The Pricing Strategies of Online Grocery Retailers *

Diego Aparicio, *IESE Business School*
Zachary Metzman, *MIT*
Roberto Rigobon, *MIT Sloan and NBER*

December 2020

Abstract

Matched product data is collected from the leading online grocers in the U.S. The same products are identified in scanner data. The paper documents pricing strategies within and across online (and offline) retailers. First, online retailers exhibit less uniform pricing than offline retailers. Second, online price dispersion across competing chains in narrow geographies is higher than offline retailers. Third, variation in offline elasticities, shipping distance, pricing frequency, and local demographics are utilized to explain price differentiation. Surprisingly, pricing flexibility (across time) magnifies price differentiation (across locations). This evidence motivates a high-frequency study to recover the patterns of algorithmic pricing. Online grocery retailers change prices very frequently and in small magnitudes, have lower menu costs, are less synchronized, constantly explore the price grid, and often match competitors' prices.

Keywords: price dispersion, algorithmic pricing, online grocers, pricing technology
JEL: E30, L81, M31

*Data and codes to reproduce the results will be publicly available. Aparicio: daparicio AT iese.edu; IESE Business School, Marketing; Metzman: zmetzman AT mit.edu; MIT, Electrical Engineering and Computer Science; Rigobon: rigobon AT mit.edu; MIT Sloan, Management and NBER. Nestor Santiago Perez provided outstanding research assistance. All errors are our own. Authors' own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the authors and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

The Internet has reduced the barriers to search, allowing consumers to explore products and prices across platforms at a lower cost (Bakos (1997); Brown and Goolsbee (2002); Ellison and Ellison (2009)). One would therefore imagine that this has led to greater price transparency and price convergence, both within a retail chain and across competing chains. However, the Internet also fostered information and communications technologies (Brynjolfsson and McAfee (2014); Acemoglu and Restrepo (2018); Forman and Goldfarb (2020)) that exploit customization opportunities. In fact, a recent technology referred to as algorithmic pricing, in which computer programs are constantly trained to optimize prices, allows firms to introduce remarkable flexibility in price setting.

Online groceries represent a meaningful part of the CPI expenditures and of the U.S. economy. In 2019, the U.S. retail e-commerce industry reached $600 billion in sales; and despite representing 11% of total retail sales, it grows at an annual rate of 16%, compared to 3% in offline sales (U.S. Census Bureau (2020)). Even faster growth, close to 20%, is taking place in online groceries (New York Times (2018)). The COVID-19 pandemic exacerbated this trend: online groceries reached record sales in May 2020, increasing 450% with respect to August 2019 (Financial Times (2020)).

In this paper, we document that online grocery retailers follow pricing strategies that question the price convergence intuition. Features that signal advances in pricing technology account for a larger online price dispersion. That pricing technology magnifies price dispersion is surprising: algorithmic pricing is typically associated with high-frequency price changes and does not imply anything for price differentiation across consumers, making purchase decisions in different locations for the same products *in a given point in time*.

We make a methodological contribution to study pricing strategies in the context of online groceries, which can be summarized as follows. First, we collect price data from the leading U.S. online grocery retailers; critically, data for a given product is collected at (nearly) the same time across retailers and across locations. Additionally, we collect price data intra-day to capture the patterns of algorithmic pricing. Second, online products are carefully matched with Nielsen's scanner data. This allows us to study the three key dimensions of price setting between online and offline grocers: pricing across locations, pricing across retailers, and pricing across time.

We begin the paper by documenting that online price dispersion is larger than offline price dispersion. In particular, we show that online grocers have higher measures of non-uniform pricing within the chain and across locations; our estimates of offline price dispersion closely follow DellaVigna and Gentzkow (2019).[1] We also show that online price dispersion, across competing chains and within a delivery zipcode, is higher than the offline equivalent specification. These results indicate that internet price competition is far from achieving price convergence in supermarket products, or at least not like we have witnessed in electronics, durables, ride sharing, or flights.

We decompose price dispersion in relative prices between chain and location effects. We find that over half of the price variation is explained by chain effects, but there is a meaningful residual explained by the retailer-zipcode. In contrast, the retailer-store effect in offline data accounts for a small portion of price variation (Nakamura (2008);

[1] Recent studies related to uniform pricing include, for example, Nakamura (2008); DellaVigna and Gentzkow (2019); Hitsch, Hortacsu and Lin (2019) using scanner data, and Cavallo, Neiman and Rigobon (2014); Aparicio and Cavallo (2017); Cavallo (2018*a*) using online data. See also Orbach and Einav (2007); Aparicio and Rigobon (2020) for uniform pricing across differentiated goods.

DellaVigna and Gentzkow (2019); Hitsch, Hortacsu and Lin (2019)). We then proceed to understand variation in relative prices within a chain. We estimate offline elasticities (for the same product and city), and find that offline elasticities are informative for offline price dispersion and, importantly, for online price dispersion. Intuitively, while we lack online demand data, the variation in offline elasticities is arguably informative about the variation in online elasticities. We also find that variation in shipping costs explain price differentiation across locations. Interestingly, local demographics are not a critical determinant. Perhaps most surprisingly is that pricing technology explains variation in prices across locations. Said differently, a lower price stickiness in a given delivery zipcode amplifies the price differentiation, for the same product and time, between two delivery zipcodes.

Figure 1 provides a compelling visual perspective. The higher the intensity of algorithmic pricing (as determined by the daily price changes) magnifies the price differentiation, for the same product and time, across two delivery zipcodes.

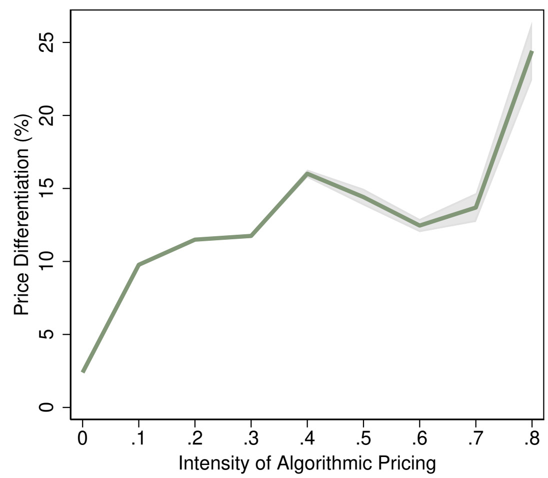


Figure 1: Algorithmic Pricing and Price Differentiation

Notes: Figure shows the average probability of a daily price change and the average price dispersion between two delivery zipcodes. Price dispersion is defined as the percent price difference, for the same product-product-time combination, between all zipcode pairs. The details of the computation are discussed in Section 5. The shaded band indicates a 95% confidence interval.

Taking a step back from the singularities of online groceries, the evidence informs that advances in pricing technology have more implications than commonly assumed. To better understand the scope and patterns of this technology, and more precisely of algorithmic pricing, we collect matched product data in high-frequency intervals for the two leading online grocers in the U.S.

First of all, we document that prices change very frequently and with great flexibility. The probability of a price change intra-day is 7% and in two consecutive days is over 11%. In Amazon, 48% of the products have experienced at least one price change during a week. These estimates suggest that price durations are decreasing considerably, relative to studies using online or offline data in the past decade, and illustrate the rise of algorithmic pricing. Relatedly, the sizes of the price changes are significantly smaller, consistent with the premise that algorithmic pricing overcomes some menu costs: close to 30% of the daily price changes are within 5% in absolute value. In addition to increasing the frequency and lowering the sizes of price fluctuations, the technology allows to expand the price grid. We show that online grocers tend to constantly explore distinct prices. That digital platforms augment the price grid may not be itself surprising, but it does contrast the evidence of "discrete" pricing in retail (Levy et al. (2011); DellaVigna and Gentzkow (2019); Ilut, Valchev and Vincent (2020); Aparicio and Rigobon (2020); Stevens (2020)) and merits further research to understand price setting frictions across channels.

The high-frequency data allows to study synchronization of price changes. Several results are noteworthy. First, synchronization is nearly zero across retailers. In other words, a given retailer-zipcode-hour does not seem more likely to change a price when the competing retailer changes the price for the same zipcode-product, even when looking at 24-hour windows. Second, there is some degree of synchronization within the same retailer, across locations and for the same product, within hours. However, those price changes are often in the *opposite* direction. In contrast, price changes in offline retailers are remarkably synchronized, i.e. stores of the same chain tend to increase (or decrease) prices together. This flexibility in price setting is, once again, another novel scope of algorithmic pricing.

The lack of price convergence or the lack of synchronization across competing retailers might give the impression that retailers optimize prices somewhat in isolation, e.g. their technology is not mindful of competitor prices. This is incorrect. We document that retailers often match each other's price for the same product and delivery zipcode. The patterns of price matching are also interesting. Price matching tends to occur on prices that are on average lower (for both the retailer matching and the retailer being matched). In particular, approximately 83% of the matching events take place on prices that are below the median price. Moreover, price matching is associated with lowering prices in the ballpark of 2%. While this suggestive evidence should not be generalized, it complements Miklós-Thal and Tucker (2019)'s theoretical work that algorithmic pricing can sometimes lead to lower prices and thereby increase consumer surplus.

The rest of the paper is organized as follows. Section 1.1 reviews the literature. Section 2 describes the data and the collection methodology. Section 3 documents facts about online and offline price dispersion. Section 4 explains online price dispersion. Section 5 documents patterns of algorithmic pricing. Section 6 concludes.

1.1 Related Literature

This paper relates to two main bodies of literature. We relate to an abundant empirical literature on supermarket pricing. In the area of price stickiness, see Bils and Klenow (2004); Nakamura and Steinsson (2008, 2013) using BLS micro data data, Cavallo and Rigobon (2016); Gorodnichenko and Talavera (2017); Cavallo (2018*b*) using online prices, and Klenow and Malin (2010); Eichenbaum, Jaimovich and Rebelo (2011); Campbell and Eden (2014); Anderson et al. (2017) using scanner data. In the area of price dispersion, see Ellickson and Misra (2008); Arcidiacono et al. (2019); Eizenberg, Lach and Yiftach (2016); Kaplan et al. (2019); DellaVigna and Gentzkow (2019); Hitsch, Hortacsu and Lin (2019); Adams and Williams (2019) using offline data and Baylis and Perloff (2002); Chevalier and Goolsbee (2003); Boivin, Clark and Vincent (2012); Overby and Forman (2015); Aparicio and Cavallo (2017); Cavallo (2018*a*) using online data. These studies examine in great detail one dimension of price setting (e.g., competition across sellers), and the offline and online channels separately. We build upon these studies by documenting stylized facts about pricing in online groceries within and across chains, across channels, and over time. Our dataset is, to the authors' knowledge, the first effort in combining time precision (the same product collected at the same time across locations and retailers) and product precision (the same product matched across retailers). A set of carefully matched products has several advantages, which we detail in Section 2. Critically, it allows to map online data with scanner data; and it allows to rule out concerns in pricing differences due to assortment composition. Hwang, Bronnenberg and Thomadsen (2010) discuss the importance of assortment overlap between supermarket chains using scanner data.

