Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce

By AUSTAN D. GOOLSBEE AND PETER J. KLENOW *

The e-commerce share of retail spending in the United States has almost tripled in the last 10 years and stands at close to 10% overall and more than 50% in several major categories according to U.S. Census Bureau (2018). If online pricing or inflation is fundamentally different than traditional retail, it could have a rising impact on the overall Consumer Price Index (CPI) or bias in it.[1]

We use *Adobe Analytics* data on online transactions for millions of products in many different categories from 2014 to 2017 to shed light on how online inflation compares to overall inflation, and to gauge the magnitude of new product bias online. The Adobe data is similar to the Billion Prices Project of Cavallo and Rigobon (2016), which scrapes list prices from the web, except the Adobe data contains actual transaction prices and includes quantities purchased.

We follow two literatures. One uses detailed scanner data from grocery stores to analyze new product introductions, such as Broda and Weinstein (2010). Another studies consumer surplus from the internet and e-commerce in particular – e.g., Brynjolfsson, Hu and Smith (2003), Goolsbee and Klenow (2006), Brynjolfsson and Oh (2012), and Varian (2013).

We document lower inflation online than in the CPI – 1.3 percentage points lower inflation per year– for the same categories. The data also show that the entry of new products and the exit of old products is extremely important for most categories of goods. In a CES (Constant Elasticity of Substitution) model with a standard model elasticity, the net entry of new goods during the sample implies that matched-model price indices overstate true inflation by an additional 1.5 to 2.5 percentage points per year.

I. Adobe Data

Adobe Analytics provides a variety of services to e-commerce merchants who share their transaction data for Adobe to analyze. Adobe clients include 20 of the 30 largest employers in the nation and 80% of Fortune 500 retailers.

The underlying data are quantities and revenue from individual transactions (not including taxes or shipping costs). The product codes are merchant-specific so our definition of a good will be the good-merchant combination. We use Adobe's data aggregated up to the monthly level: total quantity and average transaction price for each product for a given month. Adobe anonymizes the data so we cannot identify any retailers or customers.

We use a subset of the products and merchants from the full Adobe data set. This subset covers categories making up about 19% of the CPI relative importance weights in Bureau of Labor Statistics (2018), and the revenue amounts to about 15% of all e-commerce in U.S. Census Bureau (2018). The sample goes from January 2014 through September 2017.

Table 1 shows the number of products in the Adobe data overall and by by CPI Major Group, averaged over the January 2014 to September 2017 period. The dataset contains over 2 million products in the average month, vs. about 140,000 per month in the entire CPI. There are 211 CPI categories known as Entry Level Items (ELIs), and the Adobe data covers 65 of them.[2]

* Goolsbee: University of Chicago Booth School of Business, Chicago IL 60637, and NBER, goolsbee@chicagobooth.edu. Klenow: Department of Economics, Stanford University, Stanford, CA 94305, and NBER, klenow@stanford.edu. Mathias Jimenez and Matteo Leombroni provided superb research assistance; Alberto Cavallo, Charles Hulten, and Leonard Nakamura helpful comments; and Luis Maykot and Siddharth Kulkarni assistance in accessing and understanding the Adobe data.

[1] Economists have long known about the potential for new products to bias upward the inflation measured in the CPI. See Boskin et al. (1996) and, more recently, Groshen et al. (2017). Recent business press articles have argued that online commerce may be leading to growing problems in measuring inflation and complicating monetary policy decisions at the Federal Reserve, as in Cohan (2017), Torrey and Stevens (2017) and Gross (2017).

[2] The Adobe dataset covers home furnishings (appliances, furniture, etc.) but not rent or owner's equivalent rent, hence the label "Household Goods." Similarly, we refer to "Information

TABLE 1—ADOBE CPI COVERAGE AND NUMBER OF PRODUCTS BY CATEGORY

	CPI Coverage (percent)	# of products
Headline	**19**	**2.1 M**
Food and beverages	49	1 M
Housing	7	90 K
Apparel	100	128 K
Education and communication	9	400 K
Medical care	9	23 K
Transportation	3	125 K
Recreation	32	200 K
Other goods and services	42	90 K

Note: Average number of products from 2014 through 2017. Headline is all covered categories. The next rows are CPI Major Groups. The last columns gives the monthly average percentage of ELI coverage for the respective category.
Source: Authors' calculations using Adobe Analytics and BLS (Bureau of Labor Statistics) data.

