American Economic Review 2017, 107(1): 283–303
https://doi.org/10.1257/aer.20160542

Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers[†]

By Alberto Cavallo*

Online prices are increasingly used for measurement and research applications, yet little is known about their relation to prices collected offline, where most retail transactions take place. I conduct the first large-scale comparison of prices simultaneously collected from the websites and physical stores of 56 large multi-channel retailers in 10 countries. I find that price levels are identical about 72 percent of the time. Price changes are not synchronized but have similar frequencies and average sizes. These results have implications for national statistical offices, researchers using online data, and anyone interested in the effect of the Internet on retail prices. (*JEL* D22, L11, L81, O14)

Online prices are increasingly used for measurement and research applications. Since 2008, the Billion Prices Project (BPP) at MIT has been experimenting with daily online price indexes in the United States and other countries.[1] National statistical offices (NSOs) have recently started to consider the use of online data in official Consumer Price Indices (CPIs).[2] In the context of academic research, online prices are being used for a wide range of topics, including the study of price competition, market segmentation, price stickiness, international relative prices, and real exchange rate dynamics.[3]

Despite their growing appeal, an open fundamental question about online prices is whether they are similar to the prices that can be collected offline in physical

* Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, and NBER (e-mail: acavallo@mit.edu). Financial support for this research was provided by the JFRAP at MIT Sloan and the NBER's Economics of Digitization and Copyright Initiative. I greatly benefited from the comments of referees and seminar participants at the NBER/CIRW, MIT Sloan, the Ottawa Group Meeting, and the UNECE/ILO Meeting of the Group of Experts on Consumer Price Indices. I thank Maria Fazzolari for her outstanding work implementing and coordinating all the data collection. I also thank the MIT and Wellesley University students that collected some of the offline prices: Vivian Xi, Maurizio Boano, Sibo Wang, Descartes Holland, Sabrina Lui, Suh Yoon, Holly Zhu, Sean Bingham, Elizabeth Krauthamer, Jeffrey Zhang, William Rodriguez, Wenxin Zhang, Jake Amereno, Ivy Jiang, Diya Mo, Qi Jin, Riley Quinn, Do Yeon Park, Jung Hyun Choi, Xiaoxi Wang, Aaroshi Sahgal, Isaiah Udotong, Giulio Capolino, Tanya Bakshi, Allison Davanzo, Karen Pulido, and Bailey Tregoning. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.
[†] Go to https://doi.org/10.1257/aer.20160542 to visit the article page for additional materials and author disclosure statement.

[1] See Cavallo (2013) and Cavallo and Rigobon (2016).

[2] See Horrigan (2013); Griffioen, de Haan, and Willenborg (2014); Boettcher (2015); Breton et al. (2015); Krsinich (2015); Nygaard (2015); and Krsinich (2016).

[3] See Chevalier and Goolsbee (2003); Brynjolfsson, Hu, and Simester (2011); Edelman (2012); Cavallo, Neiman, and Rigobon (2014); Gorodnichenko, Sheremirov, and Talavera (2014); Simonovska (2015); Alvarez, Le Bihan, and Lippi (2016); Cavallo (forthcoming); and Gorodnichenko and Talavera (2017).

stores. The question is important because relatively few retail transactions take place online. In fact, according to Euromonitor (2014), online purchases are currently less than 10 percent of all retail transactions in the United States, and even lower in other countries.

This paper provides the first large-scale comparison of online and offline prices in large multi-channel retailers designed to answer this question. Using a combination of crowdsourcing platforms, a mobile phone app, and web scraping methods, I simultaneously collected prices in both the online and offline stores of 56 of the largest retailers in 10 countries: Argentina, Australia, Brazil, Canada, China, Germany, Japan, South Africa, United Kingdom, and the United States. These data are used to compare price levels, the behavior of price changes, and the selection of products available for sale in the offline and online stores. I document country, sector, and retailer heterogeneity, and test whether online prices vary with IP address locations or persistent browsing habits. The results have implications for NSOs and researchers using online data, as well as those interested in the effect of the Internet on retail prices.

The data collection effort is unprecedented in scope and size, and was carried out as part of the BPP. I first selected the retailers to be sampled by focusing on the top 20 companies by market shares in each country that sell both online and offline ("multi-channel"), and have product barcodes that can be matched across samples. Next, I used crowdsourcing platforms such as Amazon Mechanical Turk, Elance, and UpWork to hire 323 workers to collect the offline data. Each worker was assigned a simple task: to scan the barcodes and collect prices for a random set of 10 to 50 products in any physical store of a given retailer. In some cases they had to return to the same store multiple times to scan the same set of products. Using a special app for android phones developed to simplify and standardize the data collection process, these workers scanned each product's barcode, manually entered the price, took a photo of the price tag, and sent all the information via e-mail to the BPP servers, where it was automatically processed and cleaned. A scraping software then used the barcode numbers to look for the same product in the website of each retailer, and collected the online price within a period of seven days. The matched online-offline dataset contains prices for more than 24,000 products and 38,000 observations sampled between December 2014 and March 2016.

The main finding is that online and offline price levels are identical about 72 percent of the time, with significant heterogeneity at the country, sector, and retailer level. These percentages range from 42 percent in Brazil to 91 percent in Canada and the United Kingdom. The United States is close to the average, with 69 percent. At the sector level, drugstores and office-product retailers have the lowest share of identical prices, with 38 percent and 25 percent, respectively, while in electronics and clothing these numbers rise to 83 percent and 92 percent, respectively. When there is a price difference, the online markup tends to be small, with a magnitude of −4 percent in the full sample. If I include observations with identical prices, the online price difference is only −1 percent on average.

I also find that price *changes* have similar frequencies and sizes in the online and offline data. However, only 19 percent of weekly price changes occur at the same time. While this is higher than the unconditional probability of a simultaneous price change, the individual price series are not well synchronized.

The reasons for the online-offline price differences seem to vary across retailers and countries. Sales tend to create some discrepancies, with only 36 percent of sale prices being identical across samples, but they have a small impact in the aggregate results because the number of sale observations is relatively small (11 percent of the total dataset). A similar thing happens with *offline* price dispersion across physical stores, which tends to be low. Using a small sample of offline prices collected for multiple zip codes on the same day, I find that about 78 percent of goods have a single price within stores of the same retailer. I also found no evidence of "dynamic pricing" strategies that could potentially cause online-offline differences. At least in the United States, online prices do not change with the location of the IP address of the computer connecting to the website or when the scraping robot repeatedly browses the same web page of a particular good for a prolonged period of time. There is also no evidence that online-offline price differences are being driven by attempts to match the prices of Amazon.com, which are identical to the online prices in multi-channel retailers about 38 percent of the time.

In terms of product selection, 76 percent of the products sampled offline were also found online by either using the automated scraping matching or by manually searching for the product description on the website. The price comparison results for goods that can be automatically matched are similar to those that had to be manually matched. There is also no evidence that retailers try to obfuscate the online-offline price comparisons by changing the products' identification numbers.

Despite the general similarity in online and offline prices, there is significant heterogeneity in pricing behaviors across retailers. Three main types of companies stand out: those with nearly identical online and offline prices, those with stable online markups (either positive or negative), and those with different prices that are not consistently higher or lower online. Some of these patterns seem to be sector-level behaviors, while others are common for most retailers within a country.

