Who Benefits from Entry?*

Louise Guillouët
Columbia University

April 1, 2022

Abstract: The arrival of high-end grocery stores in gentrifying neighborhoods is a harbinger of gentrification. However, economic theory generally predicts that entry of firms is good for consumer welfare. This paper combines barcode-level retail data with a newly collected dataset on the opening dates of Whole Foods, a high-end grocery chain in the United States, in new neighborhoods, to estimate the effect of entry. I show that Whole Foods' entry causes prices to rise, an effect that is entirely driven by a four percent increase in prices paid by households in the lower half of the income distribution. Building on differentiated competition models, I show that this can happen because incumbent stores catering to high-income households are closer to Whole Foods' assortment and therefore behave pro-competitively when Whole Foods arrives, while incumbent stores catering to low-income households are quite differentiated and are able to rise their prices. I then show evidence supporting this mechanism and show that alternative mechanisms do not seem to be supported by the data. My finding is robust to robustness checks and a falsification test using announcement dates instead of entry dates.

JEL: D12; D63; L13; L81

Keywords: gentrification, imperfect competition, consumer goods

*Columbia University Department of Economics (email: louise.guillouet@columbia.edu). I thank Blaine Helleloid and Helena Bauer-Mitterlahner for excellent research assistance. I wish to thank John Asker, Gaia Dossi, Amit Khandelwal, Tobias Salz, Wolfram Schlenker, David Weinstein and seminar participants at Columbia University, the University of Warwick, Sciences Po, the French Association for Environmental and Resource Economics and the European Central Bank for helpful feedback. Researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

Gentrification has been accelerating in the United States, perhaps due to rising economic inequality: between 2015 and 2000, 20% of neighborhoods had experienced gentrification, compared to 8% in the 1990s.[1] It is also a concern in Europe and, perhaps with different expressions, in the Global South (see Delgadillo (2016) about Mexico City for example, and Atkinson and Bridge (2004) for a global overview). One important marker of this social phenomenon is the arrival of new businesses, which often serve the new residents as reflected in the product assortment, hours and prices (Zukin et al. (2009), Glaeser et al. (2018)). This has been decried by groups advocating for the low-income households living in these areas, who are already being displaced because of rising housing costs. Traditional economic intuition, however, suggests that the entry of new businesses is good for consumer welfare.

While the literature has studied the distributional effects of housing upgrading (Guerrieri et al. (2013)), little attention has been given to the distributional effects of business upgrading. This matters because low-income households spend a much larger share of their income on retail expenditure than other households (see Appendix Figure 8), so they may be more affected by changes in the competition structure of retail firms in their neighborhood than high-income households would be. Understanding how the business side of gentrification affects people across the income distribution is therefore of prime importance in order to inform policies that seek to address gentrification.

This paper documents one example where the entry of new firms into neighborhoods may benefit high-income households, yet hurt low-income households, further deepening economic inequality. Using a causal identification strategy, I find that the entry of a high-end firm in a neighborhood does not lead to a pro-competitive price effect overall, and causes the price of products purchased by lower-income households to increase. I find that low-income households pay four percent more for the same products in the same stores, three years after entry, compared to before entry.

In particular, I study the entry of a large, high-end grocery store chain in the United States: Whole Foods. Started in the 1980s in Texas, this chain focuses on self-defined "quality" foods, excluding a list of artificial ingredients and favoring "natural", environmentally friendly choices.[2] The products it carries are typically more expensive than similar products carried by other grocery stores, that may not meet the quality standards decided by Whole Foods. The opening of a new store is often viewed as a harbinger of gentrification.[3] I focus on a period of rapid expansion Whole Foods double its number of stores, going from 110 stores in 109 zipcodes in 2004 to 408 stores in 395 zipcodes in 2017.[4] I study the price responses of incumbent stores

[1] According to the Merriam-Webster dictionary, gentrification is "a process in which a poor area (as of a city) experiences an influx of middle-class or wealthy people who renovate and rebuild homes and businesses and which often results in an increase in property values and the displacement of earlier, usually poorer residents". See this 2015 Governing report.

[2] See the chain's own website for details: https://www.wholefoodsmarket.com/quality-standards/food-ingredient-standards

[3] http://nymag.com/intelligencer/2017/07/harlem-gentrification-whole-foods-vanishing-new-york.html

[4] In 2017, Whole Foods was purchased by Amazon, a sign of its success and the start of a new era for the chain. See the 2017 article https://www.nytimes.com/2017/06/16/business/dealbook/amazon-whole-foods.html a 2022 example of changes https://www.nytimes.com/2022/02/28/technology/whole-foods-amazon-automation.html

with an event-study across zipcodes.

