Q Open, 2022, **3**, 1–20
https://doi.org/10.1093/qopen/qoac021
Advance access publication date: 5 August 2022
Article

OXFORD

Within-retailer price dispersion in e-commerce: Prevalence, magnitude, and determinants

Svetlana Fedoseeva[1] and Judith Irek[2,*]

[1]University of Bonn, Nussallee 21, DE-53115 Bonn, Germany
[2]Agroscope, Tänikon 1, CH-8356 Ettenhausen, Switzerland

*Corresponding author: Agroscope, Tänikon 1, CH-8356 Ettenhausen, Switzerland.
E-mail: judith.irek@agroscope.admin.ch

Received: March 22, 2022. **Accepted:** July 6, 2022

Abstract

Whereas between-retailer price discrepancies are well documented, less is known about price differentiations within a single chain. This article investigates the prevalence and the magnitude of online within-retailer price dispersion over time. Amazon is the US largest online food retailer (and the second largest retailer in the entire US grocery market), and we use the rich daily price data on food and beverages sold by Amazon Fresh in New York City and Los Angeles to shed light on the prevalence, magnitude, and determinants of the within-retailer price dispersion over time. We show that differences in economic indicators, competitive pressure, and COVID-19 exposure across locations contribute to price dispersion. Once those factors are controlled for, we observe a negative linear time trend in the share and magnitude of non-identical prices confirming an increasing market integration as e-commerce matures.

Keywords: Electronic commerce, Grocery sector, Online retailing, Price dispersion.

JEL codes: L81, C55, D22

1. Introduction

Although information economics predicts that online retailing eventually brings about a reduction in price dispersion (Bakos 1997), an increased market efficiency (Biswas 2004), and a nearly competitive market (Brynjolfsson and Smith 2000), price discrepancies are still observed in domestic and international markets alike (Anania and Nisticò 2014; Duch-Brown *et al.* 2021). Strategic pricing and quality signalling (Wang and Li 2020), consumer loyalty (Reichheld and Schefter 2000), information overload (Grover, Lim, and Ayyagari 2006), price adjustment costs (Böheim, Hackl, and Hölzl-Leitner 2021), or use of price promotions (Cavallo 2017) keep prices for the same products across retailers from being identical even online.

Even if between-retailer price dispersion has not ceased with the development of e-commerce (Gorodnichenko, Sheremirov, and Talavera 2018), the online market maturation seems to have guided companies towards uniform pricing across their own physical and online stores. Although recent developments in information technology and the increasing popularity of e-commerce provide sellers with access to consumer information that might create favourable conditions for the application of third-degree price discrimination (e.g. Baker, Marn, and Zawada 2001; Fudenberg and Villas-Boas 2012), lower managerial decision-making costs, as well as online transparency and fairness concerns (Cavallo 2018; Mookherjee, Lee, and Sung 2021), seem to be powerful arguments for identical pricing in all channels of multi-channel retailers. While international companies tend to use uniform pricing within currency unions (Cavallo, Neiman, and Rigobon 2014), online disclosure of prices reduces price dispersion in traditional offline retailing (Ater and Rigbi 2018). Even for geographically segmented markets with different income levels, price variations within chains are negligible compared with price dispersion among stores of different chains (DellaVigna and Gentzkow 2019). Using a dataset of the twenty largest multi-channel companies in ten countries, Cavallo (2017) finds that most hybrid retailers have a single price online regardless of the location of the buyer. Offline stores of hybrid retailers have identical within-same-chain-store prices in 78 per cent, while online and offline price levels are uniform in 72 per cent of the cases. Cavallo (2018) further confirms that uniform pricing is also shared by many online retailers, including Amazon, whose cross-section average share of uniform prices is higher than the sample average (91 versus 78 per cent in the whole sample). In the USA, Amazon seems to be a relevant force in promoting uniform pricing: Products that can be found on Amazon are more likely to be priced identically by Walmart, Amazon's main competitor in the US market, in multiple locations. More recently, Aparicio, Metzman, and Rigobon (2021) show that offline retailers exhibit more uniform pricing than online retailers within the chain and across locations. In their sample, uniform prices offline were observed in 78 per cent of the cases within (and 63 per cent across) states, whereas the respective shares in online data were 66 and 40 per cent. The within-retailer price dispersion was higher for fresh produce and packaged food than for personal care items and cleaning supplies. According to Cavallo (2018), most price differences across geographical locations occur in food and beverages, the sector with the lowest share of online sales. However, the share of uniform grocery prices at Amazon is higher than that at other retailers (84 versus 76 per cent), and the authors expect geographical price dispersion to go down as e-commerce in food retailing develops.

Back in 2018, Amazon was the eighth largest player in US grocery retailing. The relevance of electronic commerce for food retailing and the role of Amazon have dramatically increased since that time. The COVID-19 pandemic triggered a general surge in online sales in the grocery sector. On top of that, Amazon has come out as a major winner of the pandemic-driven demand boom for food and beverages, showing the highest absolute and relative increase in sales among the top fifty US grocery retailers (Troy 2021). As of 2021, Amazon is America's second-largest grocery retailer after Walmart.

Those recent developments in the US grocery sector call for an update on Amazon's price dispersion estimates. So far, price dispersion within retailers has received little attention in the literature, and there are only a few attempts to not only detect but also explain the within-retailer dispersion. The commonly assumed drivers of price dispersion that might be useful for price comparisons across retailers seem to be less relevant for explaining price dispersion within retailers. Although Amazon's presence seems relevant for reducing the price dispersion in online food retailing, there is no information on the drivers of price differences within Amazon. Has price dispersion vanished? Which factors can explain within-retailer price dispersion over time?

This study provides insights into price dispersion for a wide range of grocery products offered by Amazon, the US largest online retailer, in two geographic locations over time. Using web scraping methods, we collected price quotes for products that were simultaneously available at Amazon Fresh in New York City (ZIP Code 10001) and Los Angeles (ZIP Code 90001). Amazon Fresh is Amazon's full-assortment grocery subsidiary that operates in the USA and worldwide and is now used to consolidate the various food-related businesses of Amazon. The matched dataset contains prices for 18,470 products simultaneously offered in

Table 1. Distribution of prices, US dollars.

Location	Mean	Standard deviation	Min.	Max.	Percentile				
					5th	25th	50th	75th	95th
New York City	4.37	2.98	0.13	196.06	1.27	2.69	3.92	5.24	8.99
Los Angeles	4.21	2.93	0.17	131.60	1.19	2.50	3.69	4.99	8.79

both locations. With twelve grocery product groups included, and roughly 2,200,000 pairs of price quotes sampled between November 2017 and September 2020, our data belong to the most recent, detailed, and comprehensive datasets for online food prices available at the moment.