We also relate to a growing literature on high-frequency pricing. Jank and Kannan (2005); Shiller et al. (2014); Fisher, Gallino and Li (2017); Dubé and Misra (2019) discuss how dynamic or personalized pricing can increase revenue. Miklós-Thal and Tucker (2019); Calvano et al. (2020); Brown and MacKay (2019) discuss competition incentives due to machine-based algorithms. Chen, Mislove and Wilson (2016) discuss evidence of algorithmic pricing in the Amazon marketplace, Hannak et al. (2014) describe steering in e-commerce sites, and Cohen et al. (2016) describe surge pricing at Uber. While these studies focus on a different industry, our results provide complementary perspectives to the advances of algorithmic pricing. We describe novel facts using prices collected in high-frequency in online groceries, including estimates of price stickiness, synchronization, price matching, and price grid exploration.

2 Data

This paper uses three main datasets: online, offline, and zipcode demographics data.

We collect web-scraped price data from the leading online grocery retailers in the Unites States: Amazon Fresh, Walmart Grocery, FreshDirect, Peapod, Jet, and Instacart. In the case of Instacart, we collected prices for Safeway, CVS, and Whole Foods.[2] These retailers have various market shares and geographic footprint. For instance, Amazon Fresh accounts for about 15-20% of the online grocery market; FreshDirect holds close to 60% of the market in New York City (New York Times (2020)). Throughout the paper we report estimates where price observations are weighted by market shares. As per industry reports, we use Amazon (0.35), Walmart (0.25), Peapod (0.13), FreshDirect (0.07), Jet (0.04), and Instacart (0.18). The latter we divide equally between Safeway, CSV, and Whole Foods. The use of market share weights is preferred because it allows to utilize all price observations. Robustness specifications are discussed in the Appendix.

We collect prices on 88 distinct products across 30 distinct zipcodes. We focus on a set of popular products across the main grocery categories. Our choice of popular products resembles related work on supermarket pricing (Eizenberg, Lach and Yiftach (2016); DellaVigna and Gentzkow (2019)). The data covers fresh produce, packaged food, and cleaning and personal care products. See Appendix A.1 for a list of products. In order to avoid too much traffic for websites, we focus on a set of zipcodes which are among the most populated cities in the U.S. However, we also choose cities that maximize geographic coverage. If we only focused on the most populated cities, nine out of the top fifteen cities would be within two states. Results throughout the paper primarily focus on one zipcode per city because retailers rarely vary prices within a city. In the Appendix we show robustness results using data collected from 109 zipcodes.

For each retailer, we created scripts that would systematically enter a zipcode into the website and then collect prices. A random VPN was also used to test robustness of

[2] In cases where a retailer does not offer their own grocery website, we collect price data from Instacart. Retailers set the prices of products available on the Instacart platform (Instacart (2019)).

data collection from different originating IP Addresses. Data was collected at the end of each month, and each retailer-zipcode data was collected within minutes. In some cases we collected the URL of the product directly; when such link is not available, we typed the product description (e.g., "1 bunch of bananas") in the search engine of the retailer, and collected all items in the results. We then matched each of the products across all retailers. See Appendix A.2 for methodological details on collecting matched products.

We collected two additional online datasets. We collected price data in high-frequency intervals (hours difference within a day) for Amazon and Walmart during about three months. In addition, we collected category-wide data for all retailers on the same day. We web-scraped all products in narrow categories, e.g. sodas, condiments and sauces. This allows to utilize a retailer's entire price distribution, which requires normalizing prices across categories and units of measurement. Therefore, scripts parse the words of the product description (prices, flavors, package sizes) and retailer-specific semi-supervised decision tree classifiers convert prices to price per ounce. For example, 1 pint of ice-cream would be converted to 16 ounces; therefore, if the price is $4.49, its unit price is $0.28 per ounce. Ounces was chosen because it is the most common measurement for unit prices.

Table 1: Data Description

		Online data			Scanner data
		Monthly	High-Frequency	Category-Wide	
(1)	Time period	05/2018 to 01/2019	12/2018 to 03/2019	03/2019	01/2017 to 12/2017
(2)	Price observations[a]	23,734	147,517	13,386	302,537
(3)	Distinct chains[b]	8	2	8	37
(4)	Distinct products[c]	88	75	42	82
(5)	Distinct locations[d]	30	8	2	22
(6)	Distinct states	23	8	2	18
(7)	Products per retailer[e]	54	64	286	63

Notes: [a]Price observations is the number of non-missing price observations. [b]Distinct chains is the number of distinct supermarket chains in the data. [c]Distinct products is the number of distinct products matched across all retailers. [d]Distinct locations is the number of distinct cities. [e]Average distinct products per retailer.

The second main dataset is Nielsen's Retail Scanner (RMS) data, which is provided by the Kilts Center at the University of Chicago.[3] This data covers sales and prices at the week, store, and UPC level. In most specifications we use the 2017 RMS dataset which is the latest available, but we also complement the analyses using all 2006-2017 RMS datasets. We restrict the sample to the set of matched products, to stores located in the same cities as those in the online data (Nielsen's data includes the city but not the zipcode of the store), and to grocery retailers. See Appendix A.3 for methodological details. Our baseline dataset covers over 302,000 price observations from 37 distinct grocery retail chains. While we focus on a selected but widespread geography, these cities account for approximately 40% of the observations in the RMS data. Moreover, we use the 2017 RMS dataset but the results are similar using the 2016 RMS dataset. In the Appendix we report robustness results using

[3]The data is available through the Marketing Data Center at the University of Chicago Booth School of Business. See https://research.chicagobooth.edu/kilts/marketing-databases/nielsen.

all retail formats (not just grocery chains). None of the chains are merged with the online data because retailer identifiers are masked in the Nielsen data.

Third, we collected zipcode-level covariates. We obtained the geographic coordinates from U.S. Census (2019*b*) and computed pairwise distances using the World GWGS 84 model (U.S. Department of Defense (2014)). In addition, we obtained home values at the zipcode level from Zillow Research (2018), income per capita and education at the zipcode level from the 2014-2018 American Community Survey (ACS) from U.S. Census (2019*a*), and population at the zipcode level from the U.S. Decennial Census of Population and Housing in 2010. We calculated the average measure of all zipcodes within a 10-mile radius of each delivery zipcode using the distances from NBER (2017).

The map in Appendix A.4 depicts all locations for which online data was collected. The average and median home values of the 30 zipcodes is \$648,437 and \$420,200, respectively.

Table 1 shows summary statistics. In total we collect data on 88 distinct matched products, of which 82 are identified in the scanner data. The remaining are either private label products or a handful of products for which a close substitute was not available (e.g., different package size). The number of products and locations is similar between channels.

3 Price Dispersion

We study price dispersion in online groceries using a set of matched products. We distinguish between two forms of price dispersion: within the same retailer (across locations), and across retailers (within the same location or across locations). We find that price dispersion across retailers is at least three times the price dispersion within retailers, and that online price dispersion is larger than offline price dispersion.

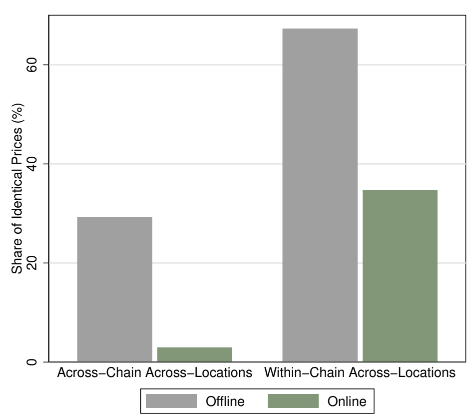


Figure 2: Price Dispersion of Oreo's Online and Offline

Notes: Figure shows the share of (almost) identical prices between all pairs of retailer-locations in different states using price observations on the same date. The share of identical prices is computed separately for locations of the same chain and for locations of different chains. A formal definition is below.

It is useful to start with one example: Oreo's. How similar are the prices of the same exact Oreo's product across online delivery locations and across offline stores? We compute the price difference between all retailer-location pairs of the same chain, and between pairs of different chains. A measure of dispersion is the percent of pairs that are (almost) identical. Figure 2 indicates that the share of identical prices is larger within chains than across chains; and in both cases identical prices are less likely online than offline.

3.1 Uniform Pricing

Uniform pricing is often defined as the practice of setting the same prices across locations (or even across products) within the same retail chain. Uniform prices have been documented in scanner data (Anderson, Jaimovich and Simester (2015); DellaVigna and Gentzkow (2019); Hitsch, Hortacsu and Lin (2019)) and in durable products in the online channel (e.g., Cavallo, Neiman and Rigobon (2014); Cavallo (2018*a*); Aparicio and Rigobon (2020)). However, there is no comprehensive evidence of pricing behaviors across geographies and across retailers in the online grocery market, or about the extent to which those behaviors are similar online and offline for the same set of matched products.

We measure uniform pricing following standard methods in the literature. We first compute pairwise price differentials at the product, time, and retailer level across all locations. We then compute the percent difference in absolute value between two prices:

$$\textit{Price Difference}_{s,s'}^{t,i} = \frac{|p_{s,r}^{t,i} - p_{s',r'}^{t,i}|}{(p_{s,r}^{t,i} + p_{s',r'}^{t,i})/2} * 100 \tag{1}$$

Where $p_{s,r}^{t,i}$ denotes the price of item i in location s, retailer r, at time t. For notation simplicity we define a retailer x location as a retailer-zipcode (retailer-store) in the case of online (offline) data. $\textit{Price Difference}_{s,s'}^{t,i}$ in equation (1) denotes the percent difference, in absolute value, for item i between a retailer location s and s' at time t. Note that in the case of the online data t stands for (nearly) the same timestamp; in the case of scanner data, t stands for the same week. We now focus on within-retailer price pairs and therefore $r = r'$. However, equation (1) allows the specification for price pairs across retailers in either the same location or in different locations.

