristics of Scraped Data. Section 3, uses the daily data from 5 countries to document the impact of measurement error by comparing the duration of prices, the distribution of the size of price changes, and the hazard functions with previous results in the literature, sampling simulations, and a comparable scanner dataset. Section 4 provides further robustness checks for exclusion of sales, using data from different retailers and sectors, and the effects of the

sampling interval. Section 5 discusses implications for the literature and uses Scraped Data from 31 countries to document the duration and size of price changes. Section 6 concludes.

2 Description of Scraped Data

2.1 The Data Collection Methodology

A large and growing share of retail prices are 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 who want to study price dynamics, yet it has been largely untapped because the information is widely dispersed across thousands of webpages and retailers. Furthermore, there is no historical record of these prices, so they must be continually collected over time.

The technology to periodically record online prices on a large scale is now more widely available. Using a combination of web programming languages, I built an automated procedure that scans the code of publicly available webpages every day, identifies relevant pieces of information, and stores the data. This technique is commonly called "web scraping", so I use the term *Scraped Data* to describe the information collected for this paper.

The scraping methodology has three steps. First, at a fixed time each day, a software program downloads a selected list of public webpages where product and price information are shown. These pages are individually retrieved using the same web address (URL) every day. Second, the underlying code is analyzed to locate each piece of relevant information. This is done by using special characters in the code that identify the start and end of each variable, which have been placed by the page programmers to give the website a particular look and feel. For example, prices may be shown with a dollar sign in front of them and enclosed within <price> and </price> tags. Third, the software stores the scraped information in a database that contains one record per product per day. These variables include the product's price, the date, category information, and sometimes an indicator for whether the item was on sale or not.

2.2 Advantages and Disadvantages

The main 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.

Table 1: Alternative Data Sources

	Scraped Data	CPI Data	Scanner Data
Data Frequency	Daily	Monthly - Bi-Monthly	Weekly
All Products in Retailer (Census)	Yes	No	No
Product Details (size, brand, sale)	Yes	Limited	Yes
Uncensored Price Spells	Yes	No	Yes
Countries Available for Research	~60	~20	<5
Comparable Data Across Countries	Yes	Limited	Limited
Real-Time Availability	Yes	No	No
Product Categories Covered	Few	Many	Few
Retailers Covered	Few	Many	Few
Quantities Sold	No	No	Yes

Notes: The Billion Prices Project (bpp.mit.edu) datasets contain information from over 60 countries with varying degrees of sector coverage. Nielsen U.S. scanner datasets are available for research application through the Kilts Center for Marketing at the University of Chicago. Klenow and Malin (2010) provide stickiness results with CPI data sourced from 27 papers in 23 countries.

Scraped Data have some important advantages. First, these datasets contain posted *daily* prices that are free from unit values, time-averaging, and imputations that can greatly affect some stickiness statistics, as is shown later in this paper. The daily data are also useful to better identify sales and other price changes that might be missed with monthly data. Second, detailed information can be obtained for *all* products sold by the sampled retailers, instead of a few (as in CPI data) or selected categories (as in scanner data). Third, there are no censored or imputed price spells in Scraped Data. Prices are recorded from the first day they are offered to consumers until the day they are discontinued from the store. In CPI, by contrast, there are frequent imputations and forced substitutions when the agent surveying prices cannot find the item. Fourth, Scraped Data can be collected remotely in any country where price information can be found online. In particular, in this paper I use data for four *developing* countries, where scanner data are scarce and product-level CPI prices are seldom disclosed.[8] Fifth, scraped datasets are comparable across countries, with prices that can be collected for the same categories of goods and time period using identical techniques. This makes it easier to perform simultaneous cross-country analyses, as I discuss in Section 5.[9] Finally, Scraped Data are available in real time, without any delays in accessing and processing the information. Eventually, this could be used by central banks to obtain

---

[8]The study of stickiness in developing countries is rare in the literature. An exception is Gagnon (2009), who provides a detailed analysis of sticky prices in Mexico using disaggregated CPI data manually digitalized from printed books. Another is Alvarez et al. (2015), who use CPI data from Argentina to document the behavior of price stickiness from the hyperinflation in the late 80s to the period of low inflation in the 90s.

[9]Past cross-country comparisons in the literature have had to rely on results provided by different papers, often with different data sources, time periods, event methods, and data treatments. See, for example, Klenow and Malin (2010). An exception is Dhyne et al. (2006), who were able to use similar data from multiple countries thanks to the coordination provided by the European Inflation Persistence Network at the European Central Bank. They describe, however, how each national statistical agency was unwilling to share the micro data, even with Eurostat, so the frequency analysis had to be conducted independently in each country

real-time estimates of stickiness and related statistics.

Table 1 also shows the main disadvantages of scraped prices. First, they typically cover a much smaller set of product categories than CPI prices. In particular, the prices used in the paper cover only between 40-70% of all CPI expenditure weights in these countries. While this is enough to demonstrate the effect of measurement errors on pricing statistics, the quantitative findings on stickiness and size of changes shown here should not be viewed as representative of services and other sectors that cannot yet be covered with online data. Second, the data comes only from large multi-channel retailers that sell both online and offline. Currently the vast majority of retail sales take place in this type of retailer, but in principle this may represent a form of sampling bias compared to the CPI (though not due to the "online" nature of the data, as I show below). Finally, a major disadvantage of Scraped Data relative to scanner datasets is the lack of information on quantities sold. In measuring stickiness, quantities are useful in obtaining detailed expenditure weights for narrowly defined categories, so in this paper I use CPI category weights when needed.

2.3 Eight Large Retailers in Five Countries

The main dataset used in this paper has more than 60 million daily prices in five countries: Argentina, Brazil, Chile, Colombia, and the United States. It is available for download at bpp.mit.edu. Table 2 provides details on each country's database. The data come from the websites of eight different companies, with prices collected on a daily basis between 2007 and 2010.

Table 2: Database Description

	USA	Argentina	Brazil	Chile	Colombia
Retailers	4	1	1	1	1
Observations (Millions)	28M	11M	10M	10M	4M
Products (Thousands)	172K	28K	22K	24K	9K
Days	865	1041	1026	1024	992
Initial Date	03/08	10/07	10/07	10/07	11/07
Final Date	08/10	08/10	08/10	08/10	08/10
Categories	49	74	72	72	59
URLs	16188	993	322	292	123
Daily Missing Obs. (%)	37	32	26	33	22

Notes: The number of observations in the second row does not include missing values within a price series. The data contains missing values caused by items that go out of stock or by failures in the scraping software that tend to last for a few days. Following Nakamura & Steinsson (2008), missing prices are replaced for the first five months of the price gap with the previous price available for each product. I also ignore all price changes exceeding +200% and -90%, which represent a negligible number but can significantly bias statistics related to the magnitude of price changes. See the Appendix for more details on data treatments.

by a different team, and each dataset had different characteristics. For example, some countries allowed researchers to categorize price changes due to product replacement, sale prices and other discounts, while others did not.

For the United States, I have data from four of the largest retailers in the country: a supermarket, a hypermarket/department store, a drugstore/pharmacy retailer, and a retailer that sells mostly electronics.[10] In the other countries, I have data from a large supermarket in each country. All these retailers included are leaders in their respective countries, with market shares of approximately 28% in Argentina, 15% in Brazil, 27% in Chile, and 30% in Colombia. The market shares for the U.S.-based retailers are not revealed for confidentiality reasons.[11]

In the United States, the data are categorized under the United Nations COICOP structure, which is used by most countries to classify CPI information. A narrower category indicator is the URL (or web address) where the products are found on the website. The retailer's website design and menu pages determines the number of URLs available in each country.

Missing values are common in daily data because products may be out of stock or not correctly scraped on a particular day. Depending on the country, the percentage of these missing values is between 22% and 37% of all observations, as shown in Table 2.[12] Price gaps, however, do not last for more than a few days. Following the literature, I therefore complete missing values by carrying forward the last recorded price until a new price is available. I do this only for the first five months of the price gap to match the approach taken by Nakamura and Steinsson (2008). In the Appendix, I show that my results are similar if I do not impute any missing values and focus exclusively on consecutive observations.

There are a few price changes in each country that seem too large and are most likely the result of scraping mistakes. Although these are a negligible part of all observations, they can affect statistics related to the magnitude of price change. Consequently, all daily price changes that exceed 200% or -70% are excluded from all duration and size calculations.