The central identification challenge is that entry is endogenous since Whole Foods does not open stores randomly. I therefore exploit the rapid expansion of Whole Foods across the study period to argue that such fast growth must have partly been achieved through the opening of stores at quasi-random times, if not at quasi-random locations. To define entry, I construct a new dataset based on the list of Whole Foods stores ever in operation in the United States, each of which I link to their opening dates thanks to an extensive research based on newspaper data. I estimate an event-study specification that compares purchases at the barcode-store level made by households in zipcodes where a Whole Foods opened during the study period, before and after the store actually opened.

I find that when Whole Foods opens a new store in a zipcode, prices paid by households in the bottom half of the income distribution increase by four per cent on average three years after entry, after controlling for barcode by store and month fixed effects. By contrast, prices paid by households in the top half of the income distribution do not appear to increase. Overall, I find that the entry of Whole Foods in a new zipcode causes prices to increase by two percent after entry, an effect concentrated on the low-income households. This result is robust to changing the store definition and changing the control variables. I also run a falsification test, using the earliest date at which local newspapers reported the opening of Whole Foods as a false treatment date. There when newspapers announce Whole Foods' intention to open a store, there is no effect on prices.

I rationalize this surprising finding with a model of competition with differentiated products. Building on Perloff et al. (1995)'s paper, I show that in a a world of horizontal differentiation, the entry of an additional firm may raise the price of the product proposed by the incumbent firms when the price sensitivity effect: the fact that consumers who stay with the incumbent reveal themselves to prefer its product, dominates the market share effect; which would encourage the incumbent to lower prices. Allowing for asymmetry, I show that the effect of entry can depend on the initial location of the incumbent. Applying it to our context, since Whole Foods presents itself as "America's healthiest grocery store", consumers who prefer healthy products will gain welfare when switching to the new entrant, and may benefit from lower prices at stores that are relatively close (the market share effect dominates) while consumers who prefer other types of products will stick to incumbent stores that are quite different from Whole Foods, and end up paying more (the price-sensitivity effect dominates) and losing welfare.

This relatively rare phenomenon in economics is supported by product differentiation. The intuition is that Whole Foods is not so different from grocery stores catering to high-income households, therefore the effect on these households in mostly pro-competitive. By contrast, there is more difference between Whole foods and the stores initially catering to low-income households. Some of them might decide to leave anyway for Whole Foods, but those who stay reveal themselves to strongly prefer the incumbemt, which is then able to raise its prices. The anonymity of the data does not allow me to directly compare Whole Foods to other stores, but I document this mechanism by showing that there is a substantial difference in prices between the chains favored by low-income households and the chains favored by high-income households.

One could still think that there are other mechanisms explaining the main result. I address the hypothesis of displacement by showing that the entry of Whole Foods causes the number of grocery stores in a zipcode to increase by a precise 0.5 three years after entry, the number of other types of stores selling groceries in a zipcode does not change. I address the hypothesis of quality upgrading by showing that when I remove the barcode fixed effect, which effectively controls for quality, I can no longer detect an effect on prices, confirming that the main result driven by a change of prices at the barcode-store level, not at the product-store level, which could have implied an increase in quality.

In studying the effect of Whole Foods' entry into a new market, I contribute to a large literature on entry. Some papers have shown that in the context of quality differentiation, entry of new firms does not necessarily lead to better outcomes for consumers: Perloff et al. (1995) show that entry leads to higher prices but also higher consumer welfare in a context of horizontal differentiation, while Ershov (2018) shows that the entry of extremely popular firms ("superstar") in a market leads to higher (excessive) entry and lower prices, but also lower quality in a context of horizontal and vertical differentiation and search costs. Chen and Riordan (2008) explain that the market share effect (decrease in quantity sold) creates an incentive for firms to decrease their price after entry, but it may be compensated by the price sensitivity effect. The latter emerges from the fact that when there are more firms in the market, each offering a different quality level, consumers who shift to the new entrant reveal a preference for their quality, and are therefore less willing to switch back to another product for a slight price increase. The authors identify the conditions under which the price under duopoly is higher than the price under monopoly in a symmetric two-product, two-firm market, namely that the joint distribution of buyer values for the two products are negatively correlated. Deck and Gu (2012) provide an experimental test of this condition.

This price-increasing mechanism has been tested empirically in several contexts: Ward et al. (2002) show that the entry of private-label goods can cause the price of branded goods to increase. They suggest three main explanations. The first, quality-upgrading, is often put forward by brand manufacturers. It is hard to verify using regular data. The second relies on search models to explain that when there are more products, the cost of searching increases and so consumers might end up buying a more expensive product on average even if lower priced products are available. The third is a story of differentiation which most closely aligns with the context I study.

In this paper, I take this idea to the data by studying empirically an oligopolistic market with entry of a clearly differentiated firm. In this, I also contribute to showing that households across the income distribution have quite different preferences when it comes to food, an observation which helps to understand Allcott et al. (2019)'s result that the entry of supermarkets does not increase the nutritional quality of low-income households' diets.