For research economists using online data, these results provide evidence that most large multi-channel retailers price similarly online and offline. There are both advantages and disadvantages of using online data, as I discuss in Cavallo (forthcoming), but the ability to collect a massive amount of prices so cheaply provides unprecedented opportunities for economic research. My results suggest that these prices are valid sources of information for retail transactions, even those that take place offline. Retailer heterogeneity, however, implies that researchers using relatively few sources of data should be cautious to understand particular pricing patterns and control for any sampling biases.

For NSOs, these results imply that the web can be effectively used as an alternative data-collection technology to obtain the same prices found offline. Prices collected through the web are very similar to those that can be obtained at a much higher cost by physically walking into a store. While many challenges to the use of online data in CPIs remain, such as the more limited sectoral coverage or the lack of quantity data, my results should help alleviate concerns about the peculiarities of prices collected online. The BPP app and methodology developed in this paper are also publicly available at bpp.mit.edu to be used for more country and retailer-specific validation tests, which are sensible given the high degree of heterogeneity in pricing behaviors.

Lastly, my findings have implications for people interested in the effects of the Internet on retail prices. The fact that online prices are the same for all locations and also similar to offline prices collected from many different zip codes implies there is little *within-retailer* price dispersion. I also show this explicitly with some offline data in multiple zip codes in Section IVB. In practice, most retailers seem to have a single price for the majority of products, regardless of the location of the buyer and whether the product is sold online or at a particular offline store. This suggests that while the web has not reduced price dispersion across different retailers, as documented by a large literature surveyed by Baye, Morgan, and Scholten (2006), it may have created incentives for firms to price identically in their own stores. This type of *within-retailer* price dispersion has received little attention in the literature, even though it could have large welfare implications within countries.

This paper is related to a literature that studies the behavior of online prices. Some papers written in the early 2000s compared manually-collected prices of online retailers and traditional brick-and-mortar stores in a few narrow categories of goods. For example, Brynjolfsson and Smith (2000) compared prices for CDs and books in both online-only and multi-channel retailers ("hybrids" in their notation). They report that online prices were 9–16 percent lower and had smaller price changes, but note that "findings would be strengthened if we excluded hybrid retailers from our comparisons of price levels," which implies that online and offline prices for multi-channel retailers were closer together (Brynjolfsson and Smith 2000, p. 572). Clay et al. (2002) also found similar prices for 107 books in both the websites and some physical stores of Barnes & Noble and Borders, which is consistent with my results.[4] More recent comparisons of online and offline prices expanded on the categories covered but were limited to small ad-hoc samples in a few stores. Examples include Cavallo, Neiman, and Rigobon (2014, 2015); Borraz et al. (2015); and Cavallo (forthcoming). A separate branch of the literature uses online prices from "shopbots," or price comparison websites, which are easier to collect. Examples include Brynjolfsson and Smith (2001); Brynjolfsson, Dick, and Smith (2009); Ellison and Ellison (2009a, b); Lunnemann and Wintr (2011); Gorodnichenko, Sheremirov, and Talavera (2014); and Gorodnichenko and Talavera (2017). Although these papers do not directly compare prices with offline data, their results suggest that online prices change more frequently and with smaller sizes than comparable findings in papers with offline CPI prices. The difference with my findings is likely caused by their focus on retailers that participate in price-comparison websites. As Ellison and Ellison (2009a) discuss, such retailers face a uniquely competitive environment that can significantly affect their pricing behaviors.

[4] For other papers in this literature, see OECD (1998); Tang and Xing (2001); Clemons, Hann, and Hitt (2002); and Xing, Yang, and Tang (2006).

I. Simultaneous Online-Offline Data Collection

A. *Multi-Channel Retailers*

There are many types of "online prices," from those in marketplaces such as eBay, online-only retailers such as Amazon, and those at stores with both an online and offline presence. In this paper, I focus on the prices of large "multi-channel" retailers that sell both online and offline. When considering all retail sales, this type of retailer still concentrates the vast majority of all retail transactions, making them the most important source of price data for applications that require "representative" data (such as inflation measurement). Despite its importance, this is also the type of online price that has received the least attention in the academic literature due to lack of data. Furthermore, as pointed out by Brynjolfsson, Hu, and Rahman (2013), technology is blurring the distinctions between physical and online retailing, making both traditional brick-and-mortar and online-only companies behave increasingly like multi-channel retailers.

B. *Retailer Selection*

The names of the retailers included in the data collection are shown in Table 1. They satisfy three conditions. First, they are in the list of top 20 retailers by market share in their respective countries. The rank information was obtained from Euromonitor International's Passport Retailing Global Rankings. This condition helps to ensure a representative sample of the retail sector. Second, they sell both online through a country-specific website and offline through physical stores. Most large retailers satisfy this condition. Third, there is a way to perfectly match products online and offline. In practice, this means that the product id number collected offline can be used to find the product on the website.

C. *Collecting Offline Prices in Physical Stores*

Collecting prices offline is normally an expensive and complicated process. NSOs rely on a large number of trained data collectors to do it correctly. Unfortunately, the micro data collected by NSOs for CPI purposes cannot be used for my comparisons because the retailer and product details are confidential information. Lacking the budget for a traditional data collection effort, I looked for alternatives using new technologies. In particular, I relied on popular crowdsourcing platforms, such as Amazon Mechanical Turk, Elance, and UpWork, to find people willing to do simple data collection tasks. To minimize the chance of data-entry errors, I developed a custom mobile phone app that simplified the data collection process.

Crowdsourcing platforms have many advantages. First, they allowed me to hire a large number of workers and reach multiple locations and cities within each country. Second, there were enough workers to limit the number of individual prices that each had to collect. This reduced the burden on the worker and also minimized the "showrooming" concerns of the retailers. Showrooming is a term used to describe the practice of visiting a physical store to examine a product but later purchasing it online in another store. Many retailers worry about people who

Table 1—Retailers Included

Argentina	Carrefour, Coto, Easy, Sodimac, Walmart
Australia	Coles, Masters, Target, Woolworths
Brazil	Droga Raia, Extra, Magazine Luiza, Pao de Azucar, Renner
Canada	Canadian Tire, Home Depot, The Source, Toys R Us, Walmart
China	Auchan Drive, Sams Club
Germany	Galeria Kaufhof, Obi, Real, Rewe, Saturn
Japan	Bic Camera, K's Denki, Lawson, Yamada
South Africa	Clicks, Dis-Chem Pharmacy, Mr Price, Pick n Pay, Woolworths
United Kingdom	Asda, Marks and Spencer, Sainsburys, Tesco
United States	Walmart, Target, Safeway, Stop&Shop, Best Buy, Home Depot, Lowe's, CVS, Macy's, Banana Republic, Forever 21, GAP, Nike, Urban Outfitters, Old Navy, Staples, OfficeMax/Depot.

Notes: These retailers satisfy three conditions. First, they are in the list of top 20 retailers by market share in their respective countries according to Euromonitor International. Second, they sell both online through a country-specific website and offline through physical stores. Third, there is a way to perfectly match products online and offline for the price comparison. See the online Appendix for more detailed characteristics and results.

use mobile apps to scan the product's barcode and buy products online at other retailers, so if the data collectors spent too much time at each store, they might be asked to leave.[5]

Two main versions of the task were posted on the crowdsourcing websites. In the simplest case, the worker had to use a mobile app provided by the BPP team to scan 10 to 50 random offline products in any physical store, with some basic instructions to spread out the data collection across categories of goods. This provided the bulk of the data that I use to compare price levels across samples. A more complex version of the task required the worker to return to the same store every week for a full month and scan the same items. This gave me the panel of prices that I use to study price changes in Section III.