2.3.1 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. Are online prices similar to offline prices? More specifically, if we were to physically walk into a store in these retailers and collect prices for the same products, would they be similar to the prices collected on the website at the same time?

To answer this question, I conducted simultaneous surveys of *offline* and *online* prices in the retailers included in this paper. These surveys took place in Buenos Aires, Santiago, Rio de Janeiro, Bogotá, and Washington D.C. with the help of five local people. They were asked to select any branch of these supermarkets and randomly buy 100 products in 10 predefined categories. These categories were chosen to ensure some variety in the type of goods purchased. After the first purchase, I used the printed receipts to get unique product IDs and check whether the same items were sold online or not.[13] Those items that could

---

[10]See the Appendix for a similar table with details for each U.S. retailer.

[11]Revealing information on the U.S. supermarket, in particular, is strictly forbidden by the conditions of the scanner data provided by the Kilts Marketing Center at the University of Chicago Booth School of Business, used in Section 3.0.1 of the paper to compare the results with online data.

[12]The share of missing observations for monthly-sampled data is only 1.74% in the U.S.. This is lower than the 12% reported by Klenow and Kryvtsov (2008) for the U.S. CPI data.

[13]In Argentina, Brazil, and Colombia, the matching was based exclusively on product IDs. In Chile the

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 branches. The same items were bought every time in 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 during subsequent purchases.[14]

Table 3 below shows the results from this data comparison 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.

Table 3: Comparing Online vs. Offline Prices

	USA	Argentina	Brazil	Chile	Colombia
Matching IDs	Yes	Yes	Yes	No	Yes
% Available Online	89%	100%	80%	90%	74%
Price Levels					
online=offline	76%	18%	42%	93%	29%
online>offline	4%	78%	34%	4%	32%
Price Difference* (Mean %)	-4	5	9	2	0
Price Changes					
Ratio of Changes over Observations					
Offline	0.442	0.215	0.356	0.274	0.433
Online	0.469	0.215	0.411	0.249	0.433
Mean Abs Size of Changes (%)					
Offline	22	12	10	13	11
Online	23	12	11	14	10

Notes: Data from randomly selected products in each retailer. Offline and online prices were collected within a 7-day time window. In Latin America, offline prices were collected four times every 15 days in 2009. In the U.S., prices were collected in 2015 at intervals that varied from a week to a month. *Excludes identical prices.

I compared prices in terms of both 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 the United States, the share of identical prices is 79%. The price change series are highly synchronized, and both the frequency and size of price changes are similar across

---

matching was based on the item's name, description, and package size, because the online product ID did not match the offline ID printed on the ticket.

[14]The offline data collection for the Latin American supermarkets took place in 2009. The U.S. data was collected in 2015, as part of a much larger validation exercise using mobile phones and crowd-sourcing websites in 10 countries. For details see Cavallo (2016).

samples. 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 among only 12 goods (mostly raw food products), meaning that 89% of the products had identical price levels across samples. In Argentina, price levels are typically higher online, yet in nearly every case there was a difference of 5% across samples. A constant markup implies that price *changes* are highly correlated, with similar frequency and size of changes.

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

The comparison between online and offline prices is explored in great detail in my related paper, Cavallo (2016), where I simultaneously collect online and offline prices for over 50 of the largest multichannel retailers in 10 countries. The results, consistent with those in this paper, are that on average 72% of the prices are identical across samples, and that price changes have similar frequencies and sizes, even though they are not perfectly synchronized.[15]

3   How Sampling and Measurement Error Affect Pricing Statistics

Measurement error has been discussed in the literature on price stickiness before. For example, Campbell and Eden (2014) identified and removed prices which could not be expressed in whole cents in an Nielsen scanner dataset. They noted that technical errors and time aggregation could be the cause for those "fractional prices". Cavallo and Rigobon (2011) further discussed the potential effect of time-averaging and unit values on the distribution of size changes, and simulated the impact on the distribution of size changes using online data in a large number of countries. Eichenbaum et al. (2014) used CPI and scanner data from multiple stores to show how unit-value prices, reported as the ratio of a product's sales revenue to the quantity sold, affected the prevalence of small price changes.

In this paper, I focus on two topics that have the biggest impact: time averages and price imputations. An advantage relative to previous papers is that I rely on a data source unaffected by these issues. I am therefore able to re-compute some classic statistics in the

---

[15]An alternative way to test the validity of Scraped Data is to see if the inflation dynamics obtained from this small sample of retailers can resemble those in CPI statistics, which are constructed using surveys from a large number of offline stores. In a related paper, Cavallo (2013), I show that online price indexes can closely match inflation rates in most of these countries.

literature, see the effect of each sampling characteristic, and compare them to results in previous papers. In particular, I can simulate some of the sampling characteristics in Scanner and CPI data to show that they generate more frequent and smaller prices changes. More explicitly, in the case of supermarket data, I can directly compare both online and scanner data from the same retailer, geographic location, and time period, and show that time-averaging accounts for all the differences observed with scanner prices.

3.0.1 Time Averages and Scanner Data

Scanner datasets have two main potential sources of measurement error. First, prices are sometimes reported as "unit values" (the ratio of total sales over quantities). As Eichenbaum et al. (2014) show, even with daily data, if items are sometimes purchased with or without coupons or at different prices across stores of the same retailer, the unit values reported by scanner datasets can generate spurious small price changes. Although this was common in early datasets available in the literature, most of the scanner data used today, such as Nielsen's "Retailer Scanner Data", do not report unit values. Instead, they provide weekly prices for individual goods sold by stores in multiple locations.

A second major source of measurement error in scanner data is the use of weekly-averaged prices. The potential effect of time-averaging in scanner data was first discussed by Campbell and Eden (2014). Their focus was not on the size of changes, but they described some complications caused by weekly averages using an example of a three-week period with a single price change in the middle of the second week. Instead of a single price change, the weekly-averaged price data produces two price changes of smaller magnitude. The prevalence of examples such as this can potentially double the frequency of changes and greatly reduce the size of price changes. The measurement bias, however, can also go in the opposite direction. For example, if there are many small increases (or decreases) during a week, then using weekly-averaged prices would reduce the measured frequency and increase the absolute size of price changes.

To empirically estimate the effects of time-averaging, consider the results in Table 4 below, where I compare the implied monthly duration and the mean size of absolute price changes for both weekly-sampled online data and a simulated weekly-average dataset that resembles the type of data available in scanner datasets. The sampling simulation is run on the original online data by simply computing the weekly-average price for each good.[16] In all cases, I compute monthly durations using standard methods in the literature.

---

[16]Note that these prices are naturally free from unit values and averaging across stores, as they come from a single retailer and location. I am simply averaging prices over a week, and then calculating the weekly frequency of price changes and its implied monthly duration.

Table 4: Duration and the Mean Absolute Size of Changes

		USA	Arg.	Brazil	Chile	Col.
Duration (months)	Weekly Online Data	2.91	2.43	1.48	2.92	1.99
	Weekly Average	1.69	1.4	.91	1.69	1.1
Mean Abs. Size	Weekly Online Data	21.98	12.22	11.46	14.66	10.74
	Weekly Average	11.06	6.09	6.57	8.24	5.95

Notes: I first obtain the frequency per individual good by calculating the number of price changes over the number of total valid change observations for a particular product. Next, I calculate the *mean* frequency per good category, and, finally, the *median* frequency across all categories. I then compute implied durations using $-1/ln(1 - frequency)$, and convert them to monthly durations for comparisons across samples.

Both the duration and the absolute size of the price change fall by approximately 50% when weekly-averaged data are used, and the effect is similar in all countries. This effect can explain why the literature that uses scanner data has tended to find that prices are so flexible. Given that scanner data are available mainly for groceries and supermarket products, in Table 5 I show the results obtained only from the supermarket in my U.S. sample, and compare them to estimates obtained from previous papers in the literature. In particular, I include results from Dominick's Supermarket and another "Large U.S. Supermarket", both reported by Eichenbaum et al. (2011).