II. DPI vs. CPI Inflation

We call the matched model price index we construct from the Adobe data the Digital Price Index (DPI) to distinguish it from the CPI. We start with price changes for overlapping products in months $t-1$ and t. These are products selling positive quantities in both months. We take log first difference of average unit prices. To aggregate price changes across products within an ELI, we use Tornqvist weights. These are the average spending share of the product in the ELI in months t-1 and t. The spending shares are based on Adobe data for overlapping products.[3]

To facilitate comparison with the CPI, we aggregate the Adobe ELI inflation rates using Laspeyres weights. In our base case, we use CPI relative importance weights for each ELI-month. We use the same set of ELIs to construct both the CPI and the DPI. In this way we can rule out that differences between the two indexes arise from categories which are not covered by the DPI or the weighting of categories covered.

We plot the two indices together in Figure 1. The DPI exhibits notably more deflation over the period than the CPI for the same categories. Table 2 shows the average annual inflation rates from 2014–2017. Overall (headline) DPI inflation is more than 1 percentage point per year lower than CPI inflation. Breaking out by Major Groups, inflation is lower in the DPI than in the CPI in every category other than medicine & medical supplies.

Now, excess deflation in high frequency, chain-weighted price indices can result from oscillating prices due to recurring discounts. This phenomenon is known as "chain drift." Even if the prices and quantities revert to their starting levels, a chained price index may not revert to 1. This has been documented in grocery store scanner data by, for example, de Haan and Van der Grient (2011).

To gauge chain drift in the Adobe data, in each year we tried adding an artificial "13th month" with the first months prices and quantities and asked whether the price index a returns to 1. We found no evidence of such chain drift. In fact, the bias went the other way – with the index moving above 1 in the 13th month. Thus, chain drift does not seem to be a source of deflationary bias in the DPI relative to the CPI.

III. Product Entry and Exit

Because the Adobe data include quantities as well as prices, we are able to look at spending on entering and exit products. The CPI does not have quantities for items sold within ELIs, so it cannot tell the market share of exiting products. And, because it does not sample all products at a given merchant, the BLS cannot assess product entry within merchants. The AC Nielsen scanner data also contains quantities sold, but

Technology" as the goods within Education and Communication that the Adobe data covers.

[3]Like the BLS, we do something special for apparel. We construct a simple index of average unit prices. This is to avoid extreme deflation from fashion and seasonal cycles for clothing.

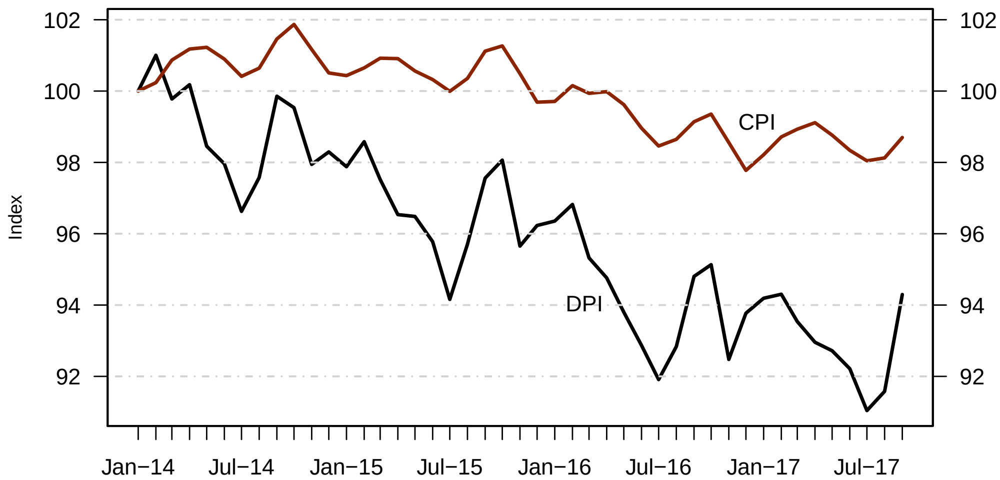

FIGURE 1. CUMULATIVE INFLATION, DPI VS. CPI.