Our results indicate a lower share of uniform prices across locations than earlier studies suggested. Whereas in Cavallo (2018), 84 per cent of grocery prices offered by Amazon in different locations were identical, in our sample the share is 27 per cent. Furthermore, we show that the prevalence and magnitude of price differences across geographic locations over time are sensitive to differences in the economic situation, competitive pressure, and pass-through of external shocks. When those factors are controlled for, there is a negative trend associated with the prevalence of price dispersion, suggesting that ceteris paribus the share and to a lesser extent the magnitude of unequal prices go down over time.

While concentrating on a single company allows us to clearly isolate within-chain from between-retailer price dispersion, the narrow focus on a single market player, however large and relevant, results in a limited generalization capacity of our findings that cannot simply be extended to other market participants or geographic locations. We acknowledge the limitations of such an approach, yet we firmly believe that understanding Amazon's food pricing is an important step in evaluating possible effects that Amazon's decisions might have on general grocery retailing prices and beyond.

2. Price data

For our analysis, we collected daily price data for foods and beverages simultaneously available[1] at Amazon Fresh in two distinct locations: New York City (ZIP Code 10001) and Los Angeles (ZIP Code 90001). The data were collected for the following twelve product groups: meat and fish, produce (fuits and vegetables), deli, dairy, frozen products, bread and bakery, pantry, snacks, beverages, baby food, sweets, and breakfast. To collect online prices, we used a Phyton scraping algorithm that accessed the Amazon Fresh website every day at the same time to avoid possible influences related to intra-day price changes (see Hillen 2019, for more technical details). The collected prices are the posted promotional prices excluding delivery costs. Any discounts that were available to all customers were included. The prices were collected for almost three years. The earliest available price quotes are from 23 November 2017, and the latest from 16 September 2020. The data are unbalanced; on some days, observations were missing owing to product availability or technical issues in the data collection process. On some days, the scraping algorithm could not fully gather the data or returned empty because the script did not run through correctly, e.g. due to changes in the Amazon Fresh website, bugs in the script, or unknown temporary technical issues. When products identified as miscellaneous and double entries are excluded, we have almost 2.2 million price quotes for 18,470 products available for each location. Table 1 shows summary statistics for price quotes in New York City and Los Angeles. The median product price in the sample is 3.92 US dollars in New York City and 3.69 US dollars in Los Angeles. About a quarter of all products have prices below 2.70 US dollars in both cities; less than 5 per cent of goods have prices higher than 9 US dollars, and less than 1 per cent have prices higher than 14 US dollars. Few products have substantially higher prices: The maximum prices in New York City and Los Angeles are 196 and 132 US dollars, respectively, and belong to the meat and fish and produce categories.

Table 2 compares price levels between New York City and Los Angeles for each product group. The share of identical prices is 27 per cent for all pooled observations across product groups (column 2). Frozen products, snacks, bread and bakery, and dairy products have the lowest share of identical prices (only 19–24 per cent of prices are the same across locations). Baby foods and sweets have the highest share of identical prices (47 and 41 per cent, respectively). When prices in New York City and Los Angeles differ, they are more often higher in New York City (columns 3–4). This is especially true for product groups with low shares of identical prices. In the pooled sample, 46 per cent of prices are higher in New York City, whereas 27 per cent of prices are higher in Los Angeles.

The average size of the price differences across cities is shown in columns 5 and 6. Column 5 reports an average price difference, which includes cases with no price difference across locations. Column 6 excludes cases where prices are identical. The price differences are relatively small, at roughly 5 per cent of the sample average price when identical prices are excluded. The positive values imply that, on average, prices are higher in New York City than in Los Angeles. This is the case for all product groups except meat and fish. When identical prices are included, the average price difference drops to roughly 4 per cent of the sample average price.

In this article, we are not merely interested to find out whether prices on average are higher in a particular location, but rather whether (and why) prices differ across locations. However, the simple price averages can hide the true prevalence of price dispersion because they might cancel out the effects of prices that are higher in New York City versus prices that are higher in Los Angeles. Imagine, for instance, a situation in which one product is 10 US dollars more expensive in New York City, and another product is 10 US dollars more expensive in Los Angeles: Although prices for both products deviate across locations, a simple average price difference equals zero. To avoid such cancel-out effects, columns 7–8 report the average values of absolute price differences across locations. These values are considerably larger than the average price differences in columns 5–6.

Columns 9–10 report the per cent difference in absolute value between two prices. The highest price dispersion is observed in bread and bakery, pantry, produce, and snacks, whereas the lowest values are in baby foods, sweets, and breakfast. The sample average relative price difference is 22 per cent when identical prices are excluded and 16 per cent when they are included. Column 11 shows the maximum value of the (absolute) price difference for each product group. Some of the differences across locations are extremely high given the magnitudes of the average and the maximum prices in our sample, and we test the sensitivity of our results to the exclusion of the extreme values in the empirical part (Section 4).

3. Determinants of price dispersion

The determinants of price dispersion have been studied intensively in the empirical literature. The most prominent finding is well summarized by Berardi, Sevestre, and Thébault (2017), who studied a large dataset of French supermarkets and came to the conclusion that price dispersion mostly results from between-retailer heterogeneity in retail chains' national pricing. In local markets, however, demand and local competitions are also important factors that affect prices and hence price dispersion, also within chains of a single retailer. According to Aparicio, Metzman, and Rigobon (2021), an increase in the demand or income gap between two locations is expected to lead to a higher share of non-uniform prices and a higher magnitude of price dispersion. Gorodnichenko, Sheremirov, and Talavera (2018) show that the intensity of direct online competition affects Amazon's regional price

Table 2. Price differences across product groups.