This paper relates to a small group of papers that study the consequences of the entry of foreign supermarkets in developing countries, where they usually disrupt the grocery sector from the top (Javorcik and Li (2013), Iacovone et al. (2015)). In particular, Atkin et al. (2018) study the impact of the opening of Wal-Mart in thousands of Mexican municipalities on consumer welfare. In Mexico, a developing country whose grocery store was until recently dominated by small-scale "mom-and-pop" stores, Wal-Mart entered as a top-quality competitor, offering cheaper prices for the same products, but higher quality and higher prices on average. They show that the entry of Wal-Mart in a municipality led incumbent stores to decrease prices of existing goods, which means that all consumers benefited from entry at the top, although high-income consumers benefited more since they were also able to upgrade quality by switching to Wal-Mart. The authors do not document the evolution of quality in the incumbent stores. In this paper, I am able to verify that the price-increasing effect exists conditional on quality (measured in the barcodes).

This paper also contributes to a literature that studies the impact of new retail stores in the United States, in particular Walmart's mass merchant format's impact on the rest of the market (Basker (2005), Hausman and Leibtag (2007), Jia (2008), Holmes (2011).)

I structure the remainder of the paper as follows: Section 2 describes the data used and provides context on the study. Section 3 presents the empirical strategy, the main results and several robustness checks. Section 4 sketches the theoretical framework that can account for the result. Section 5 describes supporting evidence for the mechanism proposed as well as some evidence against alternative mechanisms. Section 6 concludes.

2 Data and Setting

2.1 Nielsen Homescan Data

The analysis is based on Nielsen's panel dataset from 2004 (the year the dataset begins) to 2017 (the year if the purchase of Whole Foods by Amazon). About 40,000 to 60,000 households participate each year in the panel. They are incentivized to report any consumer packaged good purchased from any outlet.[5] Participants use an in-home scanner to scan the barcodes, allowing me to observe very detailed information about each product.

The strength of the homescan data compared to the alternative dataset, Retail Measurement Services (RMS), for my purposes, is that consumers report purchases from any outlet, not just the ones Nielsen has an RMS agreement with. However, stores are anonymized, I am therefore agnostic about whether the purchase I observe are at a Whole Foods or not, but assume the vast majority are incumbent stores.

I also observe some demographic information about the households, including the zipcode where they live and an income bracket. Table 1 provides information on several demographics about the households eventually in the study compared to the rest of the households in the Nielsen Homescan dataset. Variables are averaged across years 2004-2017. It is quite telling that, although the general perception is that Whole Foods increasingly opens in poorer neighborhoods, betting on gentrification, it still enters neighborhoods that are much richer than in the rest of the dataset on average, which makes sense given its positioning on the product market.

[5]A subset of households also report bulk or scale items, but I do not use this data. Allcott et al. (2019) show that packaged goods represent at least 60% of households' caloric intake.

	Control			Treated			Difference	
	mean	sd	N	mean	sd	N	diff	p
Income	61506.11	39618.77	764603	74000.22	45014.36	30222	12,494.100	0.00
Years of education	14.40	1.97	764603	15.25	1.91	30222	0.852	0.00
Age	56.18	13.06	764603	56.35	13.66	30222	0.171	0.03
Household Size	2.38	1.30	764603	2.20	1.25	30222	-0.187	0.00
White	0.83	0.38	765199	0.80	0.40	30222	-0.032	0.00
Black	0.10	0.30	765199	0.09	0.29	30222	-0.004	0.04
Married	0.63	0.48	765199	0.54	0.50	30222	-0.090	0.00
Employed	0.61	0.43	765199	0.66	0.43	30222	0.041	0.00
Weekly work hours	22.73	16.19	764603	24.25	16.21	30222	1.524	0.00

Table 1: Difference between treated and control households

Notes: Table reports summary statistics of the households who are residents of the clean, treated zipcodes eventually integrated in the main analysis (see Section 3 for details), and of households who are residents in the rest of the dataset.

2.2 Whole Foods Stores Locations and Opening Dates

I obtained the list of establishments selling "organic foods and services" (SIC 5499-35) on ReferenceUSA. In 2016 there were 2,145 such businesses in the US, 477 of which were Whole Foods Markets' stores. I cross-checked this list with the current list of stores given on the corporation's website, where I also collected the exact locations of each store.

I then collected the opening dates and announcement dates of Whole Foods Market new stores using ProQuest' news database, and occasionally Factiva. The process was as follows: for each store present in a given city, I entered in ProQuest search engine the following words "Whole Foods + opening city". I usually found the report in a local newspapers of the event hosted at the store on the opening day, which gave me the date. I then looked for the earliest piece of news mentioning the upcoming entry of Whole Foods in that city among the first 20 news piece coming up in the search "Whole Foods + opening city + year" where year is the year of opening. This news piece was either local newspapers again or comments on Whole Foods' quarterly earning reports filed by the SEC. I always stored the link to the news sources, available upon request. When I went past the first 20 results of ProQuest and Factiva for the first step, I then collected the opening date by looking at the date of the earliest entry for the store on Yelp, a consumer review website's website. For these stores I did not complete the second step, as usually these are stores that opened before the start of my data.