Table 5: Implied Duration in U.S. Supermarket Data

	Scraped	Scanner Large Retailer	Scanner Dominik's	Scraped Weekly Av.	Scanner Same Retailer
Period	2008-2010	2004-2006	1989-1997	2008-2010	2008-2010
Sampling Interval	Weekly	Weekly	Weekly	Weekly	Weekly
Duration (months)	1.53	0.6	1.0	0.8	0.8

Notes: The scanner data results in columns 3 and 4 come from Table 3 in Eichenbaum et al. (2011). The scanner data results on the last column use prices from Nielsen provided by the Kilts Center at the University of Chicago, matching with the same retailer, zip code, and time period of the online Scraped Data.

The implied duration of 1.53 months in online data are much higher in relative terms than that found in previous papers in the literature that used scanner data from U.S. supermarkets, which ranged from 0.6 to 1 month. Using weekly-averaged data produces an implied duration of 0.8 months, which is the average in the two papers mentioned above. While this is consistent with measurement error in scanner data, there are other reasons that could cause these differences. For example, the time periods of the data are quite different across papers, so it is possible that goods have become stickier in recent years. In addition, the retailers may not be the same, or even similar in their pricing behaviors and other characteristics.

To make the comparison more explicit, I purchased scanner data for the exact same retailer, location, and time period. The challenge was to find the same retailer in both datasets

because the scanner dataset, collected by Nielsen, does not explicitly identify retailers. It only provides a retailer ID, the type of store, and the zip code of each store. Fortunately, retailers tend to have a distinctive pattern of stores in different zip codes. By counting how many stores each supermarket chain in the scanner had in a given set of zip codes, I was able to find a perfect match to the retailer I used to collect the online data.[17]

The last column in Table 5 shows that the scanner data also has a duration of 0.8 which is identical to the weekly-averaged online data from the same retailer, zip code, and time period. Time-averaging is all that is needed to replicate the low duration results in scanner data. To be fair, few papers in the stickiness literature have emphasized the duration levels in scanner data, mainly because they are available only for groceries. What has received far more attention in the literature, however, is the distribution of the size of price changes.

The effects of time-averages on the size of price changes are even more evident. Figure 1 below shows the distribution of the size of price changes in the online data, in the simulated weekly-averaged data, and in the scanner data for the exact same U.S. retailer and time period.
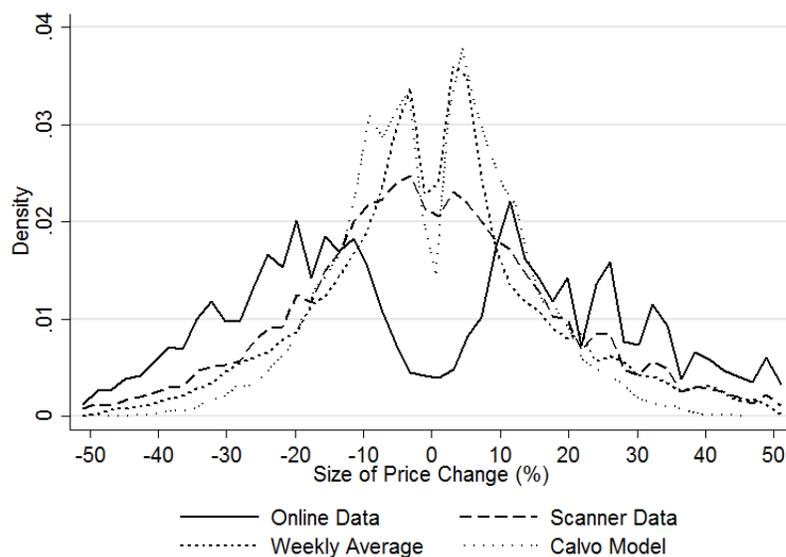

Figure 1: The Distribution of the Size of Price Changes in a U.S. Supermarket

Notes: The online and scanner data in the U.S. was collected at the same retailer during the same time period. Scanner data was collected by Nielsen and provided by the Kilts Center at the University of Chicago Booth School of Business. Results from the Calvo model were obtained by using the code and calibration parameters from Nakamura and Steinsson (2010a), adjusting the price change probability to match the observed monthly frequency of changes, and increasing the size of idiosyncratic shocks to match the absolute size of price changes (approximately 11%). See the Appendix for similar results in other U.S. sectors.

The distribution of the size of online price changes is clearly bimodal, with very few changes close to zero percent. The scanner data, by contrast, generate an unimodal distribution with a large share of small price changes close to zero percent. This is the type of

---

[17]The distribution of stores across zip codes in the online data was found by scraping the "find a store" form available on the retailer's website.

distribution that has been prevalent in the literature and motivated many papers, such as Midrigan (2011), to develop models that can account for small price changes.

Once again, weekly-averaged prices can help explain the difference. They completely change the shape of the distribution by turning larger price changes from the tails into smaller ones at the center. Indeed, the weekly-averaged and scanner dataset distributions are very similar, with the exception of the two spikes that remain in the weekly-averaged data near zero percent. One explanation for the lack of spikes in scanner data could be the effect of coupons and loyalty cards, which can create additional tiny price changes to further smooth the distribution.

In Figure 1, I also plot the distribution predicted by a simple Calvo model parametrized to match the frequency and mean absolute size of price changes in the weekly-averaged online data.[18] The resemblance with the unimodal distribution obtained from scanner datasets explains why papers such as Woodford (2009), which have models that can accommodate both state and time-dependent pricing, have tended to favor mechanisms that match the patterns predicted by the Calvo model. Instead, the actual distribution of price changes, observed with online data, has very little mass near zero percent and two modes. This is consistent with a greater role for an adjustment or "menu" cost that makes small price changes sub-optimal. In fact, the online data distribution is close to the predictions of the model in Alvarez et al. (2011), which combines both adjustment and information costs into the price-setting decision.[19] In Section 5, I provide a set of additional statistics that can be used to test and parametrize these type of models.

Finally, time-averages also have an impact on the estimated hazard rates of price adjustment. Hazard rates measure the probability of a price change as a function of the time since the previous adjustment, and different sticky-price models will have different predictions about the shape of the hazard function over time. Adjustment-cost models, for example, tend to generate upward sloping hazards if the shocks are persistent over time. Time-dependent models, by contrast, generate spikes in the hazard function at the dates when adjustment takes place.

Figure 2 below shows the daily hazard rates using the daily Scraped Data, the weekly-averaged data, and the scanner data for the same retailer. Details for the construction of these estimates are provided in the Appendix.

---

[18]The model, based on Calvo (1983), was simulated using the code from Nakamura and Steinsson (2010a). More details for the simulation are provided in the Appendix.

[19]See figure VI in that paper.

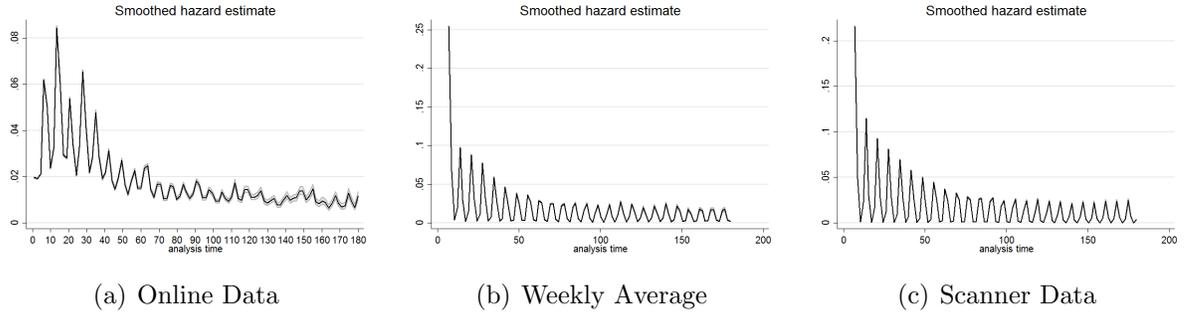| (a) Online Data | (b) Weekly Average | (c) Scanner Data |

Figure 2: Hazard Functions in a U.S. Supermarket

Notes: Initial 180 days shown. Left-censored spells are excluded.

The Scraped Data hazard function has a hump-shaped pattern, initially increasing and then gradually falling over time. By contrast, both the weekly-averaged and the scanner data produce a fully downward sloping hazard function. The reason is that with weekly averages most of the probability of a price change occurs in the first week after the previous observed change. This makes the trend of the hazard downward sloping from the start, similar to those found in Campbell and Eden (2014) with scanner data. The online data results also show weekly spikes in the hazard rates that are not observable in the other data.[20]

Once again, measurement error distorts the stylized facts in the literature. The positive trend at the beginning of the online hazard functions suggests that adjustment costs play an important role, with "older" prices having a higher probability of experiencing a change.[21] The weekly spikes are also consistent with models that have information costs, as in Alvarez et al. (2011).