The mobile app was custom-built to simplify and standardize the data collection process. It is an app for android phones called "BPP @ MIT," available for download at the Google Play Store.[6] Every time a worker visits a store, she clicks on a button to open a new file. For the first product, she has to enter the store's name, zip code, and country. Then she scans the UPC barcode of the product (or the barcode on the price tag, depending on the particular retailer instructions provided), manually enters the price shown in the price tag next to the product (including all sales), marks the price as "regular" or "sale," and takes a photograph of the price tag (which is used to detect errors and validate the data). All products are scanned in a loop which makes the process quick and simple. When done, the worker taps an icon to e-mail the data to the BPP servers. A member of the BPP team verifies the submitted data and pays the worker.

[5] I tried to conduct a similar large-scale offline data collection with MIT students in the Boston area in 2011, but most of them were asked to stop and leave the stores after some time. Collecting data this way appears to be easier now that more people use smartphones inside stores. Indeed, Fitzgerald (2013) reports that the fear of showrooming has faded for many US retailers. See Balakrishnan, Sundaresan, and Zhang (2013) for an economic analysis of showrooming practices.

[6] See https://play.google.com/store/apps/details?id=com.mit.bpp. The app can be downloaded for free, but a "project code" must be requested from the BPP team. This code is used to separate the data from different projects. See http://bpp.mit.edu/offline-data-collection/ for more details.

Every few hours, the BPP servers automatically processed the incoming offline files to clean and consolidate the data for each retailer. The offline barcode information was then used to collect the online price from the retailer's website, as described below.

D. *Collecting Online Prices on Each Retailer's Website*

To collect online prices, I built a custom scraping "robot" for each retailer. These robots are specialized software that are programmed to use the product barcode to query the retailer's website and collect the online price and other product information. In most cases, the robot was designed to use the website's search box to enter the product id obtained offline. For more general details on the BPP's online scraping methods, see Cavallo and Rigobon (2016).

The price collected online is the posted price for the product on the retailer's website, including any sales or discounts that apply to all customers. Whether taxes are added or not depends on the display conventions for prices in each country, but the same condition applies both online and offline. For example, US prices include sales but are typically shown without taxes, both on the website and the price tags found in physical stores. In all other countries, sales or value-added tax rates are usually included in the price in both locations. Shipping costs are never included in these online prices, so my comparisons are for posted prices excluding shipping costs. Retailers have different ways to charge for shipping. The most common is a set of shipping fees that varies with the total amount of the sale or weight of the products. Some retailers offer free shipping, which could mean that they adjust their online prices to compensate. The results at the retailer level provide information that can be used to determine when this occurs.

Nearly all of the online retailers in the sample have a single price online for each product, independent of the location of the buyer. In other words, someone purchasing a laptop from Best Buy in San Francisco sees the same price as someone doing it from Boston. The only exceptions are supermarkets, which sometimes require buyers to enter their zip code or location before displaying prices. There are only five retailers that do this in my sample. I always use the same zip code when collecting data online, independently of the one where the offline price was obtained, so this can cause some price-level differences between the online and offline data for those retailers. In the online Appendix, I use a scraping experiment with one of the largest US supermarkets to show that even retailers that ask for zip code information tend to price their goods identically in most locations. Furthermore, removing this type of retailer has little impact on my aggregate results.

For all benchmark results, I allow online prices to be collected within seven days of the offline price and also exclude sale prices. Results are similar for prices collected on the same day, or including sale prices, as shown in the online Appendix.

E. *The Online-Offline Matched Data*

Table 2 shows the main characteristics of the matched data. I collected prices in 56 retailers for more than a year, between December 2014 and March 2016. There are more than 24,000 products and 38,000 observations in total.

Table 2—Data by Country

	Retailers (1)	Start (2)	End (3)	Workers (4)	Zip codes (5)	Products (6)	Observations (7)
Argentina	5	02/15	08/15	18	23	2,324	3,699
Australia	4	03/15	08/15	13	22	3,073	3,797
Brazil	5	05/15	03/16	18	26	1,437	1,915
Canada	5	12/14	07/15	15	45	2,658	4,031
China	2	07/15	03/16	5	6	410	513
Germany	5	03/15	03/16	9	20	1,215	1,604
Japan	4	04/15	03/16	7	23	1,127	2,186
South Africa	5	03/15	03/16	21	31	2,336	3,212
United Kingdom	4	03/15	05/15	12	32	1,661	2,094
United States	17	12/14	03/16	206	274	7,898	15,332
All countries	56	12/14	03/16	323	499	24,132	38,383

Notes: Column 1 shows the number of retailers. Columns 2 and 3 show the start and end months of data collection. Columns 4 and 5 report the number of workers that collected the data and zip codes with offline prices. Columns 6 and 7 provide the number of products and price observations that could be matched with both online and offline information. Details by retailer are provided in the online Appendix.

The data coverage varies across countries. The effort was concentrated in the United States, with 17 retailers and about 40 percent of all observations. On the other extreme is China, with only two retailers. I was unable to expand the offline data collection in China because large retailers explicitly prohibit taking photographs and recording prices at physical locations. Apparently, showrooming is more extended in China, so retailers try to prevent the use of mobile phones in their stores. A survey conducted by IBM in 2013 found that about 24 percent of people in China admitted to having visited a physical store to buy online, compared with only 4 percent in the United States.[7]

II. Price Levels

Table 3 compares the price levels across the online and offline samples. Column 3 shows the percentage of observations that have identical online and offline prices up to the second decimal.

The percentage of identical prices is 72 percent for all pooled observations and for the average across countries. Some countries, such as Japan, have percentages close to 50 percent, while others such as Canada and the United Kingdom have over 90 percent of all prices being identical online and offline. The United States is close to the average, with 69 percent of identical prices.

Columns 4 and 5 show the share of prices that are either higher or lower online. Conditional on a price difference, most countries tend to have lower online prices, with the exception of Argentina and Australia. The three countries with the lowest percentages of identical prices, where differences matter the most, tend to have heterogeneous behaviors. In Argentina, nonidentical prices tend to be higher online,

[7] See Klena and Puleri (2013).

TABLE 3—COUNTRY: PRICE-LEVEL DIFFERENCES (*Percent*)

Country	Retailers (1)	Observations (2)	Identical (3)	Higher online (4)	Lower online (5)	Online markup (6)	Online difference (7)
Argentina	5	3,699	60	27	13	3	1
Australia	4	3,797	74	20	5	5	1
Brazil	5	1,915	42	18	40	−7	−4
Canada	5	4,031	91	3	5	−5	0
China	2	513	87	7	6	3	0
Germany	5	1,604	74	4	23	−8	−2
Japan	4	2,186	48	7	45	−13	−7
South Africa	5	3,212	85	6	9	−3	−1
United Kingdom	4	2,094	91	2	7	−8	−1
United States	17	15,332	69	8	22	−5	−1
All countries	56	38,383	72	11	18	−4	−1

Notes: Column 3 shows the percentage of observations that have identical online and offline prices. Column 4 shows the percent of observation where prices are higher online and column 5 the percentage of prices that are lower online. Column 6 shows the online markup, defined as the average price difference excluding cases that are identical. Column 7 shows the average price difference including identical prices.

with an average markup of 3 percent. In Brazil, they are lower, with a markup of −7 percent. Japan is an outlier, with prices that are lower online 45 percent of the time, with an average markup of −13 percent.