Product group	Number of observations (1)	Share of prices			Mean price difference (USD)		Absolute mean price difference (USD)		Absolute mean price difference (per cent)		Max. absolute price difference (USD) (11)
		Identical (2)	Higher in NY (3)	Higher in LA (4)	(2) incl. (5)	(2) excl. (6)	(2) incl. (7)	(2) excl. (8)	(2) incl. (9)	(2) excl. (10)	
Meat and fish	179,310	0.36	0.33	0.31	−0.02	−0.04	1.01	1.59	14.46	22.63	126.05
Produce	140,704	0.39	0.40	0.21	0.13	0.21	0.79	1.30	19.98	32.82	192.07
Deli	84,687	0.29	0.41	0.30	0.08	0.11	0.60	0.84	12.01	16.90	21.39
Dairy	513,928	0.24	0.45	0.31	0.04	0.05	0.61	0.80	14.82	19.54	112.24
Frozen	384,648	0.19	0.56	0.25	0.29	0.36	0.77	0.96	19.86	19.97	68.24
Bread and bakery	95,782	0.23	0.50	0.27	0.30	0.39	0.90	1.16	24.33	27.64	28.81
Pantry	366,038	0.30	0.45	0.25	0.17	0.24	0.59	0.83	24.39	24.94	98.01
Snacks	237,365	0.20	0.55	0.24	0.33	0.41	1.07	1.34	21.10	26.49	70.95
Beverages	97,407	0.38	0.40	0.22	0.16	0.25	0.51	0.82	10.46	16.78	164.20
Baby food	19,919	0.47	0.30	0.23	0.07	0.13	0.29	0.54	7.41	14.07	39.98
Sweets	40,793	0.41	0.37	0.22	0.08	0.14	0.33	0.58	9.58	16.35	49.95
Breakfast	33,897	0.38	0.41	0.21	0.16	0.25	0.43	0.69	10.76	17.31	25.62
Total	2,194,478	0.27	0.46	0.27	0.16	0.22	0.73	1.00	16.24	22.31	192.07

dispersion. Cavallo (2018), too, expects price dispersion to be higher when economic discrepancies across locations increase.

A brief analysis of the US grocery retailers' communications further confirms that fierce competition, consumer demand, and the COVID-19 outbreak in 2020 are the major factors that affect the sector (Amazon 2021; Costco 2021; Target 2021; Walmart 2021) and might hence influence food pricing decisions and price dispersion. We use these theories and industry insights to explain the prevalence and magnitude of within-Amazon price dispersion over time. The prevalence of price dispersion is characterized by the share of non-identical prices in each price pair. *Share* is a binary variable that takes the value of 1 if the prices for each product offered at two locations on the same day are not identical, and 0 if prices are uniform

$$Share_{i,j,t}^{NY,LA} = 1\ if\ p_{i,j,t}^{NY} \neq p_{i,j,t}^{LA};\ 0\ otherwise, \tag{1}$$

where $p_{\mathrm{NY}}^{i,j,t}$ denotes the price of product i in a product category j in location *NY (LA)* at time t.

We define *Magnitude* of price dispersion as the per cent difference in absolute value between two prices at a product and time level across two locations

$$Magnitude_{i,j,t}^{NY,LA} = \frac{\left|p_{i,j,t}^{NY} - p_{i,j,t}^{LA}\right|}{\left(p_{i,j,t}^{NY} + p_{i,j,t}^{LA}\right)/2} *100. \tag{2}$$

We regress our measures of price dispersion, *Share* or *Magnitude*, for a product pair i from a category j at a time t on the *absolute differences* in income, inflation, competitive pressure, and the intensity of the COVID-19 outbreak between the regions and a set of time-related variables

$$\begin{aligned} Dispersion_{i,j,t}^{NY,LA} = {} & a + b \cdot \left|\Delta Income_t^{NY,LA}\right| + c \cdot \left|\Delta Inflation_t^{NY,LA}\right| \\ & + d \cdot \left|\left(\sum_{1}^{3} Competition_t^{NY,LA}\right)\right| + e \cdot \left|\Delta Covid_t^{NY,LA}\right| + f \cdot Trend_t + g \\ & \cdot NoFee_t + \sum_{l=2}^{12} h_l \cdot Month_t + \sum_{m=2}^{7} k_m \cdot Weekday_t + n_i + u_{ijt} \end{aligned} \tag{3}$$

Table 3 outlines our expectations regarding the sign of the estimated coefficients. Appendix A reports some descriptive statistics for our independent variables and underlying raw data. *Income* is approximated by personal income per capita in the states of New York and California, measured in thousands of current US dollars from the US Bureau of Economic Analysis. *Inflation* (food or total) is sourced from the US Bureau of Labor Statistics. The data are monthly and reported at the city level. The *Competition* variable is calculated as the absolute difference between the total number of Walmart, Costco, and Target stores (main Amazon competitors in the US grocery sector) in the states of New York and California. It can be decomposed into variables of individual competitive pressure of those chains. Those competition-related data are annual and available at the state level only. *Covid* is the daily number of new infections. The data are taken from the Los Angeles County COVID-19 Surveillance Dashboard and the New York City Department of Health and Mental Hygiene. In the empirical part, we use the absolute difference in those indicators between both locations. *Trend* is a time trend of a form 0, 1, 2… n. The monthly Amazon Fresh subscription fee has been lifted in the USA since the last week of October 2019. *NoFee* is a binary variable that takes the value of 1 from 30 October 2019, and is 0 otherwise. Before the fee was

Table 3. Expected signs of estimated coefficients.

Coefficient	Variable	Hypothesis	Sign
b	$\Delta Income$	Higher income heterogeneity results in more discrepancies in consumer preferences and demand, and willingness-to-pay for the same products, hence more price dispersion	+
c	$\Delta Inflation$	Inflation hinders income effects, hence is expected to negatively affect price dispersion	−
d	$\Delta Competition$	Price dispersion increases with an increasing difference in the competitive pressure across locations	+
e	$\Delta Covid$	An increased demand for online grocery products due to COVID-19 restrictions depends on the timing and intensity of the COVID spread at each location. If so, the higher discrepancy in new infection numbers across locations might result in higher price dispersion over time On the other hand, nationwide shocks—e.g. delivery bottlenecks—could put additional pressure on prices, increase the degree of pass-through and negatively affect price dispersion	±
f	$Trend$	Online market maturation is expected to make grocery markets more efficient and reduce price dispersion over time Consumer loyalty or information overload online might have an opposite effect on price dispersion Which effect prevails in the current US food market is an empirical question	±
g	$NoFee$	Reducing entry barriers for new consumers is expected to boost market efficiency and transparency, affecting price dispersion negatively	−

eliminated, one had to be an Amazon Prime member and pay an additional monthly fee of 14.99 US dollars to use Amazon Fresh services.

We include the full set of product-level fixed effects, n, in our estimation model to account for time-invariant heterogeneity of prices as well as monthly and weekday dummies to capture possible seasonal/promotional effects in price dispersion. Both equations are evaluated using the *reghdfe* estimator (Correira 2015) using Stata 15. The major advantage of this approach is the ability to deal with large datasets and a high number of involved fixed effects.