3.0.2 Imputations and CPI Data

CPI micro data are not affected by time-averaging because prices are usually collected once a month.[22] There is, however, a sampling characteristic that can have an equally significant effect on price stickiness patterns: the imputation of missing prices.

Missing prices occur when the person collecting the data at the store is unable to find a particular good in stock. In some cases, a product substitution may take place, while at other times the product may be simply marked as temporarily out of stock or as a seasonal product. Many national statistical offices, including the U.S. Bureau of Labor Statistics (BLS), use an imputation method called "cell-relative imputation" to fill non-comparable

---

[20]The spikes that can be seen in figures 2(b) and 2(c) are just an artifact of the daily scale of the graphs. Plotting them on a weekly scale would make the hazard function appear smooth and completely downward sloping.

[21]These hazard functions are still affected by "survival bias", which makes hazard rates fall steadily over time as the share of stickier duration spells becomes more important. I provide some evidence of survival bias in the Appendix.

[22]There are some exceptions. The 2009 UN "Practical Guide to Producing Consumer Price Indices" (ILO et al. (2009)) notes in point 10.26 on page 151, that "in many countries, prices for at least some products are collected more than once a month". It argues that "it is inappropriate to use these individual prices [in a price index] as this will imbalance the sample of price quotations" and recommends that "instead the prices should be averaged before compiling the elementary price index".

product substitutions or temporarily missing prices. Under this approach, prices are imputed using the average observed change in the prices of goods for a similar category. This can mechanically increase the frequency of observed price changes and also reduce their size. In particular, if these average changes used in the imputation come from a large number of items, some of which may be increasing and some falling, the size of imputed price changes can be quite small in absolute value.[23]

Imputed prices due to item substitutions have been discussed in the literature before. These are usually clearly identified in the CPI datasets and can be excluded from the analysis. For example, both Klenow and Kryvtsov (2008) and Nakamura and Steinsson (2008) emphasize how removing price changes from forced substitutions reduces the frequency of price changes. Their impact on the size of price changes, however, has not been documented before. In fact, they are typically included when computing the distribution of the size of price changes in papers that use CPI data, such as Klenow and Kryvtsov (2008).

Imputed prices due to temporarily missing prices are also common in CPI datasets, though they have received less attention in the literature. They occur naturally as products go in and out of stock and can be quite prevalent in categories such as food, apparel, and electronics. Papers with U.S. CPI data tend to exclude them, as in Klenow and Kryvtsov (2008) and Nakamura and Steinsson (2008), but it is unclear how common or well identified they may be in other CPI datasets, or whether all cell-relative imputations are clearly marked in the research databases.

To illustrate the effect of cell-relative imputations on both durations and the size of price changes, I simulate the procedure with the missing prices in monthly-sampled online data. I first take the original Scraped Data (eliminating prices that had been carried forward) and keep only the price for the 15th of each month. Then, for each good, I impute missing prices within price spells by multiplying the previously available price by the geometric average of price changes for goods in the same category. The magnitude of the effect depends on the frequency of missing prices (imputations) and also on the size of the "cell" chosen to compute these average price changes. In Table 6 below, I compare the duration and size results from monthly-sampled online data with those produced by using a cell-relative simulation with both a "broad" (COICOP class) and a "narrow" category definition (URL).

---

[23]Until January 2015, the BLS used item-strata, which are relatively broad product categories, as the "cell" used for imputation. It has now moved to using more narrow elementary level items, or ELIs. See Bureau of Labor Statistics (2015b) for a description of the recent changes. In addition to the categories of goods, the imputations are applied for a given geographical aggregation level. This was traditionally the CPI index area, and is now being replaced with the narrower "Primary Sampling Unit". Geographical aggregation does not apply to the data of this paper, but it is potentially another reason for measurement bias in CPI micro data.

Table 6: Cell-Relative Imputation of Temporarily Missing Prices

		USA	Arg.	Brazil	Chile	Col.
Duration	Monthly Online Data	4.7	3.43	2.03	4.38	2.29
(months)	Cell-Relative Imputation (Broad)	1.67	2.03	1.77	3.47	1.51
	Cell-Relative Imputation (Narrow)	3.35	3.11	1.85	4.3	1.71
	Klenow and Kryvtsov (08)	3.7				
Mean Abs.	Monthly Online Data	20.82	11.54	10.07	14.29	9.92
Size	Cell-Relative Imputation (Narrow)	16.15	9.94	9.53	10.91	8.28
	Klenow and Kryvtsov (08)	14.0				
Price Changes	Monthly Online Data	1.54	4.22	7.7	3.67	7.87
below |1%|	Cell-Relative Imputation (Narrow)	7.22	11.39	12.07	13.77	16.61
	Klenow and Kryvtsov (08)	11				
Price Changes	Monthly Online Data	7.09	33.68	42.48	25.32	41.72
below |5%|	Cell-Relative Imputation (Narrow)	25.16	47.19	49.79	41.79	55.54
	Klenow and Kryvtsov (08)	40				

Notes: I first obtain the monthly frequency per individual good by calculating the number of price changes over the number of total valid change observations for a particular product. Next, I calculate the *mean* frequency per good category, and finally, the *median* frequency across all categories. I then compute implied durations using $-1/ln(1 - frequency)$, and convert them to monthly durations for comparisons across samples. Klenow and Kryvtsov (08) results taken from Tables II and III in that paper.

Table 6 shows that cell-relative imputation dramatically reduces the duration of prices. The drop is smaller when the imputation is based on the URL categories, which narrowly identify similar goods. For the rest of the paper, I will use this narrower definition.

The monthly duration of prices falls from 4.7 to 3.35 months. This is close to the 3.7 months reported by Klenow and Kryvtsov (2008) for the duration of posted prices in the U.S. CPI research database. For their estimate, Klenow and Kryvtsov (2008) excluded imputed prices for out of stock or seasonal items, but included price changes due to item substitutions, which the BLS also adjusts with cell-relative methods.[24]

The effect of imputations on the size of price changes is equally large but has received no attention in the literature before. In the lower panels of Table 6, I show the impact on the absolute size of price changes and the percentage of changes below the thresholds of 1% and 5%, which are also reported by Klenow and Kryvtsov (2008) on the U.S. CPI data. Cell relative imputation decreases the absolute size of changes from 20.82% in the U.S. to 16.15%, and makes the share of small changes increase significantly at the 1% and 5% thresholds, closer to the results reported by Klenow and Kryvtsov (2008).[25]

---

[24]Klenow and Kryvtsov (2008) report an implied median duration of 8.7 months for items excluding substitutions and all sales (regular prices). They also report a duration of 7.2 for regular prices. Assuming the only difference between these two numbers is due to substitutions, the impact of 1.5 additional months for excluding substitutions is also close to the 1.4 months I get in my estimates with posted prices.

[25]While cell-relative imputation is able to bridge the gap between my results and previous papers with CPI data, the comparison is complicated because there are many things that are potentially different between

Figure 3 compares the distribution of price changes for both online and the cell-relative imputation. The online data distribution is different from the one in Figure 1 because I include all sectors (not just supermarket products). There are more spikes than before, but a feature that is common is the lack of small price changes, particularly between -5% and +5%. Just as with time-averages, the main impact of cell-relative imputation is to increase the number of small changes. This makes the distribution unimodal, with a large mass of price changes close to zero percent and a kurtosis that rises from 3.96 to 5.45.
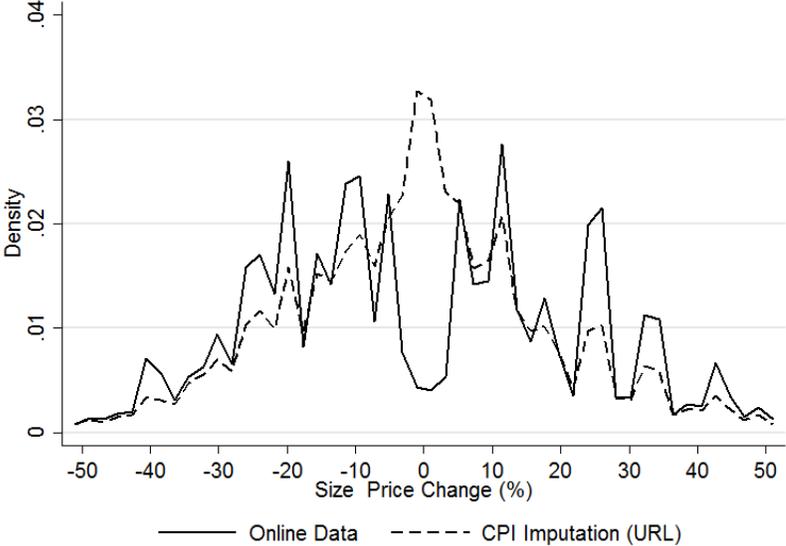
Figure 3: The Distribution of the Size of Price Changes in the U.S.