2.3 Retail Establishments Counts

I use the U.S. Census' Zipcode Business Patterns to measure the number of stores in the relevant categories: grocery stores (NAICS 445110), convenience stores (NAICS 445120), specialty food markets (4452), drug stores (446110) and supercenters / club stores (4529). Table 2 shows the number of establishments in each category present in zipcodes that end up in the study compared to all the other zipcodes in the United States, averaging across years 2004-2017. Unsurprisingly given that the residents of these zipcodes are richer than residents of control zipcodes, treated zipcodes have significantly more retail stores across all relevant categories.

	Control			Treated			Difference	
	mean	sd	N	mean	sd	N	diff	p
Grocery stores	1.53	3.92	567770	8.10	9.67	4270	6.570	0.00
Convenience stores	0.65	1.70	567770	3.10	3.50	4270	2.453	0.00
Specialty Food stores	0.94	3.14	567770	7.16	9.65	4270	6.228	0.00
Supercenters/club stores	0.10	0.39	567770	0.48	0.87	4270	0.381	0.00
Drug stores	1.01	2.26	567770	5.18	4.00	4270	4.171	0.00

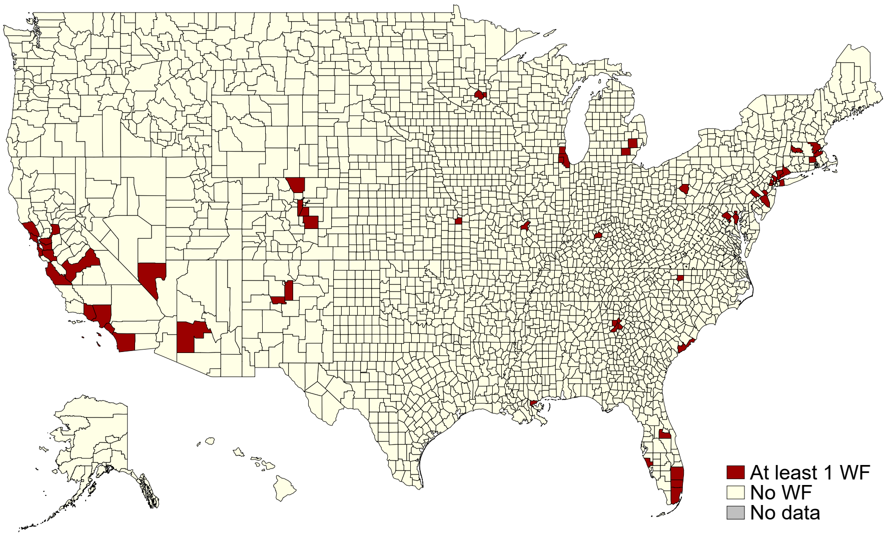
Table 2: Difference between treated and control zipcodes

Notes: Table reports summary statistics of the clean, treated zipcodes eventually integrated in the main analysis (see Section 3 for details), and of all the other zipcodes in the United States.

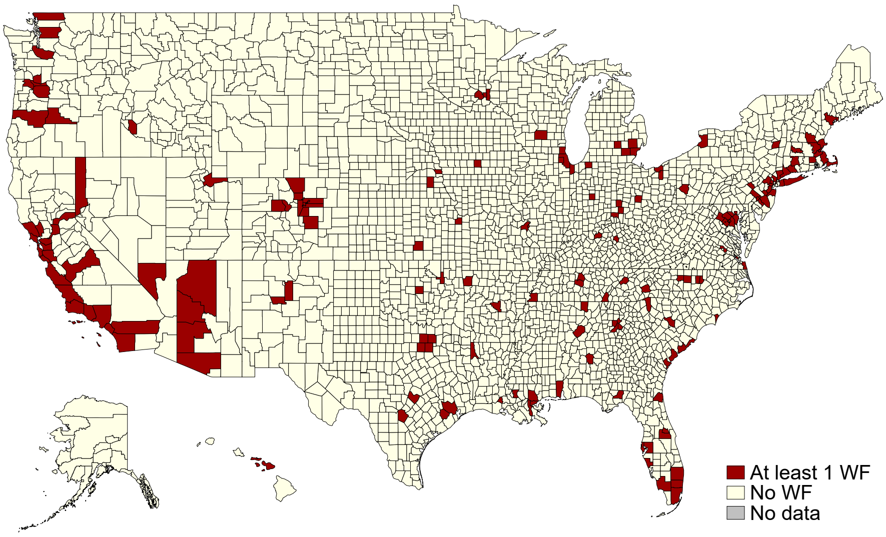
Taken together with the difference in terms of the characteristics of households, this further justifies the use of an event-study instead of a difference-in-differences, natural or synthetic.