Note: For the 65 ELIs covered by the Adobe Digital Price Index (DPI). CPI relative importance weights for each ELI.
Source: Authors' calculations using Adobe Analytics and BLS data.

this dataset is heavily tilted toward food and beverages in grocery stores – see Kaplan and Schulhofer-Wohl (2016). The Adobe data allow us to quantify the importance of new varieties outside of grocery stores.

We classify a product as new if the product-merchant combination did not exist in the data in the previous calendar year, and as exiting if it does not appear in the following calendar year. We present the entry and exit rates of products by category, weighting by sales of each product in Table 3.[4] In apparel, fashion and seasonal cycles depress sales of outgoing products and inflate sales of new products. We therefore report results with and without apparel.

As shown in Table 3, roughly half of the sales volume online is for products that did not exist in the previous year. Even without apparel, the figure is 44%. The products that disappear, meanwhile, had about 24% of total sales before they left the market (22% excluding apparel). Note that, if all that was happening in the data was relabeling of the same products each year, then we would expect both rates to be inflated by equal amounts. Such relabeling therefore cannot explain the high share of entering relative to exiting products. Importantly, in the online data the food and beverage category shows much less dynamism than other categories – an entry rate only around one third that that for the full universe of products and an exit rate around 2/5ths. Thus previous studies focusing on grocery store items, such as Broda and Weinstein (2010), may even understate the importance of new products.[5]

IV. The Impact of New Products on Inflation

Feenstra (1994) showed that a direct way to gauge the importance of new products in a CES framework is to look at the growth rate of overall spending in a category minus the growth rate of spending for products that exist in both time periods. The higher this net growth rate, the lower the true inflation rate relative to the matched model inflation rate. As shown in Table 3, entering products do tend to have significantly bigger market shares than outgoing products in the Adobe data, even outside apparel.[6]

Feenstra (1994) showed, further, that the reduction in true inflation equals the net growth in spending on new varieties times $1/(\sigma - 1)$,

[4]We weight by the average monthly sales of a product during the calendar year across the months the product was available.

[5]Based on CPI product counts, Bils and Klenow (2004) also report a markedly lower exit rate for food items than other items.

[6]This should capture improvements in product quality in addition to brand new types of products, because both are associated with new product ID codes in the Adobe data.

TABLE 2—AVERAGE ANNUAL INFLATION

	DPI	CPI
Headline	**-1.6**	**-0.3**
Food and beverages	-0.9	0.3
Household goods	-4.9	-1.9
Apparel	0.3	0.8
Information technology	-6.1	-3.7
Medicines and medical supplies	1.2	-0.2
Transportation accessories and parts	-1.3	-0.4
Recreation goods	-6.2	-3.0
Other goods and services	0.7	1.7

Note: Percentage points per year in annual average inflation for 2014-2017.

Source: Authors' calculation using Adobe Analytics and BLS data.

TABLE 3—ADOBE ENTRY AND EXIT RATES

	Entry	Exit
Headline	**51.4**	**24.3**
Headline ex. Apparel	**43.7**	**21.9**
Food and beverages	15.7	9.2
Household goods	30.5	19.0
Apparel	70.8	30.3
Information technology	60.8	31.7
Medicines and medical supplies	11.1	7.6
Transportation accessories and parts	24.6	16.9
Recreation goods	61.1	20.7
Other goods and services	49.9	13.4

Note: Sales-weighted within ELIs. CPI relative importance weights across ELIs. Average % points for 2014-2015 and 2015-2016.

Source: Authors' calculation using Adobe Analytics and BLS data

where σ is the elasticity of substitution between varieties. We use a baseline value of $\sigma = 4$ based on Hottman, Redding and Weinstein (2016). We also consider a higher value of $\sigma = 6$ for robustness – a more conservative value given new varieties are less valuable if they are closer substitutes for existing varieties.

Table 4 presents estimates of new goods bias in the Adobe online data. The arrival of new goods is equivalent to 1.5 to 2.5 percentage points lower inflation than what a matched-model would indicate.[7] This is much higher than the 0.6% per year new product bias estimated by the Boskin Commission, though that was for the CPI as a whole. The Adobe data may cover items with larger-than-average new goods bias.