Notes: Data for all U.S. retailers is included. See the Appendix for similar results across U.S. sectors and other countries.

Finally, Figure 4 below shows that cell-relative imputation produces a downward sloping hazard function similar to the ones found in the literature. In this case, I plot the hazard function for "Food and Non-Alcoholic Beverages" to compare with the results for "Processed Food" reported by Nakamura and Steinsson (2008), but the results are similar for all categories. The effect of imputations is analogous to that of time-averages, even though the cause is different. Using an imputed price tends to create two consecutive price changes in the data, increasing the hazard rate in the first month.

---

the datasets. In particular, Klenow and Kryvtsov (2008) used data from a larger number of sectors in the U.S., and for a different time period (1988-2004). I discuss possible sector differences later on, when I make comparisons to the Nakamura and Steinsson (2008) results.

| | (a) Online Data | (b) Cell-Relative Imputation |
| | | |

Figure 4: Hazard Function with Cell-Relative CPI Imputation
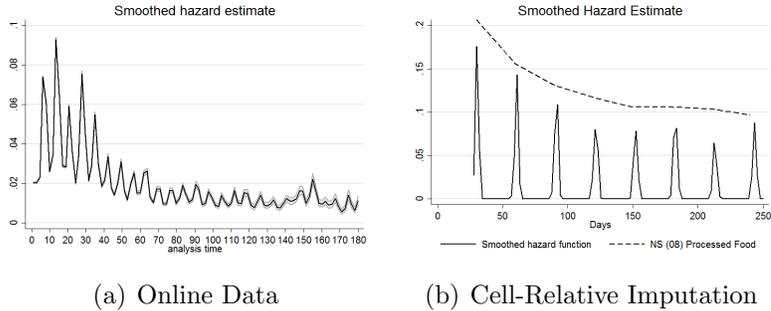
Notes: Hazard function for "Food and Non-Alcoholic Beverages" in the U.S.. Initial 180 days of smoothed hazard function shown. Cell-relative imputation of temporarily missing prices in the monthly data done at the URL level. Left-censored spells are excluded. Results labelled "NS (08)" are the monthly hazards for "Processed Food" obtained from Nakamura and Steinsson (2010b).

Although the hump-shaped pattern of the online hazard function does not match the empirical results in Nakamura and Steinsson (2008), it resembles the predictions of the *model* in that paper. In fact, the authors mention that the main difference between their model's prediction and the results with the CPI data is precisely the behavior of the hazard during the first few months, which increases in the model but not in the CPI data.[26] In categories such as 'Food and Non-Alcoholic Beverages", this initial increase is not even observable unless we use daily data.

Despite these results, the actual bias in CPI datasets may lower than what is implied in my simulations, particularly for the U.S. data. There are two main reasons for this. First, the number of temporarily missing observations generated by my simulation is higher than the 7% reported by Klenow and Kryvtsov (2008) for BLS data. Second, not every single missing price may be imputed using cell-relative methods. Still, there are other sampling characteristics in CPI data that can increase the magnitude of the bias. For example, Eichenbaum et al. (2014) show that unit-values and composite-good pricing can account for a large share of changes smaller than 1% in the U.S. CPI data. And non-missing prices can also be affected by imputations, as statistical offices often adjust price observations for coupons, rebates, loyalty cards, bonus merchandise, and quantity discounts, depending on the share of sales that had these discounts during the collection period. Examples of these and other price adjustments are described in the BLS Handbook of Methods.[27]

Overall, my results strongly suggest that imputed prices can be an important source of measurement bias for some stickiness statistics. The extent to which specific CPI datasets are affected is likely to vary across countries and time periods, but researchers need to be aware of the potential biases and should try to adjust for them.

---

[26]See Nakamura and Steinsson (2008), page 1460.
[27]See Bureau of Labor Statistics (2015a), Chapter 17, pages 30 to 33.

19

4   Robustness: Sales, Retailers, and Sectors

One of the main implications from the previous section is that differences in the sampling characteristics of the data can have a big impact on stickiness statistics. In this section, I show that these results are robust to the removal of sale prices and the use of alternative retailers or sectors. I also show how different sampling intervals can affect durations. Other robustness tests are provided in the Appendix.

4.1   Sales

In principle, we could expect the effects of the two sampling characteristics described in this paper to be particularly large in when there are many sales. For example, most of the price changes that monthly sampling miss may be caused by short-term sales or stock-outs that lead to cell-relative imputations. To see if a measurement bias exists in regular prices, I repeat the analysis after excluding sale observations. Sale prices can be identified in online data using sales flags that are sometimes captured by the scraping software from images or explicit mentions of a sale price. But not all the datasets in this paper have that information, so in this section I use a simple algorithm to identify sales based on the behavior of prices over time. In particular, I use a "v-shaped" sale algorithm used by Nakamura and Steinsson (2008) and others in the literature. It identifies sales by looking for prices that initially fall and then return to the same previous level for a period no longer than 30 days. While it may fail to identify sales that have more complicated patterns, including those that end with prices that are slightly higher than the previous price, an important advantage is that I can apply it to all retailers and countries.

Table 7 shows that the number of v-shaped sale prices as a share of all observations is highest in the United States at 4.68%. The duration and size results labelled "Ex-sales" are computed after excluding sale prices by carrying forward the last available non-sale price until the sale ends. The resulting prices are then used to compute the monthly durations and size of price changes, using the same methods as before for all sampling simulations.

As both Klenow and Kryvtsov (2008) and Nakamura and Steinsson (2008) have previously documented, durations are higher for regular prices that exclude sales. For example, the implied duration for the United States rises from 4.7 to 7.62. The effects are naturally stronger in countries with a larger share of v-shaped sales, such as the United States and Chile.

More importantly for this paper, removing sales does not eliminate the problems associated with the sampling and measurement biases in scanner and CPI data. The use of weekly averages or cell-relative imputation also tends to decrease implied durations and reduce the size of price changes. In the case of the United States, durations fall from 7.62 to 3.19 and 4.7 months, while the absolute size of changes drops from 19.12% to 10.87% and 13.93%. A similar effect can be seen in the other countries. In fact, the countries with the largest share of sale observations, the United States and Chile, are still the ones where the effect of the sampling bias is strongest.

Table 7: Sales and Regular Prices

		USA	Arg.	Brazil	Chile	Col.
Obs. with Sales		4.68%	2.55%	3.04%	3.7%	2.97%
Duration	Monthly Online Data (Posted)	4.7	3.43	2.03	4.38	2.29
(months)	Monthly Online Data (Ex-Sales)	7.62	3.92	2.26	6.03	2.6
	Weekly Average(Ex-Sales)	3.19	2.17	1.21	2.93	1.45
	Cell-Relative Imputation (Ex-Sales)	4.7	3.51	2.00	5.47	1.91
Mean Abs.	Monthly Online Data (Posted)	20.82	11.54	10.07	14.29	9.92
Size	Monthly Online Data (Ex-Sales)	19.12	10.7	9.17	12.76	8.9
	Weekly Average (Ex-Sales)	10.87	5.58	5.34	7.41	5.15
	Cell-Relative Imputation (Ex-Sales)	13.93	9.32	8.63	9.61	7.34

Notes: To obtain monthly implied durations, I first compute the monthly frequency per individual good by calculating the number of price changes over the number of total valid change observations for a particular product. Next, I calculate the *mean* frequency per good category, and then, the *median* frequency across all categories. Finally, I compute implied durations using $-1/ln(1 - frequency)$, and convert them to monthly durations for comparisons across samples.