2.4 The Expansion of Whole Foods

Whole Foods is an interesting chain to study for its positioning in terms of health and environmental quality. It also expanded very fast between 2004 and 2017, allowing us to implement the empirical strategy described below. To give a qualitative sense for this expansion, Figure 1 shows the counties where Whole Foods was present in 2004, at the beginning of the period of study, and in 2017. My analysis is at the zipcode level for reasons I explain below, but for visualization purposes I show the expansion of Whole Foods from 110 stores in 64 counties in 2004 to 408 stores in 178 counties in 2017. Part of this expansion took place through acquisitions of existing stores operating under different chains, but I do not include them in my analysis.



(a) WF stores in 2004



(b) WF stores in 2017

Figure 1: Geographical Expansion of WF stores, 2004-2017

Notes: Counties in dark red indicate the presence of a Whole Foods stores by the end of 2004 (top figure, 110 stores) and by the end of 2017 (bottom figure, 308 stores). The data come from InfoUSA and Whole Foods' website.

3 Effect of Entry

3.1 Empirical Strategy

To estimate the effect of the entry of Whole Foods on prices faced by consumers in that market, I combine information on new Whole Foods locations and opening dates with data on the prices of products purchased by the household panelists of the Nielsen homescan dataset between 2004 and 2017. I focus on the first time Whole Foods enters a given market (defined below), following Atkin et al. (2018)'s approach.

Studying the rapid expansion (252 more stores over 14 years) of a chain seems to offer the ideal setting for a staggered difference-in-difference study. However, Whole Foods does not randomly choose which areas it goes into. As shown above, it goes into areas with very different characteristics in a static sense, so there is no reason to expect that the areas would also evolve in the same way whether Whole Foods opens or not.

I focus instead on the markets where Whole Foods eventually opened a store over the period and perform an event study. The obvious identification concern with the event study is that store openings coincided with some kind of pre trends. A common intuition is that Whole Foods opened stores in locations with increasing prices, because they are experiencing income growth for example. This is the classic gentrification argument. This would lead to an upward-biased estimate of the treatment effect of the entry of Whole Foods on other stores' prices. Alternatively, Whole Foods might target markets where there is increasing competition as more and more stores are opening in these areas, so that prices in other store are decreasing prior to their entry. This would lead to a downward-biased estimate. Last, it could be the case that Whole Foods has expanded as fast as possible in areas where it wanted to go, and did not time its entry. In this scenario, there would not be a substantial bias, as neither the locations or the timing would be very correlated with pre-existing time trends.

I have found qualitative support for the latter hypothesis in the annual reports produced by Whole Foods for its investors. First, the company explains that their high sales growth (approximately 24% compounded annual growth rate between 1990 and 2014) is driven in large part by their opening of new stores, validating the hypothesis that there is a strong drive in the company to expand and to do so rapidly.[6] The reports also confirm the extreme care taken to choose the areas where the new stores will be located to maximize future profits. Second, part of this care involves finding the best real estate opportunities, which can take a long time or be surprisingly swift. When the company discloses the risks of its activities to investors, it mentions many factors (complexity of development, weather, unions, regulation, etc.) that may delay the opening of the store, even once they have taken possession of the space.[7] Taken together, these last two observations suggest that if the *location* of new stores is definitely not random, the *timing* may be considered quasi-random.

This qualitative evidence does not have to be taken at face value: I can verify the absence of pre-trends quantitatively, in the event-study. I estimate the following regression:

$$\ln p_{bgsmt} = \sum_{j=-4}^{12} \beta_j \mathbb{I}(\text{Quarters Since Entry}_{mt} = j) + \delta_{gsbm} + \eta_t + \epsilon_{gsbmt}$$

where $\ln p_{bgsmt}$is the log price of a barcode-product b in product group g, individual store s, in market m and month t. I() is an indicator function, and Quarters Since Entry$_{mt}$ counts the quarters since the first foreign entry for each market m at time t (with negative values counting quarters before entry, positive values counting quarters after entry and zero being the quarter the first Whole Foods store enters a market). I chose the event window to be within one year

[6]From the Annual Report for 2014: target growth for fiscal year is 9%, with comparable store sales growth "in the low to middle single digits" and the remainder from new stores

[7]Annual Report, 2008: The "tender period," which we define as the length of time between a store's tender date and opening date, varies depending on several factors, some of which are outside of our control. These factors include the size of the store and complexity of site development, the impact of weather and unforeseen environmental issues, and issues surrounding construction labor unions and local government authorities, among other things. Furthermore, acquired leases, ground leases and owned properties generally have longer tender periods than standard operating leases because we take possession of these locations earlier in the construction process.

before the event and three years before the event to match Atkin et al. (2018)'s specification. I include barcode by store fixed effects and month fixed effects. The coefficients on the "negative" quarters since the event measure the pre-trends, while the coefficients on the "positive" quarters since the event measure the effect of entry.