The average size of the price differences is quite small. This can be seen in columns 6 and 7, where a positive number means that prices are higher online. Column 6 shows the online "markup," excluding cases where prices are identical, while column 7 shows the online "difference," which includes cases with no price difference. The online markup tends to be small, with a magnitude of −4 percent in the full sample. Adding prices that are identical makes the online-offline price difference only −1 percent on average.

Overall, these results show little difference between online prices collected from the website of multi-channel retailers and the offline prices that can be obtained by visiting one of their physical stores.

The aggregate results, however, hide important heterogeneity at the sector level. Table 4 shows similar results for retailers grouped by the type of good they sell.

Drugstores and office-supply retailers have the lowest share of identical prices online and offline. For office products, prices are sometimes higher and sometimes lower online, without any clear patterns, as if the stores were managed independently. Drugstores, by contrast, tend to have lower prices online, possibly because they are "convenience" stores such as CVS and Walgreens in the United States that may charge higher prices to offline customers.

Electronics and clothing have the highest share of identical prices. For clothing, prices are basically the same, with most of the observed differences possibly coming from offline data collection errors. For electronics, prices are lower online 13 percent of the time, with an average markup of −9 percent (the highest in this sample).

Figure 1 shows the histograms for nonzero price differences in each country. The cases of Argentina and Australia stand out because there are spikes around the 5 percent magnitude of differences. This is caused by stable markups in online prices for

TABLE 4—SECTOR: PRICE-LEVEL DIFFERENCES (*Percent*)

Sector	Retailers (1)	Observations (2)	Identical (3)	Higher online (4)	Lower online (5)	Online markup (6)	Online difference (7)
Food	10	5,953	52	32	15	3	1
Clothing	7	2,534	92	5	3	3	0
Household	9	7,875	79	5	16	−8	−2
Drugstore	4	3,053	38	11	52	−5	−3
Electronics	5	3,712	83	4	13	−9	−1
Office	2	1,089	25	37	38	1	1
Multiple/Mix	18	14,149	80	5	15	−9	−2

Notes: Data classified into sectors at the retailer level. Column 3 shows the percentage of observations that have identical online and offline prices. Column 4 shows the percent of observation where prices are higher online and column 5 the percentage of prices that are lower online. Column 6 shows the online markup, defined as the average price difference excluding cases that are identical. Column 7 shows the average price difference including identical prices.

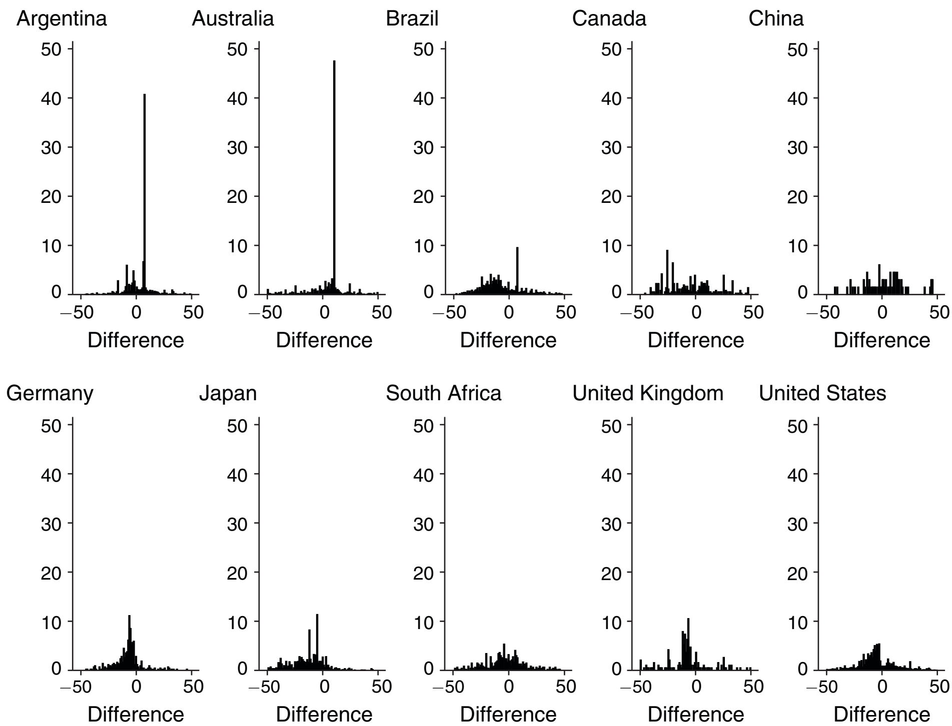

FIGURE 1. HISTOGRAMS OF NONZERO PRICE-LEVEL DIFFERENCES (*Percent*)

Notes: Price differences excluding identical prices. A positive number means that the online price is higher than the offline price. Histogram scales are matched across countries. Bin width is 1 percent.

some of the largest retailers. In all other countries, the price differences are more dispersed in the range of −50 percent to 50 percent.

As pointed out by Nakamura and Steinsson (2008), sale events are frequent in some countries, and the magnitude of the price changes that they generate can be large. I do find that sale prices create more differences between online and offline samples, the share of identical online and offline prices for sale observations being only 36 percent. But this has little impact on the full-sample results because the number of sales is small: only 11 percent of all matched observations have either an online sale (4.12 percent), an offline sale (5.03 percent), or both (1.92 percent).[8]

Similarly, restricting the sample to include only prices collected on the same day (instead of a seven-day window) has little impact on the main results. The reason is that prices do not typically change more than once a week. Details are provided in the online Appendix.

Another potential reason for some of the price-level differences is that goods have prices with similar time series that are not synchronized. I look for direct evidence of this in the next section, by comparing online and offline *changes* for a smaller sample of goods for which I have multiple weekly observations.

III. Price Changes

This section compares the behavior of price changes in the online and offline samples. A price change is computed here as a nonzero log difference in the price between weeks t and $t + 1$. I study the frequency, size, and timing of price changes.

Table 5 shows that the frequency of online and offline price change is quite similar. The first two columns show the number of observations and price changes. There are fewer observations than in previous sections because I have a short time series for a limited subset of goods, and only about 10 percent of those observations have a price change. The frequency statistics reported in columns 3 and 4 are computed for each individual good first (as the share of observations with a price change), and then averaged across countries. Column 5 shows the p-value of a two-sided t-test with a null hypothesis of equal average frequencies in the online and offline samples. I can only reject the null of equality with some confidence in the cases of Australia and Japan. Although the full sample results appear to have slightly more frequent changes online, this is entirely driven by the data from Japan.

In addition to similar frequencies, online and offline price changes tend to have similar sizes. This can be seen in columns 6 and 7, where I report the mean absolute size of price changes. Column 8 is again the p-value of a two-sided t-test of equality in the online and offline means. The null hypothesis can only be rejected in Canada, where online price changes seem to be larger. In all other countries, the difference is not statistically significant.

Similar frequencies and sizes do not imply that price changes are perfectly synchronized. This can be seen in Table 6, which focuses on the timing of changes. Price changes can occur online, offline, or in both locations. Column 3 reports the percentage of price changes for a given product that occur both online and offline at the same time, which I refer to as "synchronized." Only 19 percent of the 1,328 price changes are synchronized across online and offline samples. While this is higher than the unconditional probability of a simultaneous price change shown in column 4 (using the unconditional frequencies and assuming independence), these price series are still far from being perfectly synchronized.

[8] My ability to control for sales is somewhat limited because workers could not identify offline sales with the app until October 2015, and some of the scrape jobs were not able to include online sale indicators. It is therefore possible that the main results still contain a lot of sales that I cannot control for, and the share of identical prices would rise significantly if these observations were removed.