A second measure of uniform pricing is the share of identical prices:

$$\mathbb{1}_{s,s'}^{t,i} = 1 \text{ if } p_{s,r}^{t,i} = p_{s',r'}^{t,i}; 0 \text{ otherwise} \tag{2}$$

Where $p_{s,r}^{t,i}$ is defined similarly. In the case of within-retailer pairs, the indicator $\mathbb{1}_{s,s'}^{t,i}$ takes value one when the price of the item i, retailer r, at time t is the same between two locations s and s'.[4]

The results are shown in Table 2. We report the median and mean of all price differences, as defined in equation (1). We also report the average share of identical prices, as defined in equation (2). We distinguish between price dispersion computed on price pairs of retailer-locations within and across states. Appendix B shows robustness results using data collected from multiple zipcodes within cities.

We find that online retailers have higher measures of non-uniform pricing. The mean share of identical prices across states is 40.3% online and 63.0% offline. The median

[4]In contrast to weekly-average scanner data, online data allows to compute exact price differentials of matched products in the same day. We consider two prices as identical when the percent difference is within 0.01%. Prices at Nielsen's scanner data are available at the weekly level and weighted by units sold. Due to measurement error, rounding, liquidation, or noise in the actual price points, we bin prices to 5% intervals. We find similar estimates rounding prices to 10 cents.

and average percent difference in pairwise prices is 4.9% and 9.8% online, respectively; while the equivalent measures are 0% and 7.0% offline. Our estimates of offline price dispersion follow those estimated in DellaVigna and Gentzkow (2019); for instance, they report a share of 68% identical log prices within a metropolitan area using all retail formats. Similarly, we find a share of 73.8% identical prices within the same state using all formats of retail chains (Appendix B.2) and 78.2% identical prices within the same state using grocery chains. Appendix B.1.4 shows robustness results controlling for product- and time- fixed effects.

Table 2: Price Dispersion Within Retailers

		Online data		Scanner data	
		Within-State	Across-State	Within-State	Across-State
(1)	Share of identical prices (%)	66.1 (0.65)	40.3 (0.12)	78.2 (0.21)	63.0 (0.11)
(2)	Median price difference (%)	0	4.9	0	0
(3)	Mean price difference (%)	5.2 (0.15)	9.8 (0.03)	3.4 (0.04)	7.0 (0.04)
	Fresh	6.4 (0.34)	11.6 (.08)	3.0 (0.06)	6.6 (0.07)
	Packaged	4.9 (0.17)	9.5 (0.04)	3.8 (0.06)	7.6 (0.06)
	Cleaning	3.7 (0.27)	6.5 (0.06)	2.0 (0.13)	3.6 (0.13)
(4)	Price pairs	5,318	166,185	40,088	78,616

Notes: Price dispersion is computed for price pairs of the same product, within retailers, across locations of the same state or across locations of different states. Results using all price pairs weighted by retailers' market shares. Standard errors reported in parenthesis.

Price dispersion tends to decrease with the perishability of the product, but among each type of product online price dispersion is larger than the offline. For instance, in the case of the online price dispersion across states, price dispersion is 11.6% in fresh produce, 9.5% in packaged food, and 6.5% in personal care and cleaning products. The equivalent measure is 6.6%, 7.6%, and 3.6% in the scanner data, respectively. While this is consistent with price transparency being greater in more durable products, there may also be more constraints in distribution or inventories replenishment.

Interestingly, the share of identical prices is over 90% for private labels in the online data. Although the sample is small, these are products for which one might expect the greatest price flexibility (i.e., grocers have more control over prices). These findings complement McShane et al. (2016)'s evidence of significantly higher price stickiness for private label products. Further research is needed to understand how wholesale price negotiation with upstream producers or brand-image concerns affect decisions for private labels.

3.2 Pricing Across Retailers

The online grocery industry is reportedly under increasing competition (Wall Street Journal (2016); New York Times (2018); Bloomberg (2018*a*)). The industry has recently experienced large acquisitions; two prominent examples are Walmart's acquisition of Jet for $3.3 billion, and Amazon's acquisition of Whole Foods for $13.7 billion. And it is experiencing a surge of partnerships in a race to make delivery faster and wider.[5]

[5]See Wall Street Journal (2018); CNBC (2019); Fortune (2019). Jet recently launched a new online grocery platform in New York (Bloomberg (2018*b*)). We collected prices for this new platform and found

It is therefore natural to wonder how price dispersion across competing retailers compares with that of within retailers. Once again, online price dispersion is found to be larger than offline.

The results are shown in Table 3. Price dispersion is computed within a location (price pair between two online retailers, in the same zipcode, at the same time) and across states (price pair between two retailers, in two cities in different states). Two results are noteworthy. First, online price dispersion across retailers is larger than the offline equivalent. This fact is observed both within and across locations. Consider two retailers located in the same location (columns (1) and (3)). In the online data, the share of identical prices and the average price difference is 6.7% and 25.8%, respectively. The equivalent measures are 31.5% and 15.7% offline, respectively. Now consider retailers in different states (columns (2) and (4)). The share of identical prices and the average price difference is 5.0% and 26.3% online, respectively; and they are 16.5% and 20.5% offline, respectively.

Moreover, price dispersion is found to increase for perishable items, and among each type the estimates are larger online than for the same products offline. For instance, within a narrow location, the average price difference is 28.3% for fresh products, 25.8% for packaged products, and 18.9% for cleaning and personal care products. When computed offline, the estimates are 16.5%, 15.4%, and 11.5%, respectively.

Table 3: Price Dispersion Across Retailers

		Online data		Scanner data	
		Within-Zipcode	Across-State	Within-City	Across-State
(1)	Share of identical prices (%)	6.7 (0.16)	5.0 (0.04)	31.5 (0.12)	16.5 (0.02)
(2)	Median price difference (%)	21.6	22.3	12.4	18.2
(3)	Mean price difference (%)	25.8 (0.14)	26.3 (0.04)	15.7 (0.04)	20.5 (0.01)
	Fresh	28.3 (0.36)	29.0 (0.09)	16.5 (0.07)	20.6 (0.02)
	Packaged	25.8 (0.16)	26.2 (0.04)	15.4 (0.05)	20.9 (0.01)
	Cleaning	18.9 (0.25)	19.1 (0.07)	11.5 (0.15)	16.2 (0.04)
(4)	Price pairs	25,239	357,560	139,790	2,512,808

Notes: Price dispersion is computed for price pairs of the same product, across retailers, within the same zipcode and across zipcodes (from different states). Results using all price pairs weighted by retailers' market shares. Standard errors reported in parenthesis.

The second finding is that, if we compare Table 2 and Table 3, price dispersion across chains is substantially larger than within chains. In fact, price dispersion across chains in the same city is between three to five times the price dispersion within chains in the same state. Consider the online case. The average price difference, within chains, within states and across states, is 5.2% and 9.8%, respectively; and the equivalent measures, across chains, are 25.8% and 26.3%. A similar behavior is observed in the offline data. The average price difference, within chains, is 3.4% and 7.0% for the corresponding specification; and across chains it is 15.7% and 20.5%.

These magnitudes illustrate the importance of chain effects in explaining price dispersion. In Appendix B we present results from several robustness specifications.[6] In fact, retailers' price segmentation is remarkably stable across locations and products. A machine

significantly larger estimates of non-uniform pricing, compared to the older platform.

[6]We replicate the analysis using data from 109 zipcodes. In this case, the data includes multiple zipcodes

learning classification perspective is illustrative: five random products from two retailers can predict which retailer is more expensive with 75% accuracy. Appendix D shows additional evidence about price segmentation across chains.

4 Explaining Online Price Dispersion

4.1 Decomposition

We decompose price variation into retailer, product, and retailer x delivery zipcode effects. We start with the following model of variance decomposition.

$$p^r_{i,s} = \alpha_i + \beta_r + \gamma_{i,r} + \epsilon_{i,s,r} \tag{3}$$

The price $p^r_{i,s}$ (in natural logs) of product i in location s of retailer r can be decomposed into four components: a product effect, α_i; a retailer effect, β_r; a retailer-product effect, $\gamma_{i,r}$; and a residual $\epsilon_{i,s,r}$ that captures variation in prices for the same product across different delivery zipcodes of the same chain.

This model is similar to Kaplan et al. (2019) using scanner data, but ignoring autocovariances that would exist with time lags in equation (3). We explain variation in relative prices, and therefore we subtract the mean price of each product. Each term in equation (3) is computed through sequential linear regressions (i.e., first with retailer indicators, then retailer-product indicators), obtaining the corresponding residual variance, computing its share (subtracted from that of the previous regression, and as a share of price variance). Robustness specifications are discussed in Appendix C.1.

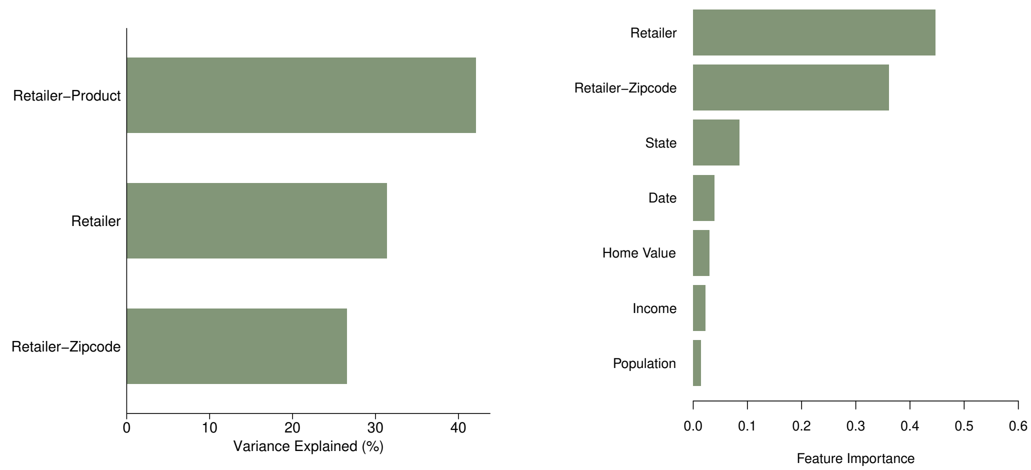
Results are shown in Panel (a) in Figure 3. The retailer effect explains about 30% of the variation in prices, and the retailer-product effect explains about 40%. Therefore, the residual, which is attributed to the retailer-zipcode effect (chains setting different prices across zipcodes) explains close to 30%. Note this is a lower bound because the retailer-zipcode effect is included last in the sequential regressions.