Table 4 reports estimation results for the full sample. The increasing income heterogeneity between two locations is associated with higher price dispersion. An increase in the personal income gap between states by 1,000 US dollars is linked to an increase in the share of non-identical prices (magnitude of price dispersion) between New York City and Los Angeles by about 2.7 (4.7) per cent. The income gap between the two locations declined by about 2,000 US dollars over the sample period, contributing to a lower price dispersion over time, ceteris paribus. A one-unit change in the difference in inflation across the cities results in a decline in the magnitude of price dispersion of 1.18 per cent. The effect of inflation on the share of non-identical products is negligible although statistically significant (−0.3 per cent).

The competition intensity variable is positive and statistically significant in both equations, suggesting a positive link between price dispersion and the difference in the number of competitors' outlets available in both locations. The price dispersion prevalence and magnitude increase by 1.4–1.6 per cent, other factors held constant, with each additional store that augments the difference in the competitor's presence in both locations. In an additional

Table 4. Determinants of price dispersion.

	Share	Magnitude
Δ*Income*	0.027***	4.696***
	(0.00)	(0.05)
Δ*Inflation*	−0.003***	−1.184***
	(0.00)	(0.01)
Δ*Competition*	0.016***	1.390***
	(0.00)	(0.01)
Δ*Covid* ($\times$1000)	−0.010***	−1.855***
	(0.00)	(0.02)
Trend ($\times$1000)	−0.193***	−1.256***
	(0.00)	(0.36)
NoFee	−0.207***	−23.441***
	(0.00)	(0.09)
Constant	−7.869***	−720.967***
	(0.14)	(6.27)
Adj. R-squared	0.43	0.28
Number of products	18,216	18,216
Number of observations	2,194,224	2,194,224

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variables are the share of non-identical prices (*Share*) and the absolute mean price difference in per cent (*Magnitude*) for each price pair. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported. The estimated coefficients for Trend and COVID are small, and we multiply them by 1,000 to facilitate their visibility here and in the following tables. The resulting effect is to be interpreted as the change in the measure of price dispersion due to a 1,000 case increase in the difference in new infections (or over 1,000 days for the trend).

regression, we substituted the aggregated competition variable by individual covariates for Walmart, Target, and Costco (results are not reported here but are available on request). The difference in the number of Walmart and Costco outlets seems to have the highest impact on the share and the size of the price difference per store (about 2 per cent in the share equation and 3.0–3.5 per cent in the magnitude equation, respectively). The number of Costco stores in New York did not change over the sample period, whereas California added an extra six stores, increasing the gap in competition intensity between both locations and contributing to a larger magnitude of price differences and a higher share of non-identical prices. New Walmart stores opened and closed in both locations over the sample period. The difference in the number of stores declined by two between 2017 and 2020, contributing, ceteris paribus, to a lower magnitude of price dispersion over time. Target has the lowest coefficient estimate (below 1 per cent in both equations) and has been expanding rapidly in both states. New York added eight stores and California added twenty-four stores over the sample period. The difference across locations hence increased by sixteen stores, adding to the frequency of non-identical prices and magnitude of price differences.

The impact of the COVID-19 spread is statistically significant, and its effect on price dispersion is negative. When the difference in new infections across locations increases by 1,000 cases, the share of non-identical prices goes down by about 1 per cent, and the magnitude of price dispersion decreases by about 1.85 per cent, other factors held constant. During the first outbreak in New York, the difference in new daily infections reached over 5,000 cases. The negative sign of the Covid variable estimate suggests the COVID-19 outbreak contributed to a more complete aggregate shock pass-through to consumer prices, leading to more uniform pricing and smaller price gaps between New York and Los Angeles.

The negative coefficient of the trend variable supports the economics of information approach predictions that expect more efficient markets and less price dispersion as online markets mature. Both measures of price dispersion seem to be declining over time. The

Table 5. Results from regressions without multivariate outliers (BACON).

	Share		Magnitude	
	$P = 0.30$	$P = 0.40$	$P = 0.30$	$P = 0.40$
$\Delta Income$	0.027***	0.026***	4.717***	4.665***
	(0.00)	(0.00)	(0.05)	(0.05)
$\Delta Inflation$	−0.003***	−0.003***	−1.183***	−1.176***
	(0.00)	(0.00)	(0.01)	(0.01)
$\Delta Competition$	0.016***	0.017***	1.391***	1.373***
	(0.00)	(0.00)	(0.01)	(0.01)
$\Delta Covid$ (×1000)	−0.010***	−0.011***	−1.860***	−1.986***
	(0.00)	(0.00)	(0.02)	(0.02)
Trend (×1000)	−0.193***	−0.194***	−1.176***	−0.819***
	(0.00)	(0.00)	(0.36)	(0.35)
NoFee	−0.207***	−0.207***	−23.492***	−23.35***
	(0.00)	(0.00)	(0.09)	(0.09)
Constant	−7.868***	−7.876***	−721.330***	−711.899***
	(0.14)	(0.14)	(6.26)	(6.149)
Adj. R-squared	0.43	0.43	0.28	0.28
BACON outliers	290	2,684	290	2,684
Number of products	18,214	18209	18,214	18,209
Number of observations	2,193,935	2,191,542	2,193,935	2.191,542

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variables are the share of non-identical prices (*Share*) and the absolute mean price difference in per cent (*Magnitude*) for each price pair. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at an individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported.

share of non-identical prices goes down by roughly 19 per cent per 1,000 days or about 7 per cent per year. The decline in the magnitude of price differences is lower, about 1.25 per cent per 1,000 days or 0.5 per cent per year.

The negative association is also observed following the Amazon Fresh fee elimination: The magnitude of price differences declined on average by 23 per cent compared with the period before the fee elimination. The share of non-identical prices declined by about 21 per cent. In the next section, we assess the robustness of our results to the elimination of extreme price differences or the omission of a particular product group from the sample.

4. Robustness

Our sample contains some extremely large price differences. For instance, while the average absolute price difference is 0.73 US dollars and the average price in the sample is 4.29 US dollars, 180 price differences are larger than 50 US dollars. Sixty-eight price differences are larger than 100 US dollars. The majority of large price deviations are in pantry, meat and fish, snacks, and dairy. Almost all price differences over 100 US dollars are from just two product groups: meat and fish and dairy.

