Digitization Research Proposal

NBER Productivity, Innovation and Entrepreneurship Program

Differences between online and offline prices

By Alberto Cavallo

Objective

The objective of this research is to compare online and offline prices collected simultaneously (same day) at some of the largest US retailers. The results will be used to document the differences and discuss the impact that the internet is having on these large firm's pricing behaviors.

Background information

Online prices are an increasingly popular source of micro-level data.[1] Although data collected from these new online sources has potentially many advantages for research, a common concern is that online transactions are still a tiny share of all sales, even in the US (currently less than 10%). Are online prices identical to offline prices? If not, what are the differences and why? Answering these types of questions is important not only to "validate" online data as a source of micro-prices, but also to help use better understand how firms' pricing behaviors are affected by the existence of the web and mobile browsing/price-checking technologies.

Unfortunately, the differences between online and offline data are not well understood simply because data has never been collected simultaneously for the same goods and retailers in a large scale.[2] My work with the Billion Prices Project at MIT has given access to a huge amount of *online* prices from the major retailers in the world, but getting *offline* prices is far more costly and complicated. In 2012 I started experimenting with more efficient offline data collection methods for large scale comparisons. I sent MIT students to stores around Boston to scan prices with their phones, but the scale of these efforts was limited in size and retailers often stopped us after noticing that the students spent considerable time at the stores. Since then, we have streamlining the process to make it as simple and cheap as possible. We developed a new android app and recently started running a series of tests using crowdsourcing platforms. The data we collected so far is useful to answer simple price-level questions (such as whether prices are identical, or by how much they might differ), but our goal is now to obtain a large panel of

[1] See Brynjolfsson & Smith (2000) for an early paper using both online and offline data, Cavallo (2013) for the use of online prices for inflation measurement, Cavallo et al (2014) for international price comparisons, and Gorodnickenko et al (2014) for online pricing dynamics using data from price comparison websites.

[2] Cavallo (2012) contains a small-scale exercise for some Latin American supermarkets, but it was limited in the number of goods, dates, and retailers being sampled.

prices (i.e. same set of barcodes over time), which we could use to study price changes, sales, availability, and other pricing behaviors over time.

The funding requested in this proposal will be used specifically to build this panel database: with weekly prices collected during 18 weeks for a set of 3,750 goods in the 10 largest US retailers that sell both online and offline. We are ready to start in March 2015. The data collected for this project will be used to write a paper answering some of the questions above, and also will be shared publicly on the BPP website (bpp.mit.edu).

Data Collection Approach

To make offline data collection as simple and cheap as possible, we will combine the use of a custom-made barcode-scanning app for android phones with the growing popularity of crowdsourcing web services such as Amazon Mechanical Turk and Elance.

The app provides a simple way to collect data with the format we need. The crowdsourcing platforms allow us to hire hundreds of independent "workers" at a very low cost to do the data collection in a large number of retailers and locations all over the US. More details on the app and the crowdsourcing platforms are provided at the end of this document.

The data obtained offline will be compared with the daily data we currently collect as part of the Billion Prices Project at MIT. Custom scrape-jobs will use used to ensure that all the products randomly chosen offline are also sampled online on the same date.

Funding Request:

I am requesting a total of $14,850. This would cover the cost of hiring 250 people collecting prices at 10 retailers, with 15 items per weekly visit during a period of 18 weeks.

The attached budget provides all the details, but we basically estimate a cost per store visit ($3), the total cost person ($3x18week=$54), and the total cost of the project ($54x 10 Retailers x 25 workers per retailer= $13,500). With Amazon's commission (10% over funds spent) the total amount needed is $14,850.

BUDGET

Offline Data Collection Costs -250 people, 10 Retailers, $54 per person, each with 18 visits to a store and collecting 15 prices per visit. Estimated hourly wage is $3.	**$14,850**

DETAILS

Budget Proposal - Panel Data of Offline Prices in the US
Dates: March 1st to July 4th (18 Weeks)

COST PER PERSON - RETAILER	
# Items per Person per Visit	15
Estimated Time per Item per Visit	4 m
Estimated Time per Visit	1 h
# Visits to the Same Store	18
Price per visit (hourly wage)	$3.00
Price Per Person and Store	$54.00

TOTAL COST	
Number of Retailers Monitored	10
Number of Workers at Each Retailer	25
Total Workers	250
Total paid to workers	$13,500
Total Amazon Mechanical Turk Commission (10%)	$1,350
TOTAL COST	$14,850

Additional Stats	
# of goods monitored	3,750
Total # of price observations collected	67,500

Additional Details on the Tools used

Crowdsourcing

Platforms such as Amazon Mechanical Turk (AMT) and Elance are web services that allow researchers to post simple jobs to be completed by registered users. These sites have several advantages for our effort. First, the cost is low (based on our previous tests we estimate paying only $3 per visit to a store). Second, we can hire hundreds of workers from cities all over the US. Third, these sites handle individual payments to each user.

The plan is to post simple jobs to collect data at a physical store of one of the specific retailers selected by us during the course of 18 weeks. We will ask workers to download a special scanning app we developed (see below), and use it to scan 15 products each time they visit a store. Workers will get paid every week after they submit a new batch of prices, with a bonus payment at the end if they complete 18 weeks.

We will hire 25 different workers for each retailer in our list. This will ensure we get a good coverage across categories of goods, locations, and quality of. Once we receive the information via email, an RA will check the data and pay the worker via AMT or Elance.

BPP Scanning App

Being able to "hire" hundreds of people is part of the solution, but we also need them to be able to perform a task in a simple, quick, and standardized way. To do this, we have developed a special barcode-scanning app for android phones. Our "workers" can download the app from the Google Play Store for free. They will use this app to collect the data needed to complete the HIT. The app is customized to our needs, makes it very efficient to scan and input data, and ensures that all the output we get in the end is standardized (no matter who collected the data or where).

The app functionality is simple. When the person is at the store, in front of a price tag, she simply has to focus the phone's camera to scan the barcode of the price tag (or enter the code manually if it fails), manually input the price, and take a picture of the tag (for validation purposes later on by our team). The app automatically adds a date, time and device ID stamp. The final step allows the user to email the excel file generated by the app to a pre-specified address. We automatically save all files received in a folder and merge the data in a database file for later analysis.

References

Brynjolfsson, E., Smith, M.D., 2000. Frictionless commerce? A comparison of Internet and conventional retailers. Management Science 46, 563–585.

Cavallo, A., 2012. Scraped Data and Sticky Prices. MIT Sloan Working Paper 4976-12.

Cavallo, A., 2013. Online and Official Price Indexes: Measuring Argentina's Inflation. Journal of Monetary Economics 60, 152–165.

Cavallo, A., Neiman, B., Rigobon, R., 2014. Currency Unions, Product Introductions, and the Real Exchange Rate. Quarterly Journal of Economics 129.

Gorodnichenko, Y., Sheremirov, V., Talavera, O., 2014. Price Setting in Online Markets: Does IT Click? National Bureau of Economic Research.