I only observe precise store identifiers for stores belonging to chain that participate in Nielsen's flagship program, the Retail Measurement Survey (RMS). These are typically large chains. The advantage of the Homescan dataset, for this paper, is that it covers all stores consumers go to, however for the subset of stores not in the RMS I only see the code for the chain, not the specific store. I construct alternative store identifier definitions by creating unique identifiers for each retailer code and resident zipcode combination. Because these are not the zipcodes of the stores, I potentially fail to separate two stores from the same chain located in the same zipcode. But I also probably create too many identifiers when residents from different zipcodes go to the same store. Since it is rare that a chain owns several outlets in the same zipcode, this control is probably conservative.

I define markets by combining administrative data with knowledge about consumers' transportation habits. I observe consumers' zipcode of residence. According to the 2017 National Household Travel Survey[8], the mean shopping trip distance was 7.1 miles, while the median and 75th percentile of shopping travel distance were 3 and 7 miles, respectively. Around each Whole Foods store, I therefore create a circle of radius 10 miles that corresponds to a conservative definition of the area of influence of that store. I intersect it with the zipcode areas map. I define an area as valid for my study if it hosted exactly zero store before 2004 and acquired exactly one such store between 2004 and 2017. I exclude areas that are covered by the 10-mile radius circle but do not host the store itself. Last, to estimate the event study on a balanced sample of areas, I exclude zones where the first Whole Foods opened in the first year of my dataset (2004) or in the last 3 years of my dataset (2015-2017). I am left with 252 zipcodes.

I cluster standard errors at the zipcode level, as it is level where errors are most likely to be serially correlated given that the shock is at the zipcode level.

3.2 Main Results

Figure 2 shows the results of the event study estimation by plotting the coefficients and the 95% confidence intervals obtained on the dummies for the quarters before and after the opening of a new Whole Foods stores. Coefficients are precisely estimated zero before the Whole Foods entry event, start increasing as soon as entry occurs, and level off slightly below 2% higher than pre-entry levels between 8 and 12 quarters after entry. The first important conclusion is that I have enough controls to take care of any potential pre-trend: the coefficients on the quarters before the opening are precisely estimated zeros, as shown in te three Panels. It confirms the identification assumption that within the areas where Whole Foods chose to enter, the timing was quasi exogenous, controlling for month and location. Second, there seems to be a positive effect of Whole Foods' entry on prices overall, although it is not significant at the 5% level (three out of the twelve post-entry dummies coefficients are significant at the 10% level).

[8]https://nhts.ornl.gov/

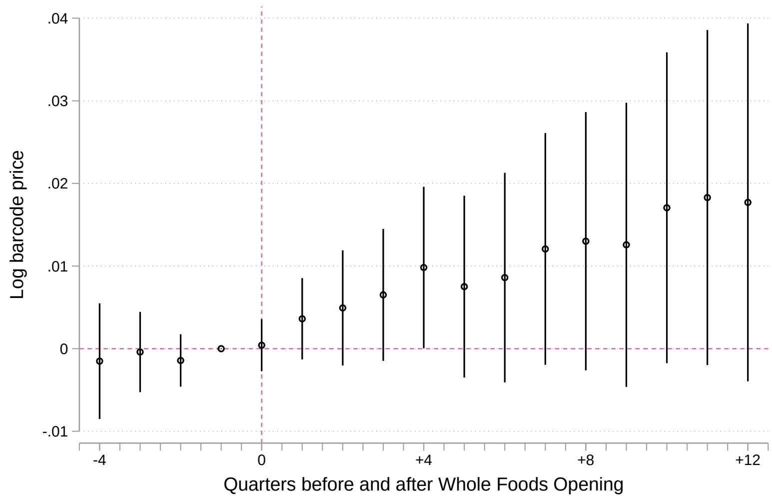
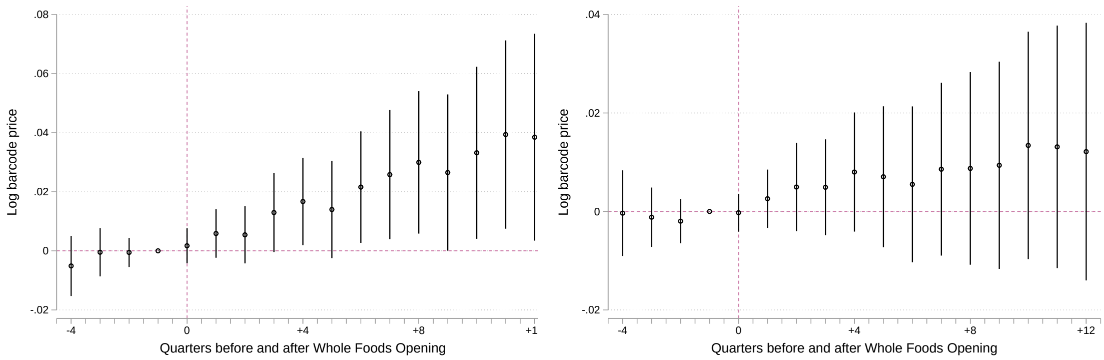


Figure 2: All households

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of log prices on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects and month fixed effects. Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See Table 4 for details.