To deal with the outliers, we first use the blocked adaptive computationally efficient outlier nominator (BACON) algorithm proposed by Billor, Hadi, and Velleman (2000) to identify outliers in multivariate data. There are no BACON outliers at the $P = 0.10$ and only three at the $P = 0.20$, and we report the results of estimations with eliminated outliers at higher percentile levels ($P = 0.30$ and $P = 0.40$) in Table 5.

Alternatively, we repeat the regressions with the sub-samples in which the largest price differences (in absolute terms) are eliminated. We set our cut-off values at 0.005, 0.05, and

Table 6. Descriptive statistics for absolute price differences (in US dollars) across locations for various cut-off values.

Top	Observations	Mean	Standard deviation	Min.	Max.
0.005 per cent	2,194,364	0.72	1.59	0.00	65.94
0.05 per cent	2,193,379	0.70	1.39	0.00	26.66
0.5 per cent	2,183,486	0.64	0.99	0.00	8.90

0.5 per cent of the sample, so that the maximum price gap between two locations fluctuates all the way from 66 US dollars to just below 10 US dollars (Table 6).

Our conclusions regarding the determinants of price dispersion prevalence and magnitude remain largely unaffected with exclusion of extreme values or alternative cut-off values (Appendix B). The outcomes in the *Share* equation remain unaffected. Eliminating the largest price differences somewhat reduces the size of estimated coefficients in the *Magnitude* equation but does not alter their signs or statistical significance.

To test if our results are driven by a particular product group, we recursively drop one product group at a time from the estimation and compare results (Appendices C and D). Most individual coefficients are robust to sample adjustments, and their sign and magnitude remain largely unchanged in comparison to the base model discussed in the previous section. The *Trend* variable in the *Magnitude* equation is the most sensitive of all model covariates to product exclusion, which hints at heterogeneous paths of price dispersion development over time for different product categories. The magnitude of price dispersion declines much faster when frozen or fresh products are excluded from the analysis. These product groups are much more difficult to arbitrage geographically, and price dispersion is the most persistent.

Finally, because not all product groups' price information is available throughout the entire sample period, we exclude products that only appeared in the dataset in 2020. Those include beverages, baby food, sweets, and bread and bakery. The results remain largely unaffected in terms of signs and magnitude, and we do not report them here.

5. Discussion and conclusions

Recent studies increasingly argue that uniform pricing is becoming the rule for US retailers (Gunn 2019; Stambor 2018), even if customized prices across different geographic locations are not generally prohibited by law. Cavallo (2018) attributes the high share of identical prices of multi-channel retailers to Amazon's presence on the market (the so-called Amazon Effect). In his sample, the majority of products had the same posted price regardless of the buyer's location, and the share of uniform prices was higher if the product could be found on Amazon. Amazon's own prices were identical at 91 per cent on average. Food prices exhibited more price dispersion at the time (84 per cent of Amazon's prices were identical), but the share of uniform prices for grocery products was expected to grow as e-commerce matures. Since then, Amazon has become America's second-largest grocery retailer, and its importance grew stronger owing to pandemic restrictions and higher shares of online purchases. Grocery e-commerce (at least as represented by its leader) seems to have matured, but have these developments also led to a decline in within-retailer price dispersion?

The goal of our analysis was to answer this question by focusing on Amazon's own prices. We used a rich dataset of daily Amazon Fresh prices for identical products in New York City and Los Angeles collected for almost three years (2017–2020) to measure the prevalence and the magnitude of price dispersion. We found that the share of non-uniform prices is higher in our dataset than in earlier findings. Whereas about 27 per cent of prices are identical in our data, the share of uniform prices for grocery products across states is 84 per cent in Cavallo (2018) and about 40 per cent in Aparicio, Metzman, and Rigobon (2021).

The highest share of uniform prices in our sample is in baby food (47 per cent), and only 19 per cent of frozen products' prices are identical. Our estimates of the magnitude of price differences across locations are also larger (the mean difference is roughly 16–22 per cent), because we did not focus on the simple price averages but instead analysed absolute price differences. Using simple means underestimates the scale of price dispersion by allowing opposite price differences in different locations to cancel each other out.

Whereas Cavallo (2018) scraped the prices from Amazon.com, we focused on the Amazon subsidiary Amazon Fresh, which is increasingly promoted by Amazon as its main food service (Schader 2021). It might be the case that Amazon Fresh allows for more regional price dispersion because it focuses on fresh and chilled products, whose prices are more difficult to arbitrage than those of pantry products and beverages, which make up a large share of the Amazon.com assortment. Amazon Fresh also requires customers to enter their ZIP code before any products or prices are shown. This might result in less price transparency than at Amazon.com, which allows for easy price comparison, including via price comparison web engines. Given that our results on the magnitude of price dispersion are closer to those of Aparicio, Metzman, and Rigobon (2021), who also used Amazon Fresh data in their analysis, the choice of a retailer's division seems to be an important determinant of price dispersion within Amazon. Unfortunately, little is known about potential price differences between Amazon's own grocery distribution channels, and more research is needed to shed light on this issue.

Going one step further, we identify possible drivers of within-retailer price dispersion and quantify their effects on the share and magnitude of price deviations. Our analysis suggests that differences in income across regions, competition intensity, as well as COVID-19 exposure contribute to regional price dispersion within Amazon. In line with Aparicio, Metzman, and Rigobon (2021), there is a positive association between the prevalence and the magnitude of price dispersion and the gap in personal income across locations: The more different in terms of income the regions become over time, the higher is the expected price dispersion across locations. In our sample, the declining gap in regional income in 2017–2020 seems to have contributed to reduced price dispersion. A note of caution is due here. Our income variables are specified at the state level. The price data, however, were collected for particular ZIP codes. For our analysis, this implies that, on average, the New York City neighbourhood is considerably richer than the chosen neighbourhood in Los Angeles. These local average income differences are higher than aggregated state differences that enter our empirical specification. As a result, the true income effect on price dispersion might be even larger than discussed, but the unavailability of the relevant ZIP-code-level income data prevents us from formally testing this issue. The real income effect on price dispersion is somewhat lower as the inflation gap is negatively associated with price dispersion.

In line with Gorodnichenko, Sheremirov, and Talavera (2018), we found that competition plays an important role in price dispersion, even within a single retailer. Not only the intensity of direct online competition affects Amazon's regional price dispersion (the result we did not directly quantify), but also the difference in the number of physical stores of main competitors available in each location.