Instead of assuming a particular form of price decomposition, we can alternatively explain variation in prices through feature importance methods in machine learning. We choose a random forest because of its widespread use, it allows to include a larger number of features (while de-correlating trees when the features are correlated), and it allows two extract two interpretable measures of feature importance (Breiman (2001); Friedman, Hastie and Tibshirani (2001)).

We train a random forest with k-fold cross-validation to obtain the relevant hyperparameters that minimize *MSE* (mean squared error), and then estimate feature importance predicting normalized prices. Prices are normalized by the mean and standard deviation within each fold. We include a richer set of covariates: retailer, retailer x delivery zipcode, state, collection date, income per capita, home values, and population. Income and population are defined as the average within a 10-mile radius of each zipcode, and home values are defined as the median home value in each zipcode (Section 2).

Results are shown in Panel (b) in Figure 3. Feature importance is measured by the node impurity, i.e. the residual sum of squares from splitting on the relevant feature, averaged over all trees. The values are scaled such that the sum of individual scores adds up

within a city. We then replicate the analysis using a random subset of products. In addition, we report results on online price dispersion using equal-sampling, equal-weights in retailers' price pairs. We report results from a fixed-effects model controlling for product- and time- fixed effects. Finally, we replicate the analysis of offline price dispersion using data from all formats (not just grocery chains).



(a) Variance Decomposition

(b) Random Forest

Figure 3: Components of Price Variation

Notes: Panel (a) shows the results of the variance decomposition following the model in equation (3). Estimates are averages across months. Panel (b) shows feature importance (node impurity) obtained through a random forest. We use 5,000 trees, two maximum features per split, and a minimum node size of 8 units. Parameters are obtained through 5-fold cross-validation.

to one. For instance, the estimates can be interpreted as indicating that the chain component is responsible for about 45% of the information gain in the training process. We also find that the retailer-zipcode accounts for a similar portion, while local demographics are less predictive of price variation. This provides additional support to the evidence that online grocers exhibit significant non-uniform pricing across locations. Appendix C.1 shows robustness results using an alternative measure of feature importance.

4.2 Price Elasticities

We now explore the degree to which price elasticities are informative of online price dispersion. There are a number of ways to compute price elasticities in the literature. We follow similar reduced-form specifications to Nijs et al. (2010); DellaVigna and Gentzkow (2019); Hitsch, Hortacsu and Lin (2019).

We pool the 2015-2017 scanner datasets together, requiring at least 80% of price observations available in a given store-product pair. We then estimate the following fixed effects model:

$$\log(q_{s,i,t}) = \alpha + \eta_{s,i}\log(p_{s,i,t}) + \gamma_{s,j,y} + \epsilon \tag{4}$$

Where we regress log units sold of product i in store s in week t ($q_{s,i,t}$) on log weekly price ($p_{s,i,t}$), and including a store-product-year fixed effect ($\gamma_{s,j,y}$). We estimate equation (4) for each store-product pair separately. Therefore, $\hat{\eta}_{s,i}$ denotes the own-price elasticity of product i in store s, which will be used in explaining offline price dispersion.[7] Because it

[7]It is also possible to follow DellaVigna and Gentzkow (2019)'s approach of instrumenting log weekly price ($p_{s,i,t}$) with the average price of the same product, in the same retail chain, but of stores located in

is not possible to perform a map between store-product elasticity and online prices, additionally we re-estimate equation (4) at the product-DMA level (e.g., Los Angeles CA is a DMA), including a fixed effect for product-DMA-week ($\delta_{i,m,t}$). The corresponding $\hat{\eta}_{i,m}$ denotes the own-price elasticity of product i in city m, which will be used in explaining online price dispersion. The histogram of the elasticities is depicted in Appendix C.2. The median own price elasticity is -2.52 when computed at the store-product level and it is -2.15 at the DMA-product level.

We then relate these price elasticities with the price dispersion estimated in Section 3. We consider a model of the form:

$$\Delta(p_{s,i,t}, p_{s',i,t}) = \alpha + \beta\Delta(\hat{\eta}_{s,i}, \hat{\eta}_{s,i}) + \zeta_i + \gamma_t + \delta_m + \epsilon \tag{5}$$

Where $\Delta(p_{s,i,t}, p_{s',i,t})$ denotes the pairwise price differential between two retailer-location pairs of the same chain (or different chains) in percentage terms, similar to equation (1); $\Delta(\hat{\eta}_{s,i}, \hat{\eta}_{s',i})$ denotes the elasticity differential between stores s and s' (or DMAs m and m'); and ζ_i, γ_t, δ_m denote a series of product-, time-, and DMA- fixed effects, respectively. We estimate alternative specifications to equation (5): the price differential in absolute terms, the elasticity differential in absolute terms, the average elasticity ($\bar{\eta}$) for a given product between two stores (or for a given product in a DMA), and the standard deviation of elasticities across retailers for a given product in a DMA ($\sigma(\eta)$).

Note that, while in the offline data we can map the price dispersion between two store-product pairs with their store-product elasticities, the same is not feasible in the online data. However, a reasonable analogous approximation is to map the delivery zipcodes with their corresponding DMAs in the scanner datasets, and then use elasticities estimated at the product-DMA level as described above.

The results are summarized in Table 4. Overall, we find precisely estimated coefficients with the correct economical sign. First, consider the set of results in columns (1)-(2) which explain price dispersion across locations within the same retail chain. We expect a positive coefficient between the elasticity differential and the price differential. Intuitively, the more price-sensitive demand in one location for a given product, the tighter its price, and controlling for locations the larger the price dispersion between retailer-location pairs. This is what the point estimates of the first model indicate. For instance, an additional elasticity differential point relates to 1.13 additional percentage points in relative prices.

Second, consider the set of results in columns (3)-(4) which explain price dispersion across retailers within a location. For the reasons just explained, we expect a positive coefficient between the elasticity differential and the price differential. This is what the 2.40 point estimate in column (3) indicates. The second and third model estimate the relationship between price dispersion and the average elasticity (between two store-product pairs for offline, or a product-DMA for online). The coefficients indicate that a lower average elasticity (a more price-sensitive demand) accounts for a lower price dispersion for a given product across competing retailers. For example, an additional elasticity point relates to 1.52 additional percentage points in online relative prices. The fourth model estimates the relationship between price dispersion and the standard deviation of (the retailer average) elasticities in a given product-DMA. Similarly, the results indicate that a larger standard deviation of elasticities across chains explains a larger price dispersion, both offline and online (the latter not statistically significant).

---

different areas. Like that study, we find that both own-price elasticities are similar (correlation of 86.3% and median absolute difference of 0.23); since focusing on matched cities between online and offline yields fewer IV elasticities, we report own-price elasticities obtained using equation (4).

Table 4: Price Elasticities and Price Dispersion

		Within Retail Chains, Across DMAs		Across Retail Chains, Within DMAs	
		Scanner	Online	Scanner	Online
Dispersion	Elasticity	(1)	(2)	(3)	(4)
$\Delta(p_s, p_{s'})$	$\Delta(\eta_s, \eta_{s'})$	1.13	0.52	2.40	-
		(0.11)	(0.11)	(0.09)	
$\|\Delta(p_s, p_{s'})\|$	$\bar{\eta}(\eta_s, \eta_{s'})$	-	-	0.56	1.52
				(0.06)	(0.32)
$\|\Delta(p_s, p_{s'})\|$	$\|\bar{\eta}(\eta_s, \eta_{s'})\|$	-	-	-0.60	-1.47
				(0.07)	(0.35)
$\|\Delta(p_s, p_{s'})\|$	$\sigma(\eta)$	-	-	1.60	0.17
				(0.15)	(0.45)
Product FE		YES	YES	YES	YES
Time FE		YES	YES	YES	YES
DMA FE		YES	YES	YES	YES

Notes: Table reports the coefficients of estimating several specifications of equation (5): the left-hand side price dispersion outcome is indicated in the Dispersion column and the right-hand side elasticity-based measure is indicated in the Elasticity column. Robust standard errors in parenthesis. Models are estimated separately using price pairs weighted by retailers' market shares and including product-, time-, and DMA-fixed effects.

Perhaps most importantly in this analysis, the offline elasticities are able to explain online price dispersion. Intuitively, while we lack data to compute online elasticities, arguably the cross-market variation in offline elasticities is informative about the cross-market variation in online elasticities. The results in Table 4 show that the variation in offline elasticity is informative of the online price dispersion, both when measured across retailers (within the same DMA) and across locations (within the same retailer).

4.3 Shipping Costs

In the same vein, we study whether shipping costs explains price differentiation within a chain. Although data on shipping costs is not available, the shipping costs are intrinsically related to the distance between the distribution facility and the delivery zipcode (Houde, Newberry and Seim (2017)). We collect data on the location of the fulfillment centers and offline stores, and define the shipping distance as the distance between the target zipcode and the closest location fulfilling the order.[8] Formally, $d_{r,z} \equiv \min_{z'} \text{dist}(z - f),\ \forall f \in F_r$.

We explain geographic price variation within the chain estimating the following model:

$$\Delta(p_{i,t}^{r,z}, p_{i,t}^{r,z'}) = \alpha + \beta \Delta(d_{r,z}, d_{r,z'}) + \zeta_i + \gamma_t + \epsilon \tag{6}$$

Where $\Delta(p_{i,t}^{r,z}, p_{i,t}^{r,z'})$ denotes the pairwise price differential for product i and retailer r at time t between zipcode z and zipcode z', in percentage terms similar to equation (1);

[8]Data on the location of the facility fulfilling an online order is not publicly available. We set up programs that enter a zipcode and retrieve the zipcodes of the offline stores or distribution centers

$\Delta(d_{r,z}, d_{r',z'})$ denotes the distance differential between their corresponding fulfillment centers (between the distance from zipcode z to its fulfillment location and the distance from zipcode z' to its fulfillment location); ζ_i and γ_t denote a series of product- and time- fixed effects, respectively. Therefore, equation (6) allows to test whether the relative distance to the fulfillment center is priced in the relative prices for the same exact product.

The results are shown in Table 5. Overall, the coefficients are precisely estimated and indicate that geographic price variation can partly be explained by delivery distances (thereby shipping costs). Consider the results in columns (2) and (4). The coefficients indicate that 10 additional miles increases the difference in pairwise prices by 0.14 percentage points, and that an additional 1 log point distance increases prices by 0.14 percentage points, respectively. Columns (1) and (2) re-estimate excluding state fixed effects and differences in income per capita. Importantly, these specifications are estimated for retailer-location pairs for which their closest (most plausible) fulfillment center is the same. In other words, we exploit the variation in prices between San Jose and San Francisco for the same set of products in the same retailer and the variation in distance to the fulfillment center, given that both locations are served by the same fulfillment center. These estimates can be interpreted as suggesting that an additional 25 miles increases prices by \$0.02 for an average product of \$4.4 between two delivery zipcodes of the same retailer.[9] In principle, the variation in distances could still be informative when two locations are served by different fulfillment centers, although in those cases equation (6) omits other factors that can interact with the shipping costs. Columns (5) and (6) show that the effect weakens precision but maintains the expected sign in retailer-location pairs of the same state.