Table 5—Country: Price-Change Frequency and Size

	Observations (1)	Price changes (2)	Mean frequency online (3)	Mean frequency offline (4)	Equality test p-value (5)	Mean absolute size online (percent) (6)	Mean absolute size offline (percent) (7)	Equality test p-value (8)
Argentina	1,392	245	0.137	0.146	0.56	13.61	12.46	0.57
Australia	759	72	0.056	0.090	0.07	45.76	42.62	0.67
Brazil	483	85	0.167	0.138	0.36	10.55	9.36	0.53
Canada	1,427	120	0.077	0.068	0.48	31.11	21.71	0.06
Germany	419	16	0.035	0.041	0.74	27.08	15.86	0.26
Japan	1,071	98	0.074	0.014	0.00	12.10	8.20	0.34
South Africa	882	109	0.100	0.077	0.17	23.33	16.99	0.11
United Kingdom	429	25	0.046	0.070	0.28	47.68	41.78	0.67
United States	7,505	563	0.052	0.046	0.33	23.78	21.31	0.20
All countries	14,367	1,328	0.076	0.068	0.07	22.02	19.94	0.10

Notes: China is excluded due to lack of price change data. The first two columns show the number of observations and price changes. The frequency statistics reported in columns 3 and 4 are computed for each individual good as the share of observations with a price change, and then averaged across countries. Column 5 shows the p-value of a two-sided t-test with a null hypothesis of equal average frequencies in the online and offline samples. Columns 6 and 7 report the mean absolute size of price changes. Column 8 shows the p-value of a two-sided t-test of equality in the online and offline means.

Table 6—Synchronized Price Changes

	Observations (1)	Price changes (2)	Synchronized price changes (percent) (3)	Unconditional probability (percent) (4)
Argentina	1,392	245	35	2.0
Australia	759	72	22	0.5
Brazil	483	85	18	2.3
Canada	1,427	120	32	0.5
Germany	419	16	31	0.1
Japan	1,071	98	1	0.1
South Africa	882	109	15	0.8
United Kingdom	429	25	44	0.3
United States	7,505	563	11	0.2
All countries	14,367	1,328	19	0.5

Notes: China is excluded due to lack of price change data. Column 3 reports the percentage of price changes for a given product that occur both online and offline at the same time, which I refer to as "synchronized." The unconditional probability of a synchronized price change in column 4 is obtained by multiplying the frequencies of price change in Table 5.

Overall, these results suggest that the online and offline price series behave similarly but are not perfectly synchronized. In a related paper, Cavallo and Rigobon (2016), we find evidence that online price inflation tends to anticipate offline CPI inflation. A faster adjustment to shocks could be the reason why online price changes are not synchronized with offline changes. Unfortunately, the limited panel data available does not allow me to explicitly test this hypothesis.

IV. Other Reasons for Online-Offline Differences

In this section, I consider three other potential reasons for the differences between online and offline prices that required a special data-collection effort: different online prices based on IP address or persistent browsing habits, multiple offline prices in different physical stores, and attempts to match prices at Amazon.com.

A. *IP Address Location and Persistent Browsing*

There have been reports suggesting that some retailers change online prices based on the browsing habits of the consumer or the location associated with the IP address of the computer being used to purchase online. See, for example, Valentino-DeVries, Singer-Vine, and Soltani (2012) and Mikians et al. (2012, 2013). If these pricing behaviors are common for the multi-channel retailers in my sample, they could help explain some of the price-level differences in the data. To test whether prices vary with browsing habits or IP address, I ran two experiments with special versions of the scraping robots for US retailers.

The first experiment was designed to test whether prices change based on the zip code associated with the IP address of the computer collecting the data. IP addresses are unique numeric identifiers for computers that are connected to a network. They are assigned by Internet service providers and have an associated geographical location that is public information. For example, MIT's campus IP addresses range from 18.0.0.0 to 18.255.255.255 and are geographically linked to the 02139 zip code in Cambridge, Massachusetts. In principle, retailers could detect the IP address of the consumer visiting a site and automatically change the prices displayed based on its geolocation information. To test if this is happening, I randomly selected 5 products in each of the 10 US retailers and scraped their prices 12 times in a consecutive loop. In each loop, I changed the IP address of the robots by using 12 different proxy servers in 9 US cities (Atlanta, Burbank, Charlotte, Chicago, Cleveland, Miami, Nashville, New York, and two proxies in Phoenix) and 2 international locations (Canada and United Kingdom).[9] I did not find any evidence of this type of price discrimination. In all cases, prices were the same for a given product, regardless of what IP address was used to connect to the retailer websites.

The second experiment was designed to test if frequent visits to the web page of a particular product could lead the retailer to change the price displayed. In this case, I scraped a single product in each retailer every five minutes for a full day. Once again, there was no evidence of price discrimination based on persistent-browsing habits: prices were always the same.

While these forms of online price discrimination may be important in other industries (for example airlines and hotels), my results suggest that they are not commonly used in large multi-channel retailers in the United States. A likely reason is that retailers may fear antagonizing their customers if reports of these tactics were to become publicized in the press, as it famously happened in 2000 with Amazon's pricing tests.[10]

[9] A proxy server is a computer that acts as an intermediary for the communications between two other computers in a network—in this case, between the machine where the scraping software runs and the server hosting the website of the retailer. From the retailer's website perspective, the request was coming from the IP address associated with the proxy server.

[10] See CNN (2000) and Valentino-DeVries, Singer-Vine, and Soltani (2012) for a more recent example. A pricing strategy that appears to be more common than price discrimination is called "steering," in which the retailer changes the order or ranking of goods shown to customers based on their location or browsing characteristics. See, for example, Mattioli (2012).

B. *Offline Price Dispersion*

Most retailers have a single price *online* regardless of the location of the buyer, so a second potential reason for online-offline differences may be that there are some price difference across physical stores.

To test for the effects of *offline* price dispersion, I use a small subset of products for which I have offline prices for multiple zip codes collected on the same day. These data include 406 observations in 9 retailers and 46 zip codes in the United States. Table 7 shows the results for the online-offline comparison restricted to this multi-zip code dataset.

There are several things to note here. First, even though the sample is small, we get roughly the same share of identical online-offline prices that are reported in Table 3 of this paper, with 60 percent of the prices being identical online and offline. Second, as expected, goods that have different offline prices across zip codes tend to have much lower probability of identical online-offline prices, about 35 percent of the time. Third, if we focus exclusively on products with the same offline price everywhere, labeled "Identical Offline," the percentage of identical online-offline prices rises from 60 percent to 67 percent.

While offline price dispersion can create online-offline price differences, the impact is limited because there is not much offline dispersion to begin with. Indeed, about 78 percent of products sampled have the same price in different physical stores within the same retailer, as seen in column 2. Sector results range from 66 percent in drugstores to 96 percent in electronics, consistent with the sectoral differences in the online-offline comparison in Section II.[11] In the online Appendix, I further show that a large multi-channel supermarket that explicitly asks online buyers to enter their zip codes also tends to limit the amount of price dispersion across locations. Overall, these results reinforce the finding that price dispersion is low for both online and offline prices within multi-channel retailers.