4.2 U.S. Retailers and Sectors

My findings are also robust when using data from different retailers. In the case of the United States, I re-calculate everything for each one of the four retailers in the data. Results are shown in Table 8. The first column with numbers shows results for the full database in the United States, which has been discussed in previous tables, while the next column labeled "USA1" is the supermarket used for Table 5.

Table 8: U.S. Retailers

Statistic		Source	USA (All)	USA1 Smkt	USA2 Dept.S	USA3 Drugs.	USA4 Elect.
Duration		Monthly Online Data	4.7	2.35	13.49	7.33	6.84
(months)		Weekly Average	1.69	.83	7.2	1.65	2.2
		Cell-Relative Imputation (Narrow)	3.35	1.41	12.49	6.53	6.3
Mean Abs.		Monthly Online Data	20.82	20.6	16.1	29.16	15.93
Size		Weekly Average	11.06	11.67	9.43	11.24	7.85
		Cell-Relative Imputation (Narrow)	16.15	16.73	11.36	25.8	13.64

Notes: To obtain monthly implied durations, I first compute the monthly frequency per individual good by calculating the number of price changes over the number of total valid change observations for a particular product. Next, I calculate the *mean* frequency per good category, and then the *median* frequency across all categories. Finally, I compute implied durations using $-1/ln(1 - frequency)$, and convert them to monthly durations for comparisons across samples.

The impact is similar in all types of retailers, and as before, weekly averages and cell-

relative imputations decrease durations and the absolute size of price changes.

The fact that these four retailers sell different types of goods suggests that my findings are not limited to a single sector. The impact across sectors can be seen more explicitly in Table 9, where I present the results for the U. S. data categorized at the first level of the UN's COICOP classification structure. These sectors are affected in similar ways. Weekly averages reduce monthly durations by approximately 70% on average, while cell-relative imputation makes them fall by about 30%. The median absolute size of price changes falls by approximately 50% with weekly averages and 25% with cell-relative imputations.

Table 9: U.S. Sectors

	Food & Bev.	Alc. Bev.	Aparrel	House-hold Goods	Health	Comm.	Elec.
*COICOP Code	100	200	300	500	600	800	900
Duration (months)							
Monthly Online Data	2.12	2.3	3.09	5.84	4.69	5.8	7.41
Weekly Averaged	0.73	0.66	1.69	1.64	1.05	1.68	2.66
Cell-Relative Imputation (Narrow)	1.45	2.07	2.35	4.11	3.35	3.82	4.27
Nakamura and Steinsson (08)*	2.6	2.4	2.69	3.8	5.91	5.2	3.18
Median Abs. Size							
Monthly Online Data	23.89	16.14	28.6	14.25	25.18	20.39	15.18
Weekly Averaged	14.16	8.73	14.2	7.89	12.21	9.46	7.99
Cell-Relative Imputation (Narrow)	19.79	13.78	24.23	10.64	15.63	14.23	11.18
Nakamura and Steinsson (08)*	27.66	11.09	30.7	12.59	18.41	22.2	14.26
Missing Observations	21.19%	26.90%	27.78%	31.16%	34.14%	36.23%	41.08%

Notes: *Weighted monthly duration and weighted median absolute size. Results computed using only the subsectors for which there is also online data. ELI-level CPI results obtained from Nakamura and Steinsson (2010b).

These sector-level results can be compared to the U.S. CPI statistics reported by Nakamura and Steinsson (2008). To make the comparison as close as possible, in Table 9 I show the *median* size of absolute changes and re-compute the CPI results reported by Nakamura and Steinsson (2008) to include only the sub-sectors that are covered by the online data. Given that Nakamura and Steinsson (2008) explicitly exclude imputed prices due to substitutions and temporarily missing observations, we can expect the monthly-sampled online prices to approximate their results. This seems to be the case for durations in "Food and Beverages", "Alcoholic Beverages", "Health", and "Communications", where their implied durations are closer to the result with online data. However, the estimates for "Household Goods" and "Electronics" appear to be quite different in terms of duration. In these cases, the Nakamura and Steinsson (08) durations are shorter and closer to the numbers produced by cell-relative imputations. The last row in Table 9 suggests that the differences might be

related to the share of missing observations in the data. In particular, electronics tend to be frequently missing and out of stock in the online data, so it is possible that some of the prices in the CPI micro data were also missing and imputed but not identified as such in the database. Nevertheless, there are many other reasons why the online results could be different from those in Nakamura and Steinsson (2008), including the fact that time periods are not the same.

4.3 Sampling Interval: Daily, Weekly, and Monthly Data

Online prices are collected every day, scanner datasets every week, and CPI data once a month (and sometimes every two months). These differences in sampling intervals can affect the measured duration and size of price changes, even if the underlying price data remain the same. The reason is that many price changes can take place within the sampling interval. For example, monthly sampled data can miss a lot of price changes that occur during the month. An advantage of using high-frequency online data is that I can quantify these sampling-interval effects.

To isolate the differences that come exclusively from the sampling interval, in Table 10 I compare the duration and absolute size of price changes in online data when sampled at daily, weekly, and monthly intervals. The weekly data are sampled each Wednesday, while the monthly data are sampled on the 15th of each month. Results are robust to picking other days of the week or month, as shown in the Appendix.

Table 10: Sampling Interval: Daily, Weekly, Monthly

		USA	Arg.	Brazil	Chile	Colombia
Duration (months)	Daily Online Data	2.68	2.12	1.27	2.5	1.7
	Weekly Online Data	2.91	2.43	1.48	2.92	1.99
	Monthly Online Data	4.7	3.43	2.03	4.38	2.29
Mean Abs. Size	Daily Online Data	22.08	12.45	11.52	15.07	10.98
	Weekly Online Data	21.98	12.22	11.46	14.66	10.74
	Monthly Online Data	20.82	11.54	10.07	14.29	9.92

Notes: Weekly data are sampled each Wednesday. Monthly data are sampled on the 15th of each month. Results are robust to picking other days of the week or month. To compute the duration of price changes, I first obtain the frequency per individual good by calculating the number of price changes over the number of total valid change observations for a particular product. Next, I calculate the *mean* frequency per good category, and finally, the *median* frequency across all categories. I then compute implied durations using $-1/ln(1 - frequency)$ at the sampled interval. Finally, I convert daily and weekly durations to monthly durations for comparisons across samples. The statistics for the U.S. are weighted using CPI expenditure weights. No weights are used in the other countries.

Daily and weekly sampled data have similar durations and sizes of price changes in every country. This is because few products have more than one price change per week. Picking a different day of the week to do the sampling can affect the timing somewhat, but does not affect the implied duration or statistics related to the size of price changes.

Monthly sampling, by contrast, tends to increase durations considerably. Prices appear stickier with monthly data because many price changes within the month are not observed. This affects durations but not the absolute size of price changes, as can be seen in the lower panel.

Whether these effects are relevant or not depends on whether we care about high-frequency price changes. Many of these temporary price changes are connected to sales, which the literature has tended to exclude in the past. At the same time, if the bias introduced by monthly sampling is stable over time, then papers that focus on the dynamic properties of stickiness are not significantly affected. On the other hand, using daily data can compensate for the other two biases mentioned in the previous section. In any case, care should be taken to acknowledge and account for any effects introduced by the use of high-frequency sampling in the data.

5   Implications for the Stickiness Literature

The results in the previous sections suggest that some empirical "stylized facts" in the stickiness literature are affected by sampling biases and measurement errors caused by the characteristics of both scanner and CPI data sources.

For the empirical literature, this implies that prices are stickier than we previously thought (conditional on the sampling interval) and that most of the "small" price changes observed in previous papers are spuriously caused by the sampling characteristics of both scanner and CPI micro datasets.

For the theoretical literature, my findings suggest that more caution is needed when interpreting the stylized facts. In particular, less emphasis should be placed in trying to explain frequent and small price changes or downward sloping hazards. Many of the pricing behaviors observed with online data resemble the ones produced by models that allow for a combination of observations and menu costs, such as the one developed by Alvarez et al. (2011). But this paper is not meant to provide empirical evidence for a particular model. In fact, for some applications the specific model used might not really matter. For example, Alvarez et al. (2016) show that a sufficient statistic for the real effects of monetary policy in a large set of models is given by the ratio of the frequency and the kurtosis of the distribution of price changes. To rely on these statistics, however, we need to make sure we measure them with as little error or bias as possible. Furthermore, the heterogeneity in pricing behaviors, documented in this and other papers in the literature, implies that we need to measure them with a large number of retailers, sectors, and even countries in order to have robust estimates.