Table 5: Price Dispersion and Shipping Costs

	Within Retailer, Across Zipcodes					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta(d_{r,z}, d_{r',z'})$	0.012	0.014			0.002	
	(0.001)	(0.002)			(0.005)	
log Distance			0.126	0.142		0.207
			(0.035)	(0.048)		(0.087)
Product FE	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES
State FE	NO	YES	NO	YES	YES	YES
Income	NO	YES	NO	YES	YES	YES
R^2	0.03	0.05	0.03	0.04	0.09	0.09
Observations	26,821	26,821	26,821	26,821	27,693	27,693

Notes: Table reports the coefficients of estimating several specifications of equation (6). Models are estimated using price pairs in the extended dataset (covering multiple zipcodes within a city, as described in Section 2), price pairs weighted by retailers' market shares and including product-, time-, and state-fixed effects. Robust standard errors in parenthesis.

[9] Houde, Newberry and Seim (2017) studies the economics of Amazon's network of fulfillment centers (not Amazon Fresh). The authors find that it costs Amazon between \$0.17 and \$0.41 to ship a box of \$30 for 100 miles. Although we study online groceries (and not just Amazon Fresh), the set of estimates in Houde, Newberry and Seim (2017) provide a benchmark which is qualitatively comparable to Table 5.

Finally, recall that Section 3.1 (and Appendix B.1.1) showed that price dispersion within the retailer and across nearby zipcodes of the same city is notably lower than the price dispersion across cities. Those findings are consistent with the estimates in Table 5, in the sense that nearby zipcodes in the same city are not only served by the same fulfillment center but also have nearly the same distance to that fulfillment center.

4.4 Pricing Technology

Improvements in pricing technology allow retailers to gain price flexibility and, thereby, increase its frequency of price changes (Brown and MacKay (2019)). We now connect the intensity of algorithmic pricing with the degree of non-uniform pricing by estimating the following model:

$$\Delta|(p_{i,t}^{r,z}, p_{i,t}^{r,z'})| = \alpha + \beta \tilde{p}(p_{r,z}, p_{r,z'}) + \zeta_i + \gamma_t + \delta_z + \epsilon \tag{7}$$

Where $|\Delta(p_{i,t}^{r,z}, p_{i,t}^{r,z'})|$ denotes the pairwise absolute price differential for product i and retailer r at time t between zipcode z and zipcode z', and $\tilde{p}(p_{r,z}, p_{r,z'})$ denotes the average (or similar) probability of a price change across both price sequences.

Table 6: Price Dispersion and Frequency of Price Changes

	Within Retailer, Across Zipcodes		
	(1)	(2)	(3)
$\overline{p}(p_{r,z}, p_{r',z'})$	9.51	3.51	4.16
	(0.40)	(0.20)	(0.16)
Product FE	YES	YES	YES
Time FE	YES	YES	YES
State FE	YES	YES	YES
R^2	0.24	0.23	0.24
Observations	117,247	117,247	117,247

Notes: Table reports the coefficients of estimating several specifications of equation (7). Models are estimated using price pairs weighted by retailers' market shares and including product-, time-, and state- fixed effects. Robust standard errors in parenthesis.

Overall, the results in Table 6 indicate that the lower price stickiness, the greater the degree of non-uniform pricing. That is, a greater intensity of price changes (across time in a given delivery zipcode) magnifies the price dispersion (for the same product and time, between two given delivery zipcodes). Columns (1), (2), and (3) show the results using the overall average probability of a price change, the average probability in a given timestamp, and whether at least one of the locations changed prices, respectively.

4.5 Demograhics

Finally, we explain price dispersion within retail chains, across locations, using a set of demographics measured at the zipcode level. In particular, we estimate a model similar to equation (5); in this case, the relationship between the relative prices in two locations s and s' for the same retailer-product pair ($\Delta(p_s, p_{s'})$), and home values, income per capita, population, and share of educated adults. We include product- and time- fixed effects.

Local demographics can explain relative prices between two delivery locations, although the point estimates are (highly significant) relatively small. For instance, prices increase with home values and income; an additional \$200,000 and \$5,000 in median home values and annual income per capita, respectively, relates to an increase of approximately 0.26 and 0.28 percentage points in prices, which amounts to about 3% in price dispersion. The results are similar in log differences. An additional 0.10 log difference in income relates to an increase of 0.29 percentage points in prices. The complete set of results is shown in Appendix C.3.

4.6 Summary

This set of findings inform our understanding of the geographic price dispersion. The online grocers exhibit a remarkable flexibility in setting prices at the zipcode level; but interestingly, price differentiation does not appear to be primarily driven by local demographics. Most importantly, a more price-sensitive demand, a greater intensity of price frequency, and variation in shipping distance magnify the price differentiation across locations.

5 Patterns of Algorithmic Pricing

In this Section, we use high-frequency hourly data collected during three months for Amazon and Walmart, which account for close to 50% of the online grocery market. We turn our attention to recovering the key patterns of algorithmic pricing, such as frequency, synchronization, price matching, and price exploration. We construct analogous statistics using scanner data of grocery chains; Appendix B.2 reports robustness results using all formats.

5.1 Price Stickiness

We begin with estimates on the frequencies of price changes. Because data is collected in high-frequency, we report price changes at various time intervals. Let L be the time interval of interest. A given product i at time t in retailer r and delivery zipcode z experiences a price change when its current price, $p_t^{r,z,i}$, is different from at least one of the prices collected throughout $p_{t-L}^{r,z,i}$. Formally,

$$\textit{Price Change}_t^{r,z,i} = 1 \;\text{ if }\; \exists h \in L : p_t^{r,z,i} \neq p_{t-h}^{r,z,i} \tag{8}$$

Where the indicator in equation (8) takes 1 one when, if we consider a weekly price change, the price of a retailer-zipcode-product experienced at least one change during the previous 168 hours (7 days).

In the case of offline data, price changes are defined similarly, but with the simplification that the average price per week for each store-product is reported. Prices in the scanner data are weekly volume-weighted averages, which have been found to overstate frequencies of price changes (Campbell and Eden (2014); Cavallo (2018*b*)). In order to better account for measurement error, liquidation, or fractional prices in scanner data, we bin prices into 5% bins. This may not alleviate all concerns because, in addition to time-averaging, scanner prices include the effects of coupons or loyalty cards, both of which generate an artificial price change. For example, a single purchase with coupons can induce a price change in that store-product-week even if the tag price remained constant. See additional discussions in Appendix E.1. In the case of online data, prices are rounded to two decimals.

The results are shown in Table 7. First, we document large estimates of daily, and even intra-day, price changes. The probability of a price change within a day is 0.07 and 0.08 in Amazon and Walmart, respectively. The probability between two consecutive days is 0.17 in Amazon and 0.12 in Walmart. The distinction between the two grocers becomes more noticeable at longer intervals. For instance, close to 50% of the products exhibit at least one price change during a given week in Amazon, while in Walmart it occurs for less than 25% of the products. The implied duration is 8.7 days and 10.1 days in Amazon and Walmart, respectively. Interestingly, price changes do not occur uniformly throughout the week. For instance, Amazon's and Walmart's frequencies in Wednesday and Thursday, respectively, are an order of magnitude larger than other week days. The results by day of the week are described in Appendix E.2.

Table 7: Price Stickiness

	Amazon	Walmart	Scanner
Prob. Price Change			
Same day	0.069 (0.002)	0.082 (0.002)	-
Daily	0.173 (0.002)	0.117 (0.002)	-
Weekly	0.479 (0.003)	0.231 (.003)	0.317 (0.003)
Monthly	0.736 (0.004)	0.500 (0.005)	0.578 (0.004)
Duration			
Median duration	1.2 weeks	1.4 weeks	3.0 weeks

Notes: Probability of price change denotes the average probability of any price change (increase or decrease) at the corresponding time interval. Estimates are an equal-weight average across retailer-zipcode-product. Median implied duration measured as $-1/(\ln(1-f))$, where f is the ratio of number of price changes to the number of price observations. When measuring online duration, the price observations are restricted to at least 12 hours difference. Standard errors reported in parenthesis.

The online frequencies can be compared, with the caveats mentioned above, with the frequencies in the scanner data. The probability that a store-product exhibits a price change over two consecutive weeks and over a month is 0.32 and 0.58, respectively. The implied duration is 3.0 weeks. These set of estimates can also be compared with the literature. Online and offline monthly frequencies have been estimated in the range of 0.30 to 0.55.[10] The time period can partly explain differences with the literature. Combining the 2006-2017 RMS scanner datasets we observe a trend of increasing price frequencies over

[10]In the case of offline frequencies of price changes, Eichenbaum, Jaimovich and Rebelo (2011) reports weekly frequencies between 0.24 and 0.43; Anderson et al. (2017) reports a weekly frequency of 0.21; Cavallo (2018*b*) reports a median weekly frequency of 0.25; Klenow and Malin (2010) reviews the literature and shows monthly frequencies between 0.35 to 0.55. In the case of online data, Boivin, Clark and Vincent (2012) reports a monthly frequency of 0.41 in online books; Gorodnichenko and Talavera (2017) reports a

time. More precisely, the probability of a weekly price change increases from below 27% in 2006 to over 30% in 2017, and the median duration decreases from 4.1 weeks to 3.7 weeks. While we lack data to test a formal hypothesis about the managerial process (unlike Zbaracki et al. (2004)), it is possible that improvements in pricing technology facilitate retailers to implement price changes more frequently. We estimate frequencies over time as follows. We use entire modules data from Nielsen's grocery chains (a module is a narrow category, e.g. potato chips), we sample at most 10 random stores per chain-year, exclude store-product pairs with less than 80% observations available, and then sample 50 random products per store-year. The amount of data is considerable: it covers 945 distinct products and 3,915,922 price observations. Additional results are discussed in Appendix E.3.