To some readers, the lack of offline price dispersion may appear to be at odds with a growing literature that uses scanner data and documents a significant price difference across physical stores. For a recent example, see Kaplan and Menzio (2015). There are many reasons that can explain the apparent differences with my results. First, many papers in this literature compare data from different retailers, so that *within* retailer price dispersion is mixed with *between* retailer price dispersion. My results focus exclusively on price differences *within* retailers. Second, the price in scanner datasets is typically a weekly average. As I discuss in Cavallo (forthcoming), this can cause significant measurement error for some applications. For example, consider a good with identical prices in two stores, a price change on a Wednesday, and a single transaction in each store. If one store sold the good on a Monday, and the other on Friday, the "weekly" price will appear to be different when in fact prices were identical on a daily basis. Similarly, some scanner datasets tend to have unit values instead of prices. These are calculated as the ratio of sales to quantities sold, and can therefore be affected by the number of coupons used or the share of transactions that take place at different prices. Of course, for some

[11] See the online Appendix for more details as well as results from a larger dataset that includes offline observations for which no online price is available.

TABLE 7—ONLINE-OFFLINE PRICE-LEVEL DIFFERENCES FOR MULTIPLE ZIP CODES (*Percent*)

Country	Retailers (1)	Observations (2)	Identical (3)	Higher online (4)	Lower online (5)	Online markup (6)	Online difference (7)
United States	9	406	60	11	29	−4	−2
Different offline	7	85	35	16	48	−5	−3
Identical offline	8	316	67	9	24	−3	−1

Notes: Column 3 shows the percentage of observations that have identical online and offline prices. Column 4 has the percent of observations where prices are higher online and column 5 the percentage of prices that are lower online. Column 6, is the online markup, defined as the average price difference excluding cases that are identical. Column 7 is the average price difference including identical prices.

purposes it makes sense to include coupons or transaction weights that affect the price actually paid by the consumer, but the fact that there is price dispersion caused by coupons should not lead us to believe that prices for the same goods are shown with different prices across stores of the same retailer. Third, price dispersion is often measured within a month or a quarter, so much of the difference in observed prices is caused by the same good being bought at different times. Finally, most scanner datasets contain prices for groceries and related goods. These are also the sectors for which I find more online-offline price dispersion, as well as offline price differences across physical stores.

C. *Amazon Pricing*

A third potential reason for differences in online and offline prices is that multi-channel retailers may be matching their online prices to those in online-only retailers such as Amazon.com, and by doing so, they create a wedge with the prices at their physical stores.

To test this possibility, I created a special dataset that contains three prices for each product: the offline price at a multi-channel retailer, the online price in the same retailer, and the price at Amazon.com. The matched data contain 1,361 observations from 455 products and 8 multi-channel retailers: Best Buy, CVS, Walmart, Target, Lowe's, Macy's, OfficeMax, and Staples. The Amazon prices considered here correspond to those products marked as "Sold by Amazon.com." To be consistent with the rest of the paper, I focus on prices collected within seven days and excluding sales. More details on how these data were collected, as well as results for products with sales or sold by third-party sellers are provided in the online Appendix.

Figure 2 compares Amazon's prices separately to both the *offline* and *online* prices from multi-channel firms. A large share of prices are identical in both cases, which is surprising given that this is comparing prices across different retailers. As expected, Amazon's prices are closer to the online prices. They are identical to the online prices approximately 38 percent of the time, and the average price difference is −5 percent. The same estimates for the Amazon-offline comparison are 31 percent and −6 percent respectively.

This finding does not mean that multi-channel retailers are making their online and offline prices different to match the online price to Amazon's. In fact, as Table 8 shows, the conditional probability of having an identical online price with Amazon

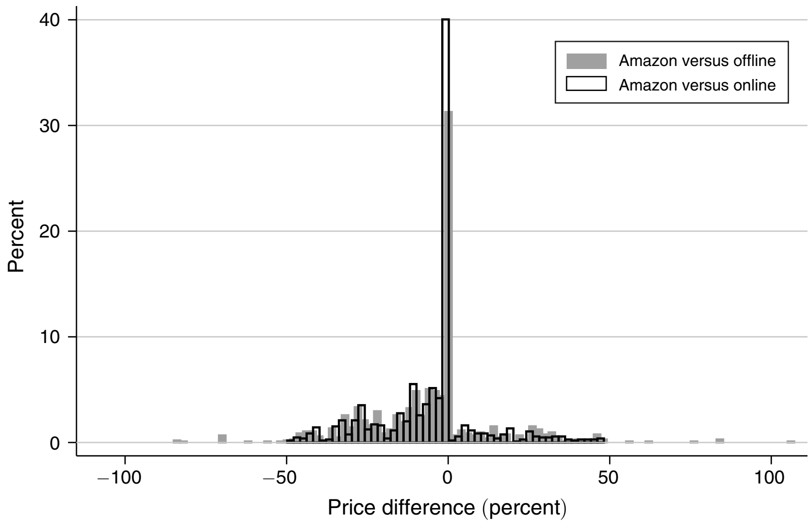

Figure 2. Price Differences with Amazon.com (*US Only*)

Notes: Price difference in Amazon.com prices relative to the *offline* and *online* prices from multi-channel retailers obtained from 1,361 observations from 455 products and 8 multi-channel retailers: Best Buy, CVS, Walmart, Target, Lowe's, Macy's, OfficeMax, and Staples. A negative number means Amazon is cheaper. Results for products marked as "Sold by Amazon.com" are shown here, with prices collected within seven days and excluding sales. More details and results for products with sales or sold by third-party sellers at Amazon's "Marketplace" are provided in the online Appendix.

is roughly the same for goods with identical online-offline prices than for those that have some online-offline price difference. The same conclusion can be obtained by running a simple probit regression of an identical online-offline price on an identical Amazon-online price. There is no statistically significant relation between these two variables. The only indication that Amazon's prices matter for the online-offline price differences is found in columns 6 and 7, which show that the difference with Amazon's prices is smaller for goods that are not identical within the multi-channel retailers.

V. Product Selection

The similarity between online and offline prices in previous sections would have different implications if most goods sold offline were not available online. I therefore now estimate the "overlap" in product selection across samples, defined as the share of offline goods that are also available online.[12]

In principle, I could use the 63 percent of offline barcodes received through the app for which the scraping software found data online. The problem with this number, however, is that the automated matching procedure can fail for many reasons: the

[12] Note that, given the data characteristics, I can estimate how many offline products are also sold online, but not the other way around. In some retailers, the online selection of goods can be larger than in a single physical store because online sales can be shipped from large centralized warehouses. See Quan and Williams (2015) for a recent discussion of the welfare effect of online and offline product variety.

TABLE 8—AMAZON: ONLINE PRICE-LEVEL DIFFERENCES (*Percent*)

Country	Retailers (1)	Observations (2)	Identical (3)	Higher Amazon (4)	Lower Amazon (5)	Amazon markup (6)	Amazon difference (7)
All observations	8	1,049	38	14	47	−9	−5
Identical on-off	8	801	38	11	51	−10	−6
Different on-off	8	248	38	25	37	−3	−2

Notes: There are 312 observations with an Amazon price and either an online or offline price, but not both, so they are excluded from these results. Column 3 shows the percentage of observations that have identical Amazon and online prices at multi-channel retailers. Column 4 has the percentage of prices that are higher in Amazon and column 5 the percentage of prices that are lower in Amazon relative to the online prices. Column 6, is the Amazon markup, defined as the average price difference excluding cases that are identical. Column 7 is the average price difference including identical prices.

worker may scan the wrong barcode, the app may incorrectly read the barcode, or the scraping robot may fail while checking the website. To get a better estimate of the overlap degree, the BPP team manually checked how many of the offline products could also be found online for a sample of 100–200 observations per retailer using all the information submitted by the workers, including the product description readable in the photo of the price tag. The results, grouped by country, are reported in Table 9.