Our results further suggest that the pandemic, online market maturation over time, and the membership fee elimination might have made grocery markets more efficient. In our sample, an increasing gap in new COVID-19 cases reduces both the share and the magnitude of price dispersion. Overall, our data suggest that price dispersion declines over time when other factors are controlled for. This trend is more pronounced in the share of non-identical prices than in the magnitude of price deviations. This finding corroborates predictions of economics of information regarding more efficient markets and less price dispersion as times go by and online markets mature. In addition to earlier findings, we observed that the change in the magnitude of price dispersion is sensitive to product categories. Food products that are more challenging to store, transport, and hence arbitrage are those for which larger price dispersion is observed.

From recent studies, we know that e-commerce maturation and online competition are expected to change retailing in many ways, including pricing behaviour that might be linked to inflation dynamics. Lower price dispersion and rigidity make prices more sensitive to aggregate shocks and increase the magnitude and speed of their pass-through. Our results confirm that price dispersion in the grocery sector declines, although they suggest a higher starting point than earlier studies. We also show that the anticipated communicated risks of American grocers are indeed associated with price differences observed across locations and over time. Our analysis might help market participants directly or indirectly connected to grocery retailing to anticipate Amazon's reactions in responses to various temporal or spatial shocks and/or to better model those using our price dispersion determinants as instruments in structural modelling.

Besides this gradual decrease in price dispersion over time, there seem to be further steps taken by Amazon to facilitate the spread of uniform prices accross its geographic locations. The price dispersion decline following the Amazon Fresh fee elimination is a case in point. Both the fee elimination and restricting its own prices from following the US inflationary trends (Hillen 2021), applied jointly with more uniform pricing, can all be a part of Amazon's expansion course in times of more price-sensitive consumers. Locked at home and income-restricted, consumers might not only be willing to spend a larger share of their income on online versus offline products but might also become more aware of existing price levels and discrepancies, fuelling the ethical debate on non-uniform pricing and price discrimination. The increasing price uniformity and synchronization at Amazon could further increase convergence and flexibility in other retailers' pricing, fostering market integration and efficiency in and beyond the grocery sector, online and offline alike.

Supplementary material

Supplementary data, including data and replication code, are available at *Q Open* online.

Funding

We are thankful for the funding by the German research foundation (DFG, project-nr. FE 1830/1-1) received by Svetlana Fedoseeva.

End Notes

1 Our prices are for posted promotional prices excluding delivery costs for products that are *simultaneously* available in both locations. As such, products that are stock-out in one or both locations are not included in the analysis. Individual stock-outs might be non-random, occurring more frequently for highly demanded products, or maybe also for niche products with low stocks, or for perishable products with a short shelf life. However, given the large number of products in the sample and because there is no information available on consumer preferences regarding those individual products across locations, we assume that overall, the remaining product pairs in the analysis are random.

Appendix A. Descriptive statistics: raw data and variables

Raw data		Mean	Min	Max	Standard deviation	Data sources (notes)
P. capita personal income:	New York	71.53	67.62	78.39	3.29	Bureau of Economic Analysis (in thousand US dollars)
	California	66.65	61.84	73.06	3.45	
Consumer price index:	New York City	271.19	262.15	284.17	6.16	US Bureau of Labor statistics (CPI food and beverages,
	Los Angeles	266.04	256.38	280.29	7.48	1982–84 = 100, not seasonally adjusted)
Walmart stores:	New York	117.17	116.00	119.00	1.18	Walmart (2021)
	California	320.69	320.00	322.00	0.96	
Costco stores:	New York	19.00	19.00	19.00	0.00	Costco (2021)
	California	128.60	125.00	131.00	1.59	
Target stores:	New York	83.80	79.00	87.00	2.29	Target (2021)
	California	295.34	283.00	307.00	8.39	
COVID new cases:	New York City	1172.56	0.00	6354.00	1559.22	NYC Department of Health and Mental Hygiene
	Los Angeles	1222.74	0.00	3555.00	884.14	LA County COVID-19 Surveillance Dashboard
Variable		Mean	Min	Max	Standard deviation	Variable definition: *Absolute difference in…*
$\Delta Income$		4.85	3.83	5.81	0.42	… per capita personal incomes in both locations
$\Delta Inflation$		5.01	1.56	10.24	2.09	… CPI in both locations
$\Delta Competition$		524.51	513.00	533.00	7.07	… total number of competitors‘ stores in both locations
$\Delta Covid$		251.01	0.00	5341.00	471.43	… new COVID-19 infections in both locations

Appendix B. Determinants of spatial price dispersion at different cut-off values

	Share			Magnitude		
	0.005 per cent	0.05 per cent	0.5 per cent	0.005 per cent	0.05 per cent	0.5 per cent
Δ*Income*	0.027***	0.027***	0.027***	4.713***	4.693***	4.551***
	(0.00)	(0.00)	(0.00)	(0.05)	(0.05)	(0.05)
Δ*Inflation*	−0.003***	−0.003***	−0.003***	−1.184***	−1.169***	−1.054***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Δ*Competition*	0.016***	0.016***	0.016***	1.387***	1.371***	1.233***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Δ*Covid* ($\times$1000)	−0.010***	−0.010***	−0.010***	−1.855***	−1.846***	−1.717***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)	(0.02)
Trend ($\times$1000)	−0.193***	−0.192***	−0.188***	−1.156***	−0.803**	−0.690**
	(0.00)	(0.00)	(0.00)	(0.36)	(0.36)	(0.35)
NoFee	−0.207***	−0.207***	−0.205***	−23.436***	−23.340***	−21.946492***
	(0.00)	(0.00)	(0.00)	(0.09)	(0.09)	(0.08)
Constant	−7.868***	−7.857***	−7.691***	−719.430***	−711.159***	−640.584***
	(0.14)	(0.14)	(0.14)	(6.26)	(6.21)	(5.95)
Adj. R-squared	0.43	0.43	0.43	0.28	0.28	0.28
Number of products	18,216	18,213	18,202	18,216	18,213	18,202
Number of observations	2,194,110	2,193,127	2,183,236	2,194,110	2,193,127	2,183,236

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variables are the share of non-identical prices (*Share*) and the absolute mean price difference in per cent (*Magnitude*) for each price pair. The cut-off level is stated in the respective column head. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at an individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported.