Returning to our estimates of online price stickiness, Walmart's frequencies are relatively similar to benchmarks using either offline or online data. In fact, it is plausible that Walmart's online behavior is influenced by its offline stores. For instance, Anderson et al. (2017) describes how coordination at the retail chain affects pricing decisions, and Ater and Rigbi (2019) describes how the existence of price disclosure online affects price dispersion at the chain level. However, Amazon exhibits a degree of price flexibility which is substantially larger than any comparable statistic previously reported in the literature. The tails are informative: 10% of the product-zipcode combinations have average *daily* probabilities of a price change above 0.45. If Amazon's results can be used as any guide about trends in price setting, they suggest that prices in online groceries are becoming increasingly flexible due to algorithmic pricing.[11]

With these ideas in mind, we test whether algorithmic pricing increases price differentiation within chain and across locations. Once we obtain the daily indicators of price changes for each retailer-zipcode-product combination (over time), we map them to the corresponding price dispersion between two zipcodes (for the same retailer-product and time). We then estimate the following model:

$$\Delta|(p_{i,t}^{r,z}, p_{i,t}^{r,z'})| = \alpha + \beta \tilde{p}_{i,t}^{r} + \gamma_t + \epsilon \tag{9}$$

Where $|\Delta(p_{i,t}^{r,z}, p_{i,t}^{r,z'})|$ denotes the pairwise absolute price differential for product i and retailer r at time t between zipcode z and zipcode z'; and $\tilde{p}_{i,t}^{r}$ denotes an indicator that takes value 1 when the price of product i at time t changed in either location. We also estimate a specification using the intensity of algorithmic pricing, defined as the average frequency of daily price changes.

The results in Table 8 indicate that when a price changes in either of the delivery zipcodes, price differentiation on average increases by 5.7 percentage points. Similarly, an increase in just 10 percentage points in the algorithmic pricing intensity, relates to an increase in price differentiation across locations of 2.7 percentage points. A visual summary to this analysis can be seen in Figure 1 in the Introduction. It shows the price differentiation (across two given zipcodes) as a function of the frequency of daily price changes (across those two zipcodes and for the same product). The results pool all product-zipcode pairs across time within the chain.

weekly frequency between 0.20 and 0.37 in electronics; Cavallo (2018*b*,*a*) documents monthly frequencies between 0.27 and 0.48 in CPI categories. Although the studies follow standard methodologies and provide informative benchmarks, there are potential differences with our data, e.g. thresholds of size of price changes, weights on food products, and time periods.

[11]The term algorithmic pricing is often substituted for robo-pricing or dynamic pricing. We note that dynamic pricing is often used in models of intertemporal price discrimination (Nair (2007)).

Table 8: Price Differentiation and Algorithmic Pricing

	Online Price Differentation	
	(1)	(2)
Price Change	5.77	
	(0.06)	
Algorithmic Intensity		0.27
		(0.06)
Time FE	YES	YES
Observations	305,230	305,220

Notes: Table reports the coefficients of estimating several specifications of equation (9). Columns (1) and (2) estimate the effect of a daily price change and the effect of the (average) algorithmic intensity of a retailer-product, respectively. Robust standard errors in parenthesis.

A note about causality. The reader might question whether a link to causality can be established. A true "experiment" is not possible: randomizing the algorithmic pricing intensity *also* entails deciding the degree of price differentiation. Instead, our identification relies on: *given* the pricing algorithms that retailers have in place and switch on/off, exploit variation in frequency intensity (within and across, products and time) and variation in price differentiation across locations.

5.2 Synchronization and Price Matching

A related feature of algorithmic pricing is the synchronization of price changes, either across locations (within a retailer-product) or across retailers (within a zipcode-product).

Synchronization *across locations* is defined as the probability that at least one location exhibits a price change, conditional on a price change in another location. In order to be synchronized, these price changes must take place within twelve hours.[12] We focus on four probabilities: probability of a price increase (or decrease) given an increase; probability of a price decrease (or increase) given an increase. We also report the probability of any price change, conditional on at least one price change.

The results are shown in Table 9. We find some evidence of synchronization. In particular, Walmart shows relatively large conditional probabilities of a price change–although not necessarily in the same direction. When a zipcode-product exhibits a price change, the same product is likely to experience a price change in a different location within hours; but it can be in the opposite direction. Amazon, on the other hand, exhibits significantly lower measures of synchronization. Row (3) shows that the probability of any price change, conditional on observing a price change, is greater than the (unconditional) daily probability of a price change. The synchronization probability is 0.10 in Amazon and 0.38 in Walmart; and these compare to the unconditional probabilities of 0.07 and 0.08 (Table 7), respectively. Hence, these estimate indicate that price changes do tend to occur across locations at

[12]Data is partitioned into time blocks where the start and end hour is the same for a retailer-product. The interval between one collection time and another collection time is several hours (capped at twelve hours).

the same time. However, the fact that the sign of the price change is not the same suggests another scope of algorithmic pricing.

The third column in Table 9 reports analogous synchronization measures for the scanner data. In particular, we restrict attention to retail chains which have at least four stores, and randomly sample four stores for each chain-product combination. This is slightly more robust than sampling four random stores for all products in the retail chain. We then observe, conditional on a weekly price change, whether other stores of the chain also exhibit a price change for that product-week. The estimates show a remarkable degree of synchronization, i.e. stores tend to increase prices (or decrease prices) at the same time. The probability that a store increases (decreases) the price, conditional on a price increase (decrease) on another store, is 0.63 (0.67), respectively.

Table 9: Synchronization Across Locations

		Amazon	Walmart	Scanner
(1)	*Cond. on Increase*			
	Prob. of price increase	0.046 (0.018)	0.283 (0.019)	0.633 (0.006)
	Prob. of price decrease	0.062 (0.021)	0.371 (0.020)	0.245 (0.005)
(2)	*Cond. on Decrease*			
	Prob. of price increase	0.082 (0.028)	0.329 (0.018)	0.234 (0.005)
	Prob. of price decrease	0.102 (0.031)	0.197 (0.016)	0.665 (0.006)
(3)	*Any Change*			
	Prob. of price change	0.100 (0.020)	0.381 (0.015)	0.770 (0.004)

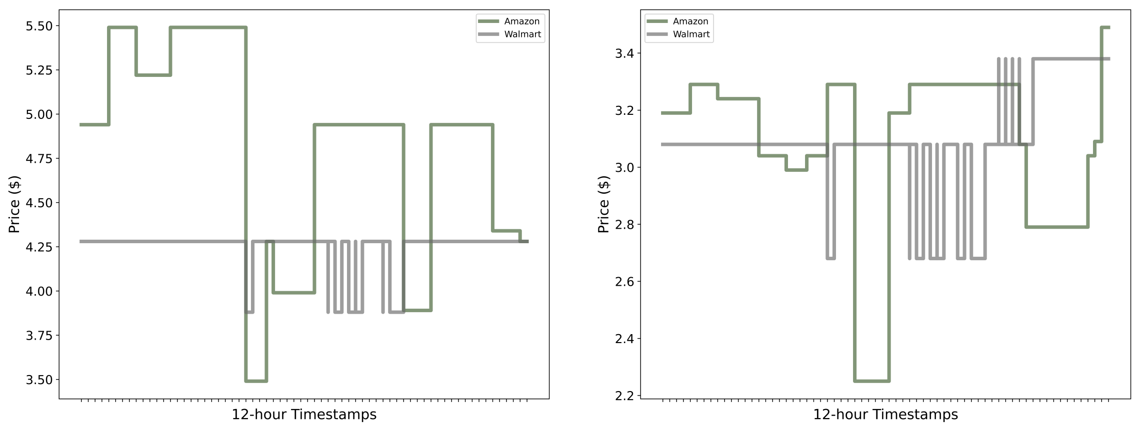
Notes: Table estimates probabilities of price change synchronization, within each retailer, across four common zipcodes. When a product experiences a price change in a given zipcode, we observe whether a different zipcode experiences a price change, for the same product, within at most 12 hours. Standard errors reported in parenthesis.

We also computed synchronization *across retailers* within the same delivery zipcode and within a window of 6 hours. We find that the probabilities of synchronization are nearly 0%, and are therefore omitted from Table 9. It is possible that a 6 hour window is too restrictive. For this reason we compute synchronization within 24 hours and ignoring the direction of the price change. In other words, we estimate the probability of a price change for a zipcode-product conditional on a price change in the competing retailer for the same zipcode and product. In this case we find probabilities that are close to 3%, which are significantly smaller than the unconditional probabilities.[13]

While price changes do not appear to be synchronized, it is still possible that retailers keep track of each others' prices and use them as input to their price setting. In fact, Walmart's policy grants consumers the possibility to match lower prices and save the difference (Walmart (2020)). The high-frequency data allows us to detect that retailers often match each other's price. We define a price matching event when Amazon and Walmart have a price within 3 cents for the same product-zipcode in a window of 24 hours. We find that in 11% of the product-zipcode pairs the average probability that Amazon and Walmart have

[13] To the best of the authors' knowledge, these are the first estimates of across-location or across-retailer synchronization in online groceries. Gorodnichenko and Talavera (2017); Gorodnichenko, Sheremirov and Talavera (2014) report no synchronization across sellers of durable goods at the weekly frequency. Cavallo (2017) reports low rates of synchronization between price changes of the same retailer in its online and offline store.

a matching price is greater than 0.05, and in 7% of the pairs the probability is over 0.15.[14] Appendix E.6 shows the distribution of the matching events and additional methodological details. We also examine the timestamps preceding the matching event to understand whether retailers coincide on a matching price or whether one of retailers actively matches an existing competitor price. In almost all cases, a retailer sets the price first. Moreover, across all product-zipcode matching events for which the price path can be recovered, the average probability that Amazon matches Walmart is 74.8%. That is to say, in 74.8% of the occasions Walmart "sets" the price and Amazon "matches".



(a) Ben & Jerry's Chocolate Fudge Brownie

(b) Philadelphia Cream Cheese

Figure 4: Price Matching Events

Notes: Panels (a) and (b) show the high-frequency price series in 12-hour timestamps for two selected products. In both panels, we focus on the same product-zipcode across retailers.

The examples in Figure 4 provide a visual perspective. Although both retailers tend to change prices relatively frequent, Amazon explores the price grid while Walmart switches between focal prices. And occasionally, throughout the price path, Amazon will match a Walmart price.