With this goal in mind, in Table 11 I show summary statistics using Scraped Data from 181 retailers in 31 countries. This large cross-section of retailer-level information can be used to evaluate the robustness of my findings in the previous section, parametrize models, and shed some light on pricing behaviors across countries. These prices were collected by PriceStats, a private company connected to the Billion Prices Project at MIT.[28] I used them to compute the monthly duration for both posted and regular prices, as well as several moments of the distribution of price changes. Additional statistics are shown in the Appendix. The last

---

[28]PriceStats LLC was co-founded by Alberto Cavallo. More details on the data collection methods and applications for these datasets can be found in Cavallo and Rigobon (2016).

column shows the monthly implied durations from the literature, summarized and reported by Klenow and Malin (2010) from different papers that use CPI data in different countries.

Comparing columns (3) and (9), it is clear that monthly-sampled online prices are stickier than those previously reported in the literature that uses CPI data. For example, in the United States the mean implied duration for monthly-sampled online prices is 9.5 months when computed with 29 retailers, much higher than the comparable implied durations from mean frequencies reported by Klenow and Malin (2010) at 3.2 months. It is also stickier than the 4.6 months from weighted medians reported by Nakamura and Steinsson (2008). If I remove symmetric v-shaped sales, the duration rises to 12 months, twice as high as the 6 months for the equivalent estimate in Nakamura and Steinsson (2008).[29] Something similar happens in other countries where mean durations have been reported by various papers in the literature. It is important to note that some of the differences may be caused by compositional differences. As mentioned before, there are many things that are different in my results relative to others in the literature, including the time periods and the sectors covered in each country. Nevertheless, the pattern of higher durations with online data seems to be quite robust.

Table 11 also shows that the results obtained for five countries in Section 3.0.2 are typical of a much larger set of economies. In particular, the median duration for posted prices is relatively high, at 9.7 months; small price changes are also rare in most countries, with a median of 4.1% below |1%| in absolute value, and the median absolute size of price changes is large, with a median of 17.6%.

There is also a lot of heterogeneity in results across countries that can be used to make cross-country comparisons. One interesting fact is that inflation rates are not correlated with the overall frequency or size of price changes. For example, Venezuela has an annual inflation rate of 37.5%, yet it is one of the stickiest countries in the sample. Turkey has twice as much inflation as Chile, even though the duration and size of changes is similar. In the Appendix, I show that inflation is correlated with the relative frequency of increases over decreases and the relative size of price increases over decreases. This suggests that factors at the country level can make prices in a country more "flexible" or "sticky", or have "big" or "small" changes, but what really matters for inflation (and therefore the real effects of monetary policy) is how much more frequent (or larger) price increases are relative to decreases.

Understanding the results for each country is beyond the scope of this paper, but Table 11 illustrates how online data can be used to produce these statistics with identical methods across countries and over time. Doing so will not only allow researchers to better understand price stickiness, but perhaps more importantly, it can eventually provide central bankers with measures of stickiness that can be used to inform policy decisions in real time.

6 Conclusions

This paper introduces a new way of collecting price data and applies it to study basic stylized facts in the price stickiness literature. Scraped data, obtained directly from online

---

[29]See Table IX in Nakamura and Steinsson (2008). I use a sales filter equivalent to the one they label "Sale Filter B, 1-month window", which has a monthly frequency of 15.3%, and therefore an implied duration of 6.02 months.

Table 11: Implied Monthly Durations and Size of Changes in 31 Countries

Country	(1) Retailers	(2) Inflation (%)	(3) Duration (months)	(4) Duration ExSales (months)	(5) Size Under |1%|	(6) Size Under |5%|	(7) Mean Abs. Size	(8) Kurt.	(9) Duration Literature (months)
Argentina	14	17.1	4	4.3	2.8	22.4	13.8	7.1	
Australia	5	2.5	6.2	7.1	5.4	16.1	25	3.7	
Austria	2	2.4	9.7	16.3	16.4	50.8	11.9	3.4	6.1
Belgium	4	1.9	11.8	12.9	8.2	49.3	9.3	6.7	5.4
Brazil	5	6.1	4.5	5.1	4.1	24.7	12.9	5.2	2.2
Canada	7	1.8	12.3	14.7	2.3	13.8	21.3	4.7	
Chile	5	3.2	7.4	8.8	3.5	19.5	18	4.8	1.6
China	3	3.2	21.9	26.5	4.7	27.5	17.2	6.1	
Colombia	4	2.9	11.3	12.3	3.5	23.2	16.7	5.7	
France	2	1.5	8.9	10.1	11.2	37.3	15.9	4.9	4.8
Germany	6	1.7	19.5	21.9	3.4	25	15	5.6	8.3
Greece	6	.5	7	8.2	12.1	28.1	15.8	4.9	
Hungary	3	2.8	11.6	15.4	4	12.3	23.8	3.2	6.1
India	3	8.7	9.8	11.8	3.2	20.5	18.2	4.4	
Indonesia	4	5.5	10.8	13.1	4.9	30.3	12.3	6.6	
Ireland	6	1	7.1	9.8	1.6	8.7	29.2	3.5	
Israel	4	1.8	8.3	10.3	3.9	10.2	23.2	3.5	3.6
Italy	4	1.9	8.7	9.7	10.1	25.2	20.3	4	9.5
Korea	6	2.2	15.2	17.9	2.2	17.4	17.5	4.4	
Netherlands	2	2	18.7	20.6	3.2	14	21.2	4.9	5.5
New Zealand	5	1.9	6.6	10.6	2.4	8	27.4	3.2	
Norway	4	1.5	6.3	7.4	7.9	21.1	19.9	5	4.0(4.2)
Portugal	2	1.7	16.4	19.1	5.5	24	15	5.1	4.0
Russia	5	7	8.5	8.9	14.4	36.7	13.1	5.6	
Singapore	7	3.3	17.8	21.7	1.5	12.8	17.4	4.1	
South Africa	5	5.6	10.2	11.5	7.5	14.9	19.8	5.3	
Spain	9	1.7	11.4	13	7	35.9	12	5.3	6.2
Turkey	7	7.9	7.5	9.8	1.9	22.3	17.6	4.1	
UK	11	2.8	8.6	21.7	4.9	14.1	26.2	4.1	4.7(6.2)
USA	29	2.1	9.5	12.1	2.5	10.1	24.8	4.3	3.2(4.2)
Venezuela	2	37.5	13.7	13.8	5.2	24	23.6	3.4	
Mean	6	4.6	10.7	13.1	5.5	22.6	18.6	4.7	
Median	5	2.4	9.7	12.1	4.1	22.3	17.6	4.8	

Note: I use monthly-sampled data collected from 181 large multichannel retailers in 31 countries selling food, groceries, electronics, apparel, furniture, household products, and related goods. Prices were collected between 2007 and 2014, with different start dates for each retailer. Each statistic is calculated at the retailer level and then averaged within countries. The average gives the same importance to each retailer within a country. The simple mean and median over all countries is reported on the last two rows. The mean duration for Eurozone countries is 11.9. The column labeled "Literature" shows the implied monthly durations computed from the *mean* monthly frequencies reported in Table 1 of Klenow and Malin (2010), with results that exclude sales in parentheses. See

retailers, provide a unique source of price information. Prices are easier to collect than in CPI and scanner data, and can be obtained with daily frequency for all products sold by retailers around the world. The data are available without any delays and the collection methodology can be customized to satisfy the specific needs of sticky-price studies. More importantly for the stickiness literature, Scraped Data are free from common sources of measurement error, such as time-averages and imputation methods that can affect traditional micro-price datasets.

I use the scraped online data to show how measurement bias affects three common stylized facts in the literature: the duration of price changes, the distribution of the size of changes, and the hazard functions. I argue that two sampling characteristics in scanner and CPI datasets can produce biased results for these statistics. Weekly averages and price imputations tend to reduce the duration of price changes (particularly in scanner data), decrease their size, and make the hazard function more downward sloping over time. I show this with sampling simulations in my own data, and confirm this explicitly in scanner data by comparing both online and scanner data collected from the same retailer and time period. I further show that these sources of measurement error account for nearly all the differences between my results and those in the existing empirical literature. The paper ends with a discussion of the implications for the stickiness literature and a table with summary statistics on stickiness that uses online data from 31 countries. These results provide confirmation that prices are stickier than previously reported in many countries and that small price changes are actually rare in the data. Furthermore, they provide cross country statistics that can be used to parametrize sticky-price models and make cross-country comparisons.