Appendix C. Determinants of spatial price dispersion with individual product groups excluded: the share of non-identical prices

	Meat and fish	Produce	Deli	Dairy	Frozen	Bread and bakery
$\Delta Income$	0.019***	0.026***	0.024***	0.025***	0.040***	0.026***
	0.00	0.00	0.00	0.00	0.00	0.00
$\Delta Inflation$	0.000	−0.002***	−0.004***	−0.006***	−0.005***	−0.003***
	0.00	0.00	0.00	0.00	0.00	0.00
$\Delta Competition$	0.013***	0.016***	0.016***	0.019***	0.020***	0.016***
	0.00	0.00	0.00	0.00	0.00	0.00
$\Delta Covid\ (\times 1000)$	−0.014***	−0.011***	−0.012***	−0.007***	−0.008***	−0.010***
	0.00	0.00	0.00	0.00	0.00	0.00
$Trend\ (\times 1000)$	−0.136***	−0.183***	−0.206***	−0.257***	−0.264***	−0.187***
	0.00	0.00	0.00	0.00	0.00	0.00
$NoFee$	−0.204***	−0.200***	−0.207***	−0.245***	−0.209***	−0.200***
	0.00	0.00	0.00	0.00	0.00	0.00
Constant	−6.234***	−7.529***	−7.532***	−9.290***	−9.708***	−7.813***
	0.14	0.14	0.14	0.15	0.15	0.14
Adj. R-squared	0.44	0.43	0.44	0.47	0.41	0.44
Number of products	17,656	17,537	17,849	16,252	15,968	17,896
Number of observations	2,014,895	2,053,523	2,109,539	1,680,299	1,809,592	2,098,435

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variable is the share of non-identical prices (*Share*). The excluded product group is stated in the respective column head. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at an individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported.

	Pantry	Snacks	Beverages	Baby food	Sweets	Breakfast
Δ*Income*	0.029***	0.024***	0.031***	0.028***	0.027***	0.026***
	0.00	0.00	0.00	0.00	0.00	0.00
Δ*Inflation*	−0.004***	−0.005***	−0.003***	−0.003***	−0.003***	−0.003***
	0.00	0.00	0.00	0.00	0.00	0.00
Δ*Competition*	0.016***	0.018***	0.016***	0.016***	0.016***	0.017***
	0.00	0.00	0.00	0.00	0.00	0.00
Δ*Covid (×1000)*	−0.008***	−0.009***	−0.011***	−0.010***	−0.010***	−0.010***
	0.00	0.00	0.00	0.00	0.00	0.00
Trend (×1000)	−0.174***	−0.222***	−0.168***	−0.190***	−0.191***	−0.196***
	0.00	0.00	0.00	0.00	0.00	0.00
NoFee	−0.198***	−0.196***	−0.208***	−0.207***	−0.207***	−0.206***
	0.00	0.00	0.00	0.00	0.00	0.00
Constant	−7.630***	−8.524***	−7.473***	−7.828***	−7.827***	−7.911***
	0.15	0.15	0.14	0.14	0.14	0.14
Adj. R-squared	0.43	0.44	0.43	0.43	0.43	0.43
Number of products	14,306	16,379	16,396	17,934	17,537	17,711
Number of observations	1,828,244	1,956,879	2,096,842	2,174,308	2,153,434	2,160,330

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variable is the share of non-identical prices (*Share*). The excluded product group is stated in the respective column head. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at an individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported.

Appendix D. Determinants of spatial price dispersion with individual product groups excluded: magnitude of price differences

	Meat and fish	Produce	Deli	Dairy	Frozen	Bread and bakery
$\Delta Income$	4.679***	4.846***	4.754***	4.050***	5.308***	4.819***
	0.06	0.06	0.06	0.06	0.06	0.06
$\Delta Inflation$	–1.129***	–1.143***	–1.227***	–1.189***	–1.203***	–1.227***
	0.01	0.01	0.01	0.01	0.01	0.01
$\Delta Competition$	1.314***	1.322***	1.392***	1.532***	1.569***	1.405***
	0.01	0.01	0.01	0.01	0.01	0.01
$\Delta Covid$ ($\times 1000$)	–1.908***	–1.811***	–1.891***	–2.003***	–1.826***	–1.870***
	0.02	0.02	0.02	0.02	0.02	0.02
$Trend$ ($\times 1000$)	–0.641*	0.020	–1.851***	–7.379***	–4.402***	–0.871**
	0.38	0.37	0.37	0.42	0.41	0.37
$NoFee$	–22.955***	–22.788***	–23.372***	–23.224***	–24.037***	–23.598***
	0.09	0.09	0.09	0.11	0.10	0.09
Constant	–681.584***	–687.009***	–721.461***	–788.615***	–815.469***	–729.402***
	6.47	6.32	6.37	7.24	7.04	6.37
Adj. R-squared	0.28	0.27	0.28	0.29	0.28	0.28
Number of products	17,656	17,537	17,849	16,252	15,968	17,896
Number of observations	2,014,895	2,053,523	2,109,539	1,680,299	1,809,592	2,098,435

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variable is the absolute mean price difference in per cent (*Magnitude*) for each price pair. The excluded product group is stated in the respective column head. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at an individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported.

	Pantry	Snacks	Beverages	Baby food	Sweets	Breakfast
Δ*Income*	4.532***	4.274***	4.875***	4.720***	4.698***	4.652***
	0.06	0.06	0.06	0.05	0.06	0.05
Δ*Inflation*	–1.195***	–1.177***	–1.209***	–1.187***	–1.201***	–1.196***
	0.01	0.01	0.01	0.01	0.01	0.01
Δ*Competition*	1.361***	1.374***	1.324***	1.382***	1.376***	1.389***
	0.01	0.01	0.01	0.01	0.01	0.01
Δ*Covid* ($\times$1000)	–1.730***	–1.641***	–1.923***	–1.873***	–1.870***	–1.870***
	0.02	0.02	0.02	0.02	0.02	0.02
Trend ($\times$1000)	0.276	–0.372	0.605	–0.996***	–0.868**	–1.230***
	0.40	0.38	0.37	0.36	0.37	0.37
NoFee	–23.642***	–23.261***	–23.648***	–23.466***	–23.514***	–23.450***
	0.10	0.09	0.09	0.09	0.09	0.09
Constant	–705.648***	–711.391***	–687.601***	–716.637***	–713.400***	–719.909***
	7.01	6.73	6.41	6.30	6.32	6.32
Adj. R-squared	0.28	0.28	0.28	0.28	0.28	0.28
Number of products	14,306	16,379	16,396	17,934	17,537	17,711
Number of observations	1,828,244	1,956,879	2,096,842	2,174,308	2,153,434	2,160,330

Notes: ***, **, and * refer to statistical significance at $P < 0.01$, 0.05, and 0.10, respectively. Robust standard errors are in parentheses. The dependent variable is the absolute mean price difference in per cent (*Magnitude*) for each price pair. The excluded product group is stated in the respective column head. The economic indicators, COVID-19-related numbers, and competition variables are absolute differences in respective values between New York City (New York) and Los Angeles (California). Fixed effects are at individual product level. Categorical variables for days of the week and months of the year are included in the estimation but not reported.