I conduct two additional event-studies, keeping the specification the same but dividing the sample into two groups: households below and above the median income, $50,000.[9] Figure 3 shows the results for these two separate regressions. I find that there is a positive and statistically significant effect of Whole Foods opening on the prices of products paid by households in the lower half of the income distribution, as shown in Panel (a). Prices increase after the opening and continue to increase for three years, stabilizing just below 4% above prices paid in the year before the opening, while we control for month fixed effects which should take care of typical inflationary trends. This seems to explain most of the effect captured in Figure 2, as Panel (b) shows that prices did not increase much for high-income households: between 0 and 1% and not statistically different from zero.

[9] In 2010 (about half-way through my study period), the median household income in the United States was $49,445 according to the Census. Income in the Nielsen Homescan data are reported using bins, and one of these is up to $49,999, so I use all the bins below and up to this one to define the lower-income group, and put all the other households in the higher-income group.



(a) Below median income

(b) Above median income

Figure 3: Effect of Entry on Prices, Depending on Income Group

Notes: Figure shows the coefficients and 95% confidence intervals obtained from two regressions of log prices on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects and month fixed effects, , run separately for households in the bottom half of the income distribution (top figure) and for households in the top half of the income distribution (bottom figure). Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category for each graph is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See Table 4 for details.

My main result is therefore that the entry of Whole Foods into a new neighborhood, defined by a zipcode, between 2004 and 2017 caused retail prices to increase by three percent on average three years after entry, although this effect is not precisely estimated. It appears most of this effect is driven by a four percent, precisely estimated effect on prices paid by households in the bottom half of the income distribution.

3.3 Robustness checks

In this section, I present several alternative specifications that reinforce my main finding.

3.3.1 Store definition

As explained above, the store definition is not exact. One may worry that the effect I am capturing comes from more granular trends not captured by the store identifiers I define above. I therefore show the results obtained by excluding all the stores that don't have identifiers (have never been in Nielsen's RMS data). I lose about a quarter of observations by doing so. Figure 9 shows that for the entire sample of households, entry has a precisely estimated zero effect on prices purchased at RMS stores. The main result, which is that entry leads prices of goods purchased by households in the bottom half of the income distribution to increase, still holds, as shown in the Panel (a) of Figure 10. All the coefficients estimated on the post-entry quarter dummies are positive, and almost all are satistically different from zero at the 5% level, with a leveling off of the price effect at about 3% three years after entry. Panel (b) of Figure 10 shows that entry may have had a negative effect on prices of purchases made by upper-income households in RMS stores, but this effect is not statistically different from zero. In addition, the fact that there is almost a negative pre-trend here is a reassuring sign that the positive result found above is not due to pre trend. Overall, the smaller effects I find using the subsample of RMS stores is consistent with the evidence in support of large chains in the United States practicing uniform pricing, as argued by DellaVigna and Gentzkow (2019). As most chains set prices nationally for a large array of popular products, a given store is less able to respond, in either direction, to a change in the competitive landscape within its zipcode.

3.3.2 Controls

One may be worried that the effect I find is due to trends that are more granular than the month fixed effects that I use. Following Atkin et al. (2018), I therefore propose to reestimate the main specification, replacing the month fixed effects by zipcode size by month fixed effects, store type by product category by month fixed effects, and region by month fixed effects. Here, zipcode sizes are just coded as the quintile in terms of the population registered in the American Community Survey in 2011. Store type is a dummy dividing stores between grocery stores, and everything else. Regions are the 9 United States "divisions" as defined by the U.S. Census and attributed by Nielsen to households based on their place of residence. The results shown in Figures 11 and 12 are remarkably similar to the one found with the baseline specification, and the estimates are more precise, which is what one would expect from adding relevant controls.

3.3.3 Falsification test

Despite the alternative specifications proposed, one may worry that the effect I am detecting is spurious. I therefore propose a data test. During the collection of the opening dates of Whole Foods stores, I also collected the earliest date on which a media source announced the upcoming opening. This was not feasible for all the "clean entry" event zipcodes, but allows for a relevant comparison. The median time between announcement date and opening date is 22 months: we don't expect incumbent stores to react to the announcement in the media of of the opening of a new store in almost two years. I therefore run again the specification shown above with the announcement date instead of the opening date. I find no detectable effect of the announcement, as shown in Figures 13 and 14. This comforts the result shown above the the entry of Whole Foods causes incumbents to raise prices, especially incumbents serving lower-income households.