I emphasize that there is nothing intrinsically wrong with scanner and CPI datasets. As with any type of data, including scraped online prices, there are advantages and disadvantages for various uses. The point I make in this paper is that researchers need to be aware of how the sampling and other characteristics of the data may affect what they are trying to measure. Scanner and CPI data were not designed for the measurement of price stickiness, so it is natural that some of the sampling decisions made by the data collectors are not ideal for this purpose. Unfortunately, it is not always possible to control for these biases because the datasets available for research may not provide enough details on how prices are treated or adjusted. One of the main advantages of online data is that we can collect the prices as posted by the retailers, and we can decide how to treat them depending on the particular statistic that we are trying to measure. The ability to access the raw and unfiltered data is a characteristic of many other new "Big Data" sources of information, from crowd-sourced data, to Twitter feeds and mobile phone sensors.

This paper focuses on how scraped online data affects the measurement of stylized facts in the stickiness literature, but the potential uses of Scraped Data in macroeconomics go far beyond those explored here. For example, scraped prices can be used to create daily price indexes that complement official statistics, compare and test theories of international prices, and better measure exchange rate and commodity shocks pass-through. As discussed in Cavallo and Rigobon (2016), many of the datasets collected by the Billion Prices Project, including those in this paper, are publicly shared at bpp.mit.edu so other researchers can further explore these and other potential uses of Scraped Data.

References

ALVAREZ, F., M. GONZALEZ-ROZADA, A. NEUMEYER, AND M. BERAJA (2015): "From hyperinflation to stable prices: Argentinas evidence on menu cost models," *SCID Working Paper 470*.

ALVAREZ, F., F. LIPPI, AND H. LE BIHAN (2016): "The real effects of monetary shocks in sticky price models: a sufficient statistic approach," *American Economic Review*, Forthcoming.

ALVAREZ, F. E. AND F. LIPPI (2014): "Price Setting with menu cost for Multi-product firms," *Econometrica*, 82, 89–135.

ALVAREZ, F. E., F. LIPPI, AND L. PACIELLO (2011): "Optimal price setting with observation and menu costs," *Quarterly Journal of Economics*, 126, 1909–1960.

BILS, M. AND P. J. KLENOW (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.

BOIVIN, J., H. E. C. MONTRAL, M. P. GIANNONI, AND I. MIHOV (2009): "Sticky Prices and Monetary Policy : Evidence from Disaggregated U . S . Data ," *American Economic Review*, No. 12824, 350–384.

BONOMO, M., C. CARVALHO, AND R. GARCIA (2011): "Time- and State-Dependent Pricing: A Unified Framework," SSRN Scholarly Paper ID 1930633, Social Science Research Network, Rochester, NY.

BRYNJOLFSSON, E., A. A. DICK, AND M. D. SMITH (2003): "Search and Product Differentiation at an Internet Shopbot," *SSRN Electronic Journal*.

BUREAU OF LABOR STATISTICS (2015a): *The Consumer Price Index*, vol. Chapter 17 of *Handbook of Methods*, BLS.

——— (2015b): "New CPI Estimation System to be Introduced," .

CALVO, G. (1983): "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, 12, 383–398.

CAMPBELL, J. R. AND B. EDEN (2014): "Rigid prices: evidence from U.S. scanner data ," *International Economic Review*, 55, 423–442.

CAVALLO, A. (2013): "Online and official price indexes: Measuring Argentina's inflation," *Journal of Monetary Economics*, 152–165.

——— (2016): "Are Online and Offline Prices Similar? Evidence from Multi-Channel Retailers," *American Economic Review*, Forthcoming.

CAVALLO, A. AND R. RIGOBON (2011): "The Distribution of the Size of Price Changes," *NBER Working Paper*, w16760, 1–40.

——— (2016): "The Billion Prices Project: Using Online Data for Measurement and Research," *Journal of Economic Perspectives*, 30, 151–78.

CECCHETTI, S. G. (1986): "The frequency of price adjustment," *Journal of Econometrics*, 31, 255–274.

COSTAIN, J. AND A. NAKOV (2011): "Distributional dynamics under smoothly state-dependent pricing," *Journal of Monetary Economics*, 58, 646–665.

DHYNE, E., L. J. LVAREZ, H. L. BIHAN, G. VERONESE, D. DIAS, J. HOFFMANN, N. JONKER, P. LNNEMANN, F. RUMLER, AND J. VILMUNEN (2006): "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data," *Journal of Economic Perspectives*, 20, 171–192.

EICHENBAUM, M., N. JAIMOVICH, AND S. REBELO (2011): "Reference prices, costs, and nominal rigidities," *American Economic Review*, 101, 234–262.

EICHENBAUM, M., N. JAIMOVICH, S. REBELO, AND J. SMITH (2014): "How frequent are small price changes?" *American Economic Journal: Macroeconomics*, 6, 137–155.

EINAV, L. AND J. LEVIN (2014): "Economics in the age of big data," *Science*, 346, 1243089.

ELLISON, G. AND S. F. ELLISON (2009): "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77, 427–452.

ELLISON, S. F., C. M. SNYDER, AND H. ZHANG (2015): "Costs of Managerial Attention and Activity as a Source of Sticky Prices: Structural Estimates from an Online Market," Tech. rep., Mimeo, Massachusetts Institute of Technology.

GAGNON, E. (2009): "Price Setting During Low and High Inflation: Evidence from Mexico," *The Quarterly Journal of Economics*, 124, 1221–1263.

GORODNICHENKO, Y. (2008): "Endogenous information, menu costs and inflation persistence," *NBER Working Paper*.

GORODNICHENKO, Y., V. SHEREMIROV, AND O. TALAVERA (2014): "Price Setting in Online Markets: Does IT Click?" *NBER Working Paper Series*.

GORODNICHENKO, Y. AND O. TALAVERA (2016): "Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-border Integration," *American Economic Review*.

ILO, IMF, W. BANK, OECD, EUROSTAT, AND U. NATIONS (2009): *Practical guide to producing consumer price indices*, UN.

KASHYAP, A. K. (1995): "Sticky Prices: New Evidence from Retail Catalogues." *Quarterly Journal of Economics*, 110, 245–274.

KLENOW, P. J. AND O. KRYVTSOV (2008): "State-Dependent or Time-Dependent Pricing: Does it Matter for Recent US Inflation?" *The Quarterly Journal of Economics*, 73(3), 863–903.

KLENOW, P. J. AND B. A. MALIN (2010): "Microeconomic evidence on price-setting," in *Handbook of Monetary Economics*, Elsevier, vol. 3.

KLENOW, P. J. AND J. L. WILLIS (2007): "Sticky information and sticky prices," *Journal of Monetary Economics*, 54, 79–99.

LACH, S. AND D. TSIDDON (1996): "Staggering and Synzhronization in {Price-Setting:} Evidence from Multiproduct Firms," *American Economic Review*, 86(5), 1175–1196.

LUNNEMANN, P. AND L. WINTR (2011): "Price Stickiness in the US and Europe Revisited: Evidence from Internet Prices*," *Oxford Bulletin of Economics and Statistics*, 73, 593–621.

MIDRIGAN, V. (2011): "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations," *Econometrica*, 79, 1139–1180.

NAKAMURA, E. AND J. STEINSSON (2008): "Five Facts About Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123(4), 1415–1464.

——— (2010a): "Monetary Non-neutrality in a Multisector Menu Cost Model," *The Quarterly Journal of Economics*, 125, 961–1013.

——— (2010b): "Supplement for: Five Facts About Prices: A Reevaluation of Menu Cost Models," .

——— (2013): "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics*, 5, 133–163.

WOODFORD, M. (2009): "Information-constrained State-dependent Pricing," *Journal of Monetary Economics*, 56, S100– S124.

WULFSBERG, F. (2009): "Price adjustments and inflation - evidence from Norwegian consumer price data 1975-2004," *Norges Bank Working Papers*.