References

Amazon (2021). *Amazon: Annual Report 2020*. https://s2.q4cdn.com/299287126/files/doc_financials/2021/ar/Amazon-2020-Annual-Report.pdf, accessed 5 Mar 2021.

Anania G. and Nisticò R. (2014). 'Price dispersion and seller heterogeneity in retail food markets', *Food Policy*, 44: 190–201.

Aparicio D., Metzman Z. and Rigobon R. (2021). 'The pricing strategies of online grocery retailers', NBER Working Paper No. w28639, Cambridge, MA: National Bureau of Economic Research, https://www.nber.org/papers/w28639, accessed Apr 2021.

Ater I. and Rigbi O. (2018). 'The effects of mandatory disclosure of supermarket prices', SSRN Scholarly Paper ID 3178561, Rochester, NY: Social Science Research Network.

Baker W., Marn M. and Zawada C. (2001). 'Price smarter on the net', *Harvard Business Review*, 79: 122–7.

Bakos J. Y. (1997). 'Reducing buyer search costs: implications for electronic marketplaces', *Management Science*, 43: 1676–92.

Barslund M. (2007). *CHECKROB: Stata Module to Perform Robustness Check of Alternative Specifications. Statistical Software Components S456837*, Boston College Department of Economics, revised 06 Jun 2018.

Berardi N., Sevestre P. and Thébault J. (2017). 'The determinants of consumer price dispersion: evidence from French supermarkets', in Matyas, L. (ed.) *The Econometrics of Multi-dimensional Panels. Advanced Studies in Theoretical and Applied Econometrics, vol 50*. Cham: Springer, https://doi.org/10.1007/978-3-319-60783-2_15

Billor N., Hadi A. S. and Velleman P. F. (2000). 'BACON: blocked adaptive computationally efficient outlier nominators', *Computational Statistics & Data Analysis*, 34: 279–98.

Biswas D. (2004). 'Economics of information in the web economy: towards a new theory?', *Journal of Business Research*, 57: 724–33.

Böheim R., Hackl F. and Hölzl-Leitner M. (2021). 'The impact of price adjustment costs on price dispersion in e-commerce', *International Journal of Industrial Organization*, 77: 102743.

Brynjolfsson E. and Smith M. D. (2000). 'Frictionless commerce? A comparison of internet and conventional retailers', *Management Science*, 46: 563–85.

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

——— (2018). 'More Amazon effects: online competition and pricing behaviors', NBER Working Paper No. 25138, Cambridge, MA: National Bureau of Economic Research. https://www.nber.org/papers/w25138.

Cavallo A., Neiman B. and Rigobon R. (2014). 'Currency unions, product introductions, and the real exchange rate', *The Quarterly Journal of Economics*, 129: 529–95.

Correia S. (2015). 'Singletons, cluster-robust standard errors and fixed effects: A bad mix', Technical Note, Duke University: 1–7.

Costco (2021). *Costco Wholesale Corporation: Annual Report 2020*. https://investor.costco.com/sec-filings/sec-filing/10-k/0000909832-20-000017

DellaVigna S. and Gentzkow M. (2019). 'Uniform pricing in U.S. retail chains', *The Quarterly Journal of Economics*, 134: 2011–84.

Duch-Brown N., Grzybowski L., Romahn A. and Verboven F. (2021). 'Are online markets more integrated than traditional markets? Evidence from consumer electronics. *Journal of International Economics*, 131: 103476.

Fudenberg D. and Villas-Boas J. M. (2012). 'Price discrimination in the digital economy', *The Oxford Handbook of the Digital Economy*, Oxford: Oxford University Press, 254.

Gorodnichenko Y., Sheremirov V. and Talavera O. (2018). 'Price setting in online markets: does it click?', *Journal of the European Economic Association*, 16: 1764–811.

Grover V., Lim J. and Ayyagari R. (2006). 'The dark side of information and market efficiency in e-markets', *Decision Sciences*, 37: 297–324.

Gunn D. (2019). 'E-commerce and the pricing behavior of traditional retailers', The Digest, NBER, No. 1, https://www.nber.org/digest/e-commerce-and-pricing-behavior-traditional-retailers

Hillen J. (2019). 'Web scraping for food price research', *British Food Journal*, 121: 3350–61.

——— (2021). 'Online food prices during the COVID-19 pandemic', *Agribusiness*, 37: 91–107.

Mookherjee S., Lee J. J. and Sung B. (2021). 'Multichannel presence, boon or curse?: a comparison in price, loyalty, regret, and disappointment', *Journal of Business Research*, 132: 429–40.

Reichheld F. F. and Schefter P. (2000). 'E-loyalty: your secret weapon on the web', *Harvard Business Review*, 78: 105–113. https://hbr.org/2000/07/e-loyalty-your-secret-weapon-on-the-web

Schader P. (2021). 'Tschüssi, servus, bye bye: Amazon macht endgültig schluss mit prime now', Supermarketblog, https://www.supermarktblog.com/2021/05/21/tschuessi-servus-bye-bye-amazon-macht-endgueltig-schluss-mit-prime-now/, accessed 21 May 2021.

Stambor Z. (2018). 'Prices are becoming more uniform and that's to Amazon's advantage', Digital Commerce, https://www.digitalcommerce360.com/2018/10/22/prices-are-becoming-more-uniform-and-thats-to-amazons-advantage/, accessed 22 Oct 2018.

Target (2021). *Target Corporation: Annual Report 2020*. https://sec.report/Document/0000027419-21-000010/

Troy M. (2021). 'The pandemic-fueled record growth in 2020', *Progressive Grocer*, 100: 30–42. https://progressivegrocer.com/pandemic-fueled-record-growth-2020-pg-100

Walmart (2021). *Walmart Inc.: Annual Report 2020*. https://sec.report/Document/0000104169-21-000033/

Wang W. and Li F. (2020). 'What determines online transaction price dispersion? Evidence from the largest online platform in China', *Electronic Commerce Research and Applications*, 42, 100968.