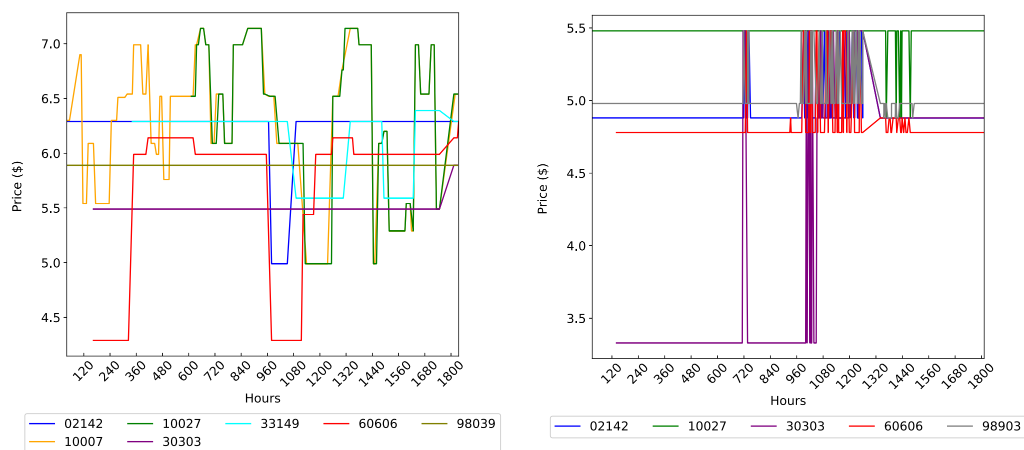
The price levels at which price matching occurs are also interesting. Tracking the price sequence leading to the event, we price matching is associated with lower prices. More precisely, 88% and 79% of the price matching events occur at prices that are below the median price (for that product in that zipcode) in Amazon and Walmart, respectively. More formally, we estimate the following model: $p_{i,t}^{r,z} = \alpha + \beta Event_{i,t}^{r,z} + \gamma_t + \zeta_{r,z,i} + \epsilon$, where $p_{i,t}^{r,z}$ is the (log) price of product i in retailer r at time t in zipcode z, $Event$ is an indicator that takes value 1 with a price matching, and $\gamma_t + \zeta_{r,z,i}$ control for day and for retailer-zipcode-product fixed effects, respectively. Price matching relates to a 2.6% decrease in prices; similarly, a 10% increase in price matching events is associated with a 1% decrease in prices ($p < 0.001$ in both cases). Interestingly, these findings bring preliminary evidence

[14]The percent of product-zipcode pairs with a probability of price matching greater than 0.05 increases from 11% to 19% if we define a price matching event when the price difference is within 10 cents. We do not consider matching events exceeding a 24-hour range or in a different delivery zipcode, although less stringent specifications that study delayed price matching or matching in baskets of products or locations are interesting.

to the recent theoretical work in Miklós-Thal and Tucker (2019), showing that machine-based pricing can sometimes toughen price competition and increase consumer surplus.

5.3 Price Grid

We examine the menu of prices used by the two leading online grocery retailers. We document that Amazon tends to use significantly more distinct prices for the exact same product than Walmart. In fact, for the same time period and product, Amazon might use six times more distinct prices than Walmart–despite that the daily frequencies are of the same order of magnitude.



(a) Amazon (b) Walmart

Figure 5: Algorithmic Pricing Example

Notes: Figure shows prices of a Diet Coke 12 fl oz 12 Pack during about three months. Data was collected multiple times a day and therefore the horizontal denotes the hours since the first collection time (e.g., 120 label is the end of the 5th day). Panel (a) shows Amazon Fresh prices across locations and hours. Panel (b) shows Walmart Grocery prices across locations and hours.

Again, a guiding example allows to visualize the sharp distinction between price choices. Figure 5 shows the typical behavior of a product, in this case Diet Coke 12 fl oz 12 Pack. While Amazon explores more distinct prices and exhibits a greater degree of non-uniform pricing across locations, Walmart tends to follow a high/low strategy between stable prices (a behavior that connects with Seim and Sinkinson (2016)'s evidence of high-/low pricing in office supplies).

We formalize these observations using a number of measures that characterize the flexibility in introducing new prices. Table 10 shows that the price of a product lasts on average 3.1 weeks on Amazon, and it lasts 5.9 weeks on Walmart. Similarly, there are 1.6 price changes per distinct price on Amazon, and 3.6 price changes on Walmart. The probability that a retailer-zipcode-product explores a new price is 0.65 in Amazon and 0.27 in Walmart. When similar measures are estimated in the scanner data, the evidence suggests a significantly tighter price grid in offline retailers. The number of weeks per distinct price and the price changes per distinct price are 12.0 and 3.7, respectively. The probability that, conditional on a price change, a store-product pair picks a price not used in the last four weeks is 0.52.

Table 10: Price Grid

		Amazon	Walmart	Scanner
	Product-Zipcode Level			
(1)	Weeks per distinct price[a]	3.1 (0.1)	5.9 (0.2)	12.0 (0.2)
(2)	Price changes per distinct price[b]	1.6 (0.1)	3.6 (0.2)	3.7 (0.04)
(3)	Prob. of new price[c]	0.65 (0.01)	0.27 (0.01)	0.52 (0.01)
	Product Level			
(4)	Daily distinct prices per product[d]	3.7 (0.01)	1.8 (0.01)	-
(5)	Distinct prices per product[e]	12.7 (0.78)	2.2 (0.16)	-

Notes: [a]Defined as the average, across zipcode-products, of the ratio of the number of weeks to the number of distinct prices. [b]Denotes the average, across zipcode-products, of the ratio of number of price changes to the number of distinct prices. [c]Denotes the probability that a retailer-zipcode, conditional on a price change, picks a price not used in that retailer-zipcode-product during the past week. [d]Defined as the average, across products, number of distinct prices on a daily basis for the same product across four common zipcodes. [e]Defined as the average, across products, number of distinct prices for the same product across four common zipcodes over the sample period. Standard errors reported in parenthesis.

Row (4) in Table 10 measures the average number of distinct prices across locations per day. The number of daily distinct prices is 3.7 in Amazon and 1.8 in Walmart. In other words, Amazon's products simultaneously sold in four zipcodes exhibit, on average, 3.7 distinct prices per day. Note that the maximum number can exceed four because prices can change within the day. Row (5) computes the average number of distinct prices per product over the full sample period. The average number of distinct prices is 12.7 and 2.2 in Amazon and Walmart, respectively.

These facts combined indicate that Amazon consistently explores more prices than Walmart. In fact, Amazon displays greater price experimentation for almost every single matched product. Appendix E.4 shows the distribution of the product-level distinct prices by retailer, as well as the product-by-product ratio of distinct prices. The average and median ratio is 7.5 and 6.0, respectively. That is to say, Amazon uses on average 6.5 more distinct prices for the same matched product than Walmart.

5.4 Menu Cost and Many Little Changes

Algorithmic pricing is often characterized as automating the price setting process (Brown and MacKay (2019)). In principle, this automation allows to break the relationship between the menu cost and the size of the price change. In other words, small or large changes are equally "costly". In order to make a connection between the menu cost and the price frequency, we follow Anderson, Jaimovich and Simester (2015)'s approach of measuring the number of variants. In our case, using the product description files (Appendix A.3), we compute the number of distinct UPCs that, in a given year, have the same brand, category, and package size. Products with more variants have a higher in-labor menu cost and thus are less likely to experience price changes (Zbaracki et al. (2004); Anderson, Jaimovich and Simester (2015)). We then map each matched product with the offline-based variants, and explore whether the menu cost is also present in the online channel.

Table 11: Size of Price Changes

		Amazon	Walmart	Scanner
Daily				
(1)	Increase size (%)	12.5 (0.30)	12.7 (0.28)	-
(2)	Decrease size (%)	-10.0 (0.21)	-10.6 (0.33)	-
Weekly				
(1)	Increase size (%)	12.6 (0.15)	13.2 (0.20)	28.9 (0.19)
(2)	Decrease size (%)	-10.6 (0.11)	-10.0 (0.14)	-21.9 (0.13)

Notes: Size of price change denotes the average size of the price change, in percentage terms, conditional on a price change. In order to measure the size of online price changes, which can be multiple for an interval L, we compute the average size of the positive and negative changes for each retailer-zipcode-product-time, and then obtain the average across products. Standard errors reported in parenthesis.

Appendix E.5 shows that a 10% increase in the number of variants reduces the probability of a price change by 9.2 percentage points in offline stores. However, the effect is close to 0 and not significant in the case of online grocers. Consistent with a reduction in the menu cost, Table 11 shows that the size of positive and negative price changes in the offline retailers are about two times larger (in absolute value) than the size of price changes in online retailers. Moreover, the histogram of daily price changes (Appendix E.5) shows that 31% of the price changes are within 5% in absolute value.

6 Conclusion

In this paper, we provide new stylized facts about price setting in the online grocery industry. We collect data from the leading online grocers in the U.S. We focus on a set of products, matched across online and offline retailers, which rules out assortment differences and allows to compare pricing strategies between both channels.

Online grocers exhibit higher non-uniform pricing and higher price differentiation across chains, compared to offline retailers. While online grocers set prices at the delivery zipcode, local demographics moderately drive those differences. Instead, price elasticities, price technology, and shipping distance amplify price dispersion. In closer examination using a high-frequency dataset, we show that algorithmic pricing allows for remarkable flexibility in price setting: intra-day price changes, little synchronization, precise price matching within hours, tiny price changes, and substantial experimentation in prices.

References

Acemoglu, Daron, and Pascual Restrepo. 2018. "Artificial intelligence, automation and work." *NBER w24196*.

Adams, Brian, and Kevin R Williams. 2019. "Zone pricing in retail oligopoly." *American Economic Journal: Microeconomics*, 11(1): 124–56.

Ailawadi, Kusum L, Jie Zhang, Aradhna Krishna, and Michael W Kruger. 2010. "When Wal-Mart enters: How incumbent retailers react and how this affects their sales outcomes." *Journal of Marketing Research*, 47(4): 577–593.

Anderson, Eric, Benjamin A Malin, Emi Nakamura, Duncan Simester, and Jón Steinsson. 2017. "Informational rigidities and the stickiness of temporary sales." *Journal of Monetary Economics*, 90: 64–83.

Anderson, Eric, Nir Jaimovich, and Duncan Simester. 2015. "Price stickiness: Empirical evidence of the menu cost channel." *Review of Economics and Statistics*, 97(4): 813–826.

Aparicio, Diego, and Alberto Cavallo. 2017. "Targeted price controls on supermarket products." *Review of Economics and Statistics*, forthcoming.

Aparicio, Diego, and Roberto Rigobon. 2020. "Quantum Prices." *NBER w26646*.

Arcidiacono, Peter, Paul B Ellickson, Carl F Mela, and John D Singleton. 2019. "The competitive effects of entry: Evidence from supercenter expansion." *American Economic Journal: Applied Economics*, forthcoming.