4 Price-increasing competition

I this section, I follow Perloff et al. (1995) in presenting a simple model of Bertrand competition with differentiated products to explain my results. Perloff et al. (1995) discuss the entry of one firm compared to a monopoly. I study the entrance of an asymmetric competitor in two-firm market.

Suppose there is a mass L of consumers. They are described by an ideal quality $\hat{t}$, distributed uniformly along the Salop (1979) circle.

Their utility from consuming good i is described by

$$u_i = v - p_i - c|\hat{t} - t_i|$$

where v is the utility from consuming any good in that category, p_i is the price of the good, t_i

is the quality of the good, and c is the marginal utility cost of consuming a good that is distant from one's ideal quality.

The consumer maximizes their utility by choosing the good that maximizes their utility, and buys it if it generates more utility than the outside option, u_0.

4.1 Local Monopolists

To fix ideas, suppose there is only one good. The consumer will choose it if

$$v - p_i - c|\hat{t} - t_i| \geq u_0$$

w.l.o.g., I normalize $u_0 = 0$. The monopolist can sell to consumers who are within the "monopoly region", that is within the distance x_m of its location, that is

$$|\hat{t} - t_i| \leq x_m = \frac{v - p_m}{c}$$

The monopolist faces total demand $2x_m L$. It chooses its price to maximize its profit taking into account its marginal cost m, where $v > m$

$$\max_{p_m} \; 2L\left(\frac{v - p_m}{c}\right)(p_m - m)$$

The F.O.C. gives

$$p_m = \frac{v + m}{2}, \; x_m = \frac{v - m}{2c}$$

If the two firms are sufficiently far apart, they will each charge p_m and attract all the consumers in their local monopoly region. I assume this is the baseline situation before Whole Foods opens.

4.2 A third firm enters

Consumers may be located in three areas: between the entrant and one incumbent, between the entrant and the other incumbent, or between the two incumbents. Suppose the store A is located z_A away from the entrant, while store B is located z_B away from the entrant, with $z_A > z_B$.

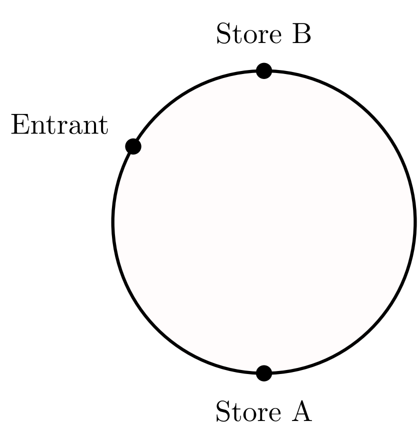


Figure 4: Competitive Landscape after Entry

Since the two incumbents are not truly competing with each other, I will treat each incumbent separately. For the entrant to compete with an incumbent, it must be that the sum of each firms' monopoly region is greater than their distance from each other.

$$\frac{v-p_e}{c} + \frac{v-p_i}{c} > z_i, \ \forall i = A, B$$

The marginal consumer is such that the demand faced by the incumbent from the consumers located between the two firms depends on x_i, defined as:

$$v - p_i - cx_i = v - c(z_i - x_i) - p_e \ \forall i = A, B$$

$$x_i = \frac{z_i}{2} + \frac{p_e - p_i}{2c} \quad \forall i = A, B$$

Therefore the demand is kinked at $\bar{p}_i = v - \frac{cz_i}{2}$. We will focus on the case above the kink.[10]

The incumbent faces demand $L(x_m + x_i)$ (x_m from the consumers between itself and the other incumbent and x_i from the consumers between itself and the entrant). Since it cannot price discriminate, it sets its price to maximize

$$\max_{p_i} L\left(\frac{z}{2} + \frac{p_e - p_i}{2c} + \frac{v - p_i}{c}\right)(p_i - m)$$

$$p_A = \frac{3m + 2v + cz_A + p_e}{6}, \ p_B = \frac{3m + 2v + cz_B + p_e}{6}$$

While the entrant faces demand from consumers located between itself and firm A and between itself and firm B, respectively:

$$\max_{p_e} L\left(\frac{z_A}{2} + \frac{p_A - p_e}{2c} + \frac{z_B}{2} + \frac{p_B - p_e}{2c}\right)(p_e - m)$$

$$p_A = \frac{2m + cz_A + cz_B + p_A + p_B}{5}$$

The solution is

$$p_i = \frac{84m + 48v + 29cz_i + 7cz_{-i}}{132} \ \forall i = A, B \ \ p_e = \frac{7cz_A + 7cz_B + 18m + 4v}{22}$$

[10]

$$D = \begin{cases} L(x_m + x_i) & \text{if } \frac{v-p_e}{2} + \frac{v-p_i}{2} > z_i \\ 2Lx_m & \text{if } \frac{v-p_e}{2} + \frac{v-p_i}{2} < z_i \ \forall i = A, B \end{cases}$$

Above the kink, the incumbent faces the lower part of the demand curve, such that (assuming symmetry) $v - p < z$.

$$