The vital role for new goods in the Adobe online data calls for more research on new varieties in traditional retail, preferably outside of just the food category.[8] If offline sales are similar to online sales, as suggested by Cavallo (2017), new products may be even more important than previously thought.

Combining the two points here, the Adobe DPI inflation rate – adjusted for new goods – is more than 3 percentage points per year lower than the CPI inflation rate for the same categories from 2014–2017.

REFERENCES

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li. 2017. "Missing growth from creative destruction." National Bureau of Economic Research.

[7]We exclude apparel because incoming items may sell a lot more than outgoing items purely due to seasonal/fashion cycles.

[8]E.g., Aghion et al. (2017) study the nonfarm business sector.

TABLE 4—NEW GOODS BIAS BASED ON THE ADOBE DATA

	$\sigma = 4$	$\sigma = 6$
Headline	**3.5**	**2.1**
Headline ex. Apparel	**2.5**	**1.5**
Food and beverages	0.4	0.2
Household goods	0.9	0.5
Apparel	7.3	4.4
Information technology	4.1	2.5
Medicines and medical supplies	0.0	0.0
Transportation accessories and parts	0.7	0.4
Recreation goods	5.4	3.2
Other goods and services	5.9	3.9

Note: Entries are percentage points per year, averaged over 2014-2015 and 2015-2016.

Source: Authors' calculation using Adobe Analytics and BLS data.

Bils, Mark, and Peter J Klenow. 2004. "Some evidence on the importance of sticky prices." *Journal of political economy*, 112(5): 947–985.

Boskin, Michael J, Ellen R Dulberger, Robert J Gordon, Zvi Griliches, and Dale W Jorgenson. 1996. *Toward a more accurate measure of the cost of living: Final report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index.* Advisory Commission to Study the Consumer Price Index.

Broda, Christian, and David E Weinstein. 2010. "Product creation and destruction: Evidence and price implications." *The American economic review*, 100(3): 691–723.

Brynjolfsson, Erik, and JooHee Oh. 2012. "The attention economy: measuring the value of free digital services on the Internet."

Brynjolfsson, Erik, Yu Hu, and Michael D Smith. 2003. "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers." *Management Science*, 49(11): 1580–1596.

Bureau of Labor Statistics. 2018. *https://www.bls.gov/cpi/tables/supplemental-files/home.htm.*

Cavallo, Alberto. 2017. "Are online and offline prices similar? Evidence from large multi-channel retailers." *The American Economic Review*, 107(1): 283–303.

Cavallo, Alberto, and Roberto Rigobon. 2016. "The billion prices project: Using online prices for measurement and research." *The Journal of Economic Perspectives*, 30(2): 151–178.

Cohan, Peter. 2017. "How Amazon Overpowers the Fed." *Forbes*.

de Haan, Jan, and Heymerik A Van der Grient. 2011. "Eliminating chain drift in price indexes based on scanner data." *Journal of Econometrics*, 161(1): 36–46.

Feenstra, Robert C. 1994. "New product varieties and the measurement of international prices." *The American Economic Review*, 157–177.

Goolsbee, Austan, and Peter J Klenow. 2006. "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet." *American Economic Review*, 96(2): 108–113.

Groshen, Erica L, Brian C Moyer, Ana M Aizcorbe, Ralph Bradley, and David M Friedman. 2017. "How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches." *Journal of Economic Perspectives*, 31(2): 187–210.

Gross, Daniel. 2017. "How Amazon Fuels Deflation." *Strategy+Business*.

Hottman, Colin J, Stephen J Redding, and David E Weinstein. 2016. "Quantifying the sources of firm heterogeneity." *The Quarterly Journal of Economics*, 131(3): 1291–1364.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2016. "Inflation at the Household Level." National Bureau of Economic Research.

Torrey, Harriet, and Laura Stevens. 2017. "As the Fed Deliberates Amazon is Making it's Job More Difficult." *Wall Street Journal*.

U.S. Census Bureau. 2018. https://www.census.gov/retail/index.html.

Varian, Hal. 2013. "The value of the internet now and in the future." *The Economist*.