Ater, Itai, and Oren Rigbi. 2019. "Price transparency, media and informative advertising." *Working Paper*.

Bakos, J Yannis. 1997. "Reducing buyer search costs: Implications for electronic marketplaces." *Management Science*, 43(12): 1676–1692.

Baylis, Kathy, and Jeffrey M Perloff. 2002. "Price dispersion on the Internet: Good firms and bad firms." *Review of industrial Organization*, 21(3): 305–324.

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

Bloomberg. 2018*a*. "Grocery Rivalry Hot as Ever a Year After Amazon-Whole Foods Deal." Craig Giammona, June 14, 2018.

Bloomberg. 2018*b*. "Walmart's Jet Revamps Site to Narrow Focus on Urban Shoppers." Matthew Boyle, September 13, 2018.

Bloomberg. 2019. "Amazon Effect Invigorates Northern Virginia's Housing Market." Shelly Hagan, July 19, 2019.

Boivin, Jean, Robert Clark, and Nicolas Vincent. 2012. "Virtual borders." *Journal of International Economics*, 86(2): 327–335.

Breiman, Leo. 2001. "Random forests." *Machine learning*, 45(1): 5–32.

Brodersen, Kay Henning, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. 2010. "The balanced accuracy and its posterior distribution." 3121–3124, IEEE.

Brown, Jeffrey R, and Austan Goolsbee. 2002. "Does the Internet make markets more competitive? Evidence from the life insurance industry." *Journal of Political Economy*, 110(3): 481–507.

Brown, Zach, and Alexander MacKay. 2019. "Competition in Pricing Algorithms." *Available at SSRN 3485024*.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & Company.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. 2020. "Artificial intelligence, algorithmic pricing and collusion." *American Economic Review*, 110(10): 3267–97.

Campbell, Jeffrey R, and Benjamin Eden. 2014. "Rigid prices: Evidence from us scanner data." *International Economic Review*, 55(2): 423–442.

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. 2018*a*. "More amazon effects: Online competition and pricing behaviors." *NBER w25138*.

Cavallo, Alberto. 2018*b*. "Scraped data and sticky prices." *Review of Economics and Statistics*, 100(1): 105–119.

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.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon. 2014. "Currency unions, product introductions, and the real exchange rate." *The Quarterly Journal of Economics*, 129(2): 529–595.

Chen, Le, Alan Mislove, and Christo Wilson. 2016. "An empirical analysis of algorithmic pricing on amazon marketplace." 1339–1349, International World Wide Web Conferences Steering Committee.

Chevalier, Judith, and Austan Goolsbee. 2003. "Measuring prices and price competition online: Amazon. com and BarnesandNoble. com." *Quantitative Marketing and Economics*, 1(2): 203–222.

Chintagunta, Pradeep K, Junhong Chu, and Javier Cebollada. 2012. "Quantifying transaction costs in online/off-line grocery channel choice." *Marketing Science*, 31(1): 96–114.

CNBC. 2019. "Walmart announces next-day delivery." Lauren Thomas, May 14, 2019.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. 2016. "Using Big Data to Estimate Consumer Surplus: The Case of Uber." *NBER w22627*.

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform pricing in us retail chains." *The Quarterly Journal of Economics*, 134(4): 2011–2084.

Dubé, Jean-Pierre, and Sanjog Misra. 2019. "Personalized Pricing and Customer Welfare." *Working Paper*.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo. 2011. "Reference prices, costs, and nominal rigidities." *American Economic Review*, 101(1): 234–62.

Eizenberg, Alon, Saul Lach, and Merav Yiftach. 2016. "Retail prices in a city." *Working Paper*.

Ellickson, Paul B, and Sanjog Misra. 2008. "Supermarket pricing strategies." *Marketing Science*, 27(5): 811–828.

Ellison, Glenn, and Sara Fisher Ellison. 2009. "Search, obfuscation, and price elasticities on the internet." *Econometrica*, 77(2): 427–452.

Financial Times. 2020. "US online grocery shopping jumps as chains rush to add capacity." Dave Lee, June 01, 2020.

Fisher, Marshall, Santiago Gallino, and Jun Li. 2017. "Competition-based dynamic pricing in online retailing: A methodology validated with field experiments." *Management Science*, 64(6): 2496–2514.

Forman, Chris, and Avi Goldfarb. 2020. "Concentration and Agglomeration of IT Innovation and Entrepreneurship: Evidence from Patenting." *NBER w27338*.

Fortune. 2019. "It's Not Just Amazon: Free One-Day Delivery Is The New Normal." Laura Stampler, April 30, 2019.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The elements of statistical learning.* Vol. 1, Springer series in statistics New York.

Goolsbee, Austan D, and Peter J Klenow. 2018. "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce." *AEA Papers and Proceedings*, 108: 488–92.

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.

Gorodnichenko, Yuriy, Viacheslav Sheremirov, and Oleksandr Talavera. 2014. "Price setting in online markets: Does IT click?" *Journal of the European Economic Association*.

Hannak, Aniko, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. "Measuring price discrimination and steering on e-commerce web sites." 305–318, ACM.

Hitsch, Günter J, Ali Hortacsu, and Xiliang Lin. 2019. "Prices and promotions in us retail markets: Evidence from big data." *NBER w26306*.

Houde, Jean-François, Peter Newberry, and Katja Seim. 2017. "Economies of density in e-commerce: A study of amazon's fulfillment center network." *NBER w23361*.

Hwang, Minha, Bart J Bronnenberg, and Raphael Thomadsen. 2010. "An empirical analysis of assortment similarities across US supermarkets." *Marketing Science*, 29(5): 858–879.

Ilut, Cosmin, Rosen Valchev, and Nicolas Vincent. 2020. "Paralyzed by fear: Rigid and discrete pricing under demand uncertainty." *Econometrica*, 88(5): 1899–1938.

Instacart. 2019. "Instacart com help." Instacart com (Help Pricing), Jan 24, 2019.

Jank, Wolfgang, and PK Kannan. 2005. "Understanding geographical markets of online firms using spatial models of customer choice." *Marketing Science*, 24(4): 623–634.

Kaplan, Greg, Guido Menzio, Leena Rudanko, and Nicholas Trachter. 2019. "Relative price dispersion: Evidence and theory." *American Economic Journal: Microeconomics*, 11(3): 68–124.

Klenow, Peter J, and Benjamin A Malin. 2010. "Microeconomic evidence on price-setting." In *Handbook of Monetary Economics*. Vol. 3, 231–284. Elsevier.

Levy, Daniel, Dongwon Lee, Haipeng Chen, Robert J Kauffman, and Mark Bergen. 2011. "Price points and price rigidity." *Review of Economics and Statistics*, 93(4): 1417–1431.

McShane, Blakeley B, Chaoqun Chen, Eric T Anderson, and Duncan I Simester. 2016. "Decision stages and asymmetries in regular retail price pass-through." *Marketing Science*, 35(4): 619–639.

Miklós-Thal, Jeanine, and Catherine Tucker. 2019. "Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers?" *Management Science*, 65(4): 1552–1561.

Nair, Harikesh. 2007. "Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games." *Quantitative Marketing and Economics*, 5(3): 239–292.

Nakamura, Emi. 2008. "Pass-through in retail and wholesale." *American Economic Review*, 98(2): 430–37.

Nakamura, Emi, and Jon Steinsson. 2008. "Five facts about prices: A reevaluation of menu cost models." *The Quarterly Journal of Economics*, 123(4): 1415–1464.

Nakamura, Emi, and Jon Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annual Review of Economics*, 5(1): 133–163.

NBER. 2017. "ZIP Code Distance Database." https://www.nber.org/data/zip-code-distance-database.html.

New York Times. 2018. "Amazon's Ripple Effect on Grocery Industry." Erin Griffith, August 21, 2018.

New York Times. 2020. "FreshDirect, After Broken Eggs and Angry Customers, Stages a Comeback." David Yaffe-Bellany, February 21, 2020.

Nijs, Vincent, Kanishka Misra, Eric T Anderson, Karsten Hansen, and Lakshman Krishnamurthi. 2010. "Channel pass-through of trade promotions." *Marketing Science*, 29(2): 250–267.

Orbach, Barak Y, and Liran Einav. 2007. "Uniform prices for differentiated goods: The case of the movie-theater industry." *International Review of Law and Economics*, 27(2): 129–153.

Overby, Eric, and Chris Forman. 2015. "The effect of electronic commerce on geographic purchasing patterns and price dispersion." *Management Science*, 61(2): 431–453.

Pozzi, Andrea. 2012. "Shopping cost and brand exploration in online grocery." *American Economic Journal: Microeconomics*, 4(3): 96–120.

Pozzi, Andrea. 2013. "The effect of Internet distribution on brick-and-mortar sales." *The RAND Journal of Economics*, 44(3): 569–583.

Reuters. 2017. "Amazon's Name Pops Up on 10% of U.S. Earnings Conference Calls." July 28, 2017.

Seim, Katja, and Michael Sinkinson. 2016. "Mixed pricing in online marketplaces." *Quantitative Marketing and Economics*, 14(2): 129–155.

Shiller, Benjamin Reed, et al. 2014. "First-degree price discrimination using big data." *Federal Trade Commission*.

Stevens, Luminita. 2020. "Coarse pricing policies." *The Review of Economic Studies*, 87(1): 420–453.

U.S. Census. 2019*a*. "American Community Survey 2014-2018." https://www.census.gov/programs-surveys/acs/.

U.S. Census. 2019*b*. "U.S. Gazetteer Files." https://www.census.gov/geo/maps-data/data/gazetteer.html.

U.S. Census Bureau. 2020. "Quarterly retail e-commerce sales: 4Q 2019." Feb 19, 2020.

U.S. Department of Defense. 2014. "World Geodetic System 1984." Its Definition and Relationships with Local Geodetic Systems.

Wall Street Journal. 2016. "Wal-Mart, Kroger Strive to Counter Amazon Grocery Challenge." Heather Haddon and Sarah Nassauer, October 12, 2016.

Wall Street Journal. 2018. "Inside FreshDirect's Big Bet to Win the Home-Delivery Fight." Jennifer Smith, July 18, 2018.

Walmart. 2020. "Walmart com price matching policy." Walmart com Help, Jan 26, 2020.

Zbaracki, Mark J, Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen. 2004. "Managerial and customer costs of price adjustment: direct evidence from industrial markets." *The Review of Economics and Statistics*, 86(2): 514–533.

Zillow Research. 2018. "Zillow Home Value Index." https://www.zillow.com/research/data/.