As can be expected given the large product variety in these websites, a large fraction of goods found offline can also be found online. On average, 76 percent of all products randomly collected at the physical stores could also be found on the retailer's website. There are important differences among countries, although they seem to be unrelated to the findings in previous sections. China and Germany have the lowest overlap, while Australia, Brazil, and the United Kingdom have the highest. In the United States, 81 percent of offline products were also found online.

Furthermore, both the automatic and manually-matched goods produced similar results for online and offline price-level comparisons, as shown in the online Appendix. This finding rules out the possibility that goods that could not be automatically matched were precisely those for which the online and offline prices are different. This would happen, for example, if retailers changed the online id number for those goods to obfuscate their price differences and prevent comparisons. The evidence suggests that this is not generally the case.

VI. Retailer Heterogeneity

The country-level results in the previous sections conceal a great deal of heterogeneity across retailers in each country. Details for each retailer can be seen in online Appendix Table A1, where I show price-level and price-changes results for all retailers with at least 100 observations.

Three main types of retailers are typical. First, there are retailers where online and offline prices are identical most of the time. These are cases where the retailer explicitly chooses to have the same online and offline price. Second, there are also some retailers with a low share of identical prices, but no clear online markups. Many retailers in Brazil, for example, exhibit this pattern. These are likely cases where the online store is simply treated as another outlet, sometimes cheaper, sometimes more expensive. Third, there are retailers with a low share of identical prices

Table 9—Retailer: Product Selection Overlap

	Sample (1)	Found automatically (2)	Found manually (3)	Total overlap (percent) (4)
Argentina	500	294	52	73
Australia	500	435	36	95
Brazil	400	331	12	86
Canada	500	279	132	85
China	100	50	3	53
Germany	400	178	23	52
Japan	500	329	61	74
South Africa	500	332	60	76
United Kingdom	500	373	59	86
United States	1,600	1,003	316	81
All countries	5,500	3,604	754	76

Notes: The BPP team took a random sample of 100–200 offline prices per retailer and manually searched for the same products in the corresponding website. Column 2 shows the number of products that were found online by the automated process used to build the matched dataset in the paper. Column 3 shows the number of products that were missed by the automated process but were found online by manually checking the website. Column 4 shows the estimate for the total overlap in product selection. Only a subset of retailers in each country are included. These numbers are lower-bound estimates for the overlap because some of our manual checks took place several days after the original offline data was collected. Results by retailer are provided in the online Appendix.

and a significant online markup (either positive or negative). There are some examples in Argentina, Brazil, Japan, and the United States. These patterns may reflect a desire to compensate for shipping costs or price-discriminate online consumers.

Whether each kind of retailer is useful as a source of data depends on the purpose of the paper or application. For example, using online prices for the retailer in Argentina where 79 percent of prices are higher online is not a problem for measuring inflation if the online markup is relatively constant over time, but it would bias the results if we were interested in comparing price-level differences across countries. Unless a correction is applied, the online data would make prices in Argentina appear higher than what they are. Identifying these special patterns and correcting for any biases is particularly important in papers or applications that use online data from one (or a few) retailers.

VII. Conclusions

This paper shows that in large, multi-channel retailers there is little difference between the online price collected from a website and the offline price obtained by visiting the physical store. Prices are identical about 72 percent of the time, and while price changes are not synchronized, they have similar frequencies and sizes. At the same time, there is considerable heterogeneity across countries, sectors, and retailers.

For research economists using online data for macro and international research questions, my results provide evidence that online prices are a representative source of retail prices, even if most transactions still take place offline. At a more micro level, the differences in behaviors can be used to better model the pricing dynamics and strategies of different types of retailers in various sectors and countries. This high degree of heterogeneity also implies that papers that use relatively few sources of data should be cautious to understand relevant pricing patterns and control for any potential sampling biases.

For NSOs considering the use of online data for consumer price indexes, my results show that the web can be effectively used as an alternative data-collection technology for multi-channel retailers. Particularly for products such as electronics or clothing, the price collected on the web will tend to be identical to the one that can be obtained by walking into a physical store. Online prices are not only cheaper to collect, but they also provide information for all goods sold by each retailer, with many details per product, uncensored price spells, and can be collected on a high-frequency basis without any delays. Of course, there are also many potential disadvantages of using online data, including limited sector coverage and the lack of information on quantities, as we discuss in Cavallo and Rigobon (2016). But from a data-collection perspective, my results suggest that the online-offline price differences should not be a major source of concern.

For those interested in the effect of the Internet on retail prices, my findings imply little within-retailer price dispersion, both online and offline. While the Internet may not have reduced dispersion across retailers, it seems to have created the incentives for companies to price identically across their own physical and online stores. More research is needed to understand the mechanisms that drive this effect. One possibility is that retailers are worried about antagonizing customers who can now easily compare prices online through the web or their mobile phones. This might even be affecting cross-country pricing, as suggested by Cavallo, Neiman, and Rigobon (2014), where we found evidence that global firms such as Apple and IKEA tend to price identically in countries that use the same currency, where it is trivial for consumers to compare prices across borders.

Future work should also try to understand why there are still some observed price-level differences. One explanation may be that online prices adjust faster to shocks. That would be consistent with the unsynchronized price change results in this paper and the anticipation in online price indices documented in Cavallo and Rigobon (2016). Another potential reason is that location-specific sales or offline price dispersion may play a larger role than I can detect in these data. In particular, the offline price comparisons for multiple zip codes in Section IVB could be expanded to cover more sectors and countries. In addition, good-level characteristics, such as the bargaining power of the manufacturer or the nature of its production and distribution costs, may help explain why some goods have identical prices while others do not.

Another limitation of my analysis is the lack of quantity information at the product and retailer levels. For some applications, such as the computations of price indices, we can use category weights in official CPI data. But other pricing statistics may change considerably when individual goods are weighted by transactions, as shown with online book sales by Chevalier and Goolsbee (2003). Future research should try to combine online prices with other micro data, such as scanner datasets, that can provide more detailed quantity information.

Finally, except for the Amazon results in Section IVC, this paper does not study the prices of online-only retailers or small companies that participate in online marketplaces. If their share of retail transactions continues to grow, a large-scale comparison with traditional multi-channel retailers will be needed to better understand how pricing strategies and dynamics are likely to evolve in the future.

REFERENCES

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi. 2016. "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach." *American Economic Review* 106 (10): 2817–51.

Balakrishnan, Anantaram, Shankar Sundaresan, and Bo Zhang. 2013. "Browse-and-Switch: Retail-Online Competition under Value Uncertainty." *Production and Operations Management* 23 (7): 1129–45.

Baye, Michael R., John Morgan, and Patrick Scholten. 2006. "Information, Search, and Price Dispersion." In *Economics and Information Systems*, edited by Terrence Hendershott, 323–76. Amsterdam: Elsevier.

Boettcher, Ingolf. 2015. "Automatic Data Collection on the Internet (Web Scraping)." *Statistics Austria.*

Borraz, Fernando, Alberto Cavallo, Roberto Rigobon, and Leandro Zipitria. 2015. "Distance and Political Boundaries: Estimating Border Effects under Inequality Constraints." *International Journal of Finance & Economics* 21 (1): 3–35.

Breton, Robert, Tanya Flower, Matthew Mayhew, Elizabeth Metcalfe, Natasha Milliken, Christopher Payne, Thomas Smith, Joe Winton, and Ainslie Woods. 2015. *Research Indices Using Web Scraped Data*. Office for National Statistics, UK.

Brynjolfsson, Erik, Astrid Andrea Dick, and Michael D. Smith. 2009. "A Nearly Perfect Market." *Quantitative Marketing and Economics* 8 (1): 1–33.

Brynjolfsson, Erik, Yu Jeffrey Hu, and Mohammad S. Rahman. 2013. "Competing in the Age of Omnichannel Retailing." *MIT Sloan Management Review* 54 (4): 152–65.

Brynjolfsson, Erik, Yu Jeffrey Hu, and Duncan Simester. 2011. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales." *Management Science* 57 (8): 1373–86.

Brynjolfsson, Erik, and Michael D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46 (4): 563–85.

Brynjolfsson, Erik, and Michael D. Smith. 2001. "The Great Equalizer? Consumer Choice Behavior at Internet Shopbots." Massachusetts Institute of Technology Sloan Working Paper 4208-01.

Cavallo, Alberto. 2013. "Online and Official Price Indexes: Measuring Argentina's Inflation." *Journal of Monetary Economics* 60 (2): 152–65.

Cavallo, Alberto. Forthcoming. "Scraped Data and Sticky Prices." *Review of Economics and Statistics*.

Cavallo, Alberto. 2017. "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers: Dataset." *American Economic Review*. https://doi.org/10.1257/aer.20160542.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon. 2014. "Currency Unions, Product Introductions, and the Real Exchange Rate." *Quarterly Journal of Economics* 129 (2): 529–95.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon. 2015. "The Price Impact of Joining a Currency Union: Evidence from Latvia." *IMF Economic Review* 63 (2): 281–97.

Cavallo, Alberto, and Roberto Rigobon. 2016. "The Billion Prices Project: Using Online Prices for Measurement and Research." *Journal of Economic Perspectives* 30 (2): 151–78.

Chevalier, Judith, and Austan Goolsbee. 2003. "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com." *Quantitative Marketing and Economics* 1 (2): 203–22.

Clay, Karen, Ramayya Krishnan, Eric Wolff, and Danny Fernandes. 2002. "Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry." *Journal of Industrial Economics* 50 (3): 351–67.

Clemons, Eric. K., Il-Horn Hann, and Lorin M. Hitt. 2002. "Price Dispersion and Differentiation in Online Travel: An Empirical Investigation." *Management Science* 48 (4): 534–49.

CNN. 2000. "Amazon Pricing Flap: Web Retailer Apologizes for Price Test, Refunds Money to Customers," September 28. http://money.cnn.com/2000/09/28/technology/amazon/index.htm.

Edelman, Benjamin. 2012. "Using Internet Data for Economic Research." *Journal of Economic Perspectives* 26 (2): 189–206.

Ellison, Glenn, and Sara Fisher Ellison. 2009a. "Search, Obfuscation, and Price Elasticities on the Internet." *Econometrica* 77 (2): 427–52.

Ellison, Glenn, and Sara Fisher Ellison. 2009b. "Tax Sensitivity and Home State Preferences in Internet Purchasing." *American Economic Journal: Economic Policy* 1 (2): 53–71.

Euromonitor International. 2014. *Internet vs Store-based Shopping: The Global Move Towards Omnichannel Retailing.* London: Euromonitor International.

Fitzgerald, Drew. 2013. "Fear of 'Showrooming' Fades: Best Buy, Other Retailers Are Optimistic Price-Matching Can Stanch Trend." *Wall Street Journal*, November 3. http://www.wsj.com/articles/SB10001424052702303661404579175690690126298.

Gorodnichenko, Yuriy, Viacheslav Sheremirov, and Oleksandr Talavera. 2014. "Price Setting in Online Markets: Does IT Click?" National Bureau of Economic Research Working Paper 20819.

Gorodnichenko, Yuriy, and Oleksandr Talavera. 2017. "Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration." *American Economic Review* 107 (1): 249–82.

Griffioen, Robert, Jan de Haan, and Leon Willenborg. 2014. *Collecting Clothing Data from the Internet.* The Hague: Statistics Netherlands.

Horrigan, Michael W. 2013. "Big Data: A Perspective from the BLS." *Amstat News*, January 1. http://magazine.amstat.org/blog/2013/01/01/sci-policy-jan2013/.

Kaplan, Greg, and Guido Menzio. 2015. "The Morphology of Price Dispersion." *International Economic Review* 56 (4): 1165–205.

Klena, Kali, and Jill Puleri. 2013. *From Transactions to Relationships: Connecting with the Transitioning Shopper: Australian Insights from the 2013 Smarter Consumer Study.* St. Leonards: IBM Institute for Business Value.

Krsinich, Frances. 2015. "Price Indexes from Online Data Using the Fixed-Effects Window-Splice (FEWS) Index." Paper presented at Ottawa Group Conference, Tokyo, Japan.

Krsinich, Frances. 2016. "The FEWS Index: Fixed Effects with a Window Splice." *Journal of Official Statistics* 32 (2): 375–404.

Lunnemann, Patrick, and Ladislav Wintr. 2011. "Price Stickiness in the US and Europe Revisited: Evidence from Internet Prices." *Oxford Bulletin of Economics and Statistics* 73 (5): 593–621.

Mattioli, Dana. 2012. "On Orbitz, Mac Users Steered to Pricier Hotels." *Wall Street Journal*, August 23. http://www.wsj.com/articles/SB10001424052702304458604577488822667325882.

Mikians, Jakub, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. "Detecting Price and Search Discrimination on the Internet." *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*: 79–84. New York: Association for Computing Machinery.

Mikians, Jakub, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2013. "Crowd-Assisted Search for Price Discrimination in E-commerce: First Results." *Proceedings of the 9th ACM Conference on Emerging Networking Experiments and Technologies*: 1–6. New York: Association for Computing Machinery.

Nakamura, Emi, and Jon Steinsson. 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics* 123 (4): 1415–64.

Nygaard, Ragnhild. 2015. "The Use of Online Prices in the Norwegian Consumer Price Index." Paper presented at the 14th Meeting of the Ottawa Group, Tokyo, Japan. Statistics Norway.

OECD. 1998. "Electronic Commerce: Price and Consumer Issues for Three Products: Books, Compact Discs, and Software." Organization for Economic Co-operation and Development Digital Economy Paper 32.

Quan, Thomas W., and Kevin R. Williams. 2015. "Product Variety, Across-Market Demand Heterogeneity and the Value of Online Retail." http://www.econ.umn.edu/~quanx039/quanJMP.pdf (accessed August 5, 2015).

Simonovska, Ina. 2015. "Income Differences and Prices of Tradables: Insights from an Online Retailer." *Review of Economic Studies* 82 (4): 1612–56.

Tang, Fang-Fang, and Xiaolin Xing. 2001. "Will the Growth of Multichannel Retailing Diminish the Pricing Efficiency of the Web?" *Journal of Retailing* 77 (3): 319–33.

Valentino-DeVries, Jennifer, Jeremy Singer-Vine, and Ashkan Soltani. 2012. "Websites Vary Prices, Deals Based on Users' Information." *Wall Street Journal*, December 24. http://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

Xing, Xiaolin, Zhenlin Yang, and Fang-Fang Tang. 2006. "A Comparison of Time-Varying Online Price and Price Dispersion between Multichannel and Dotcom DVD Retailers." *Journal of Interactive Marketing* 20 (2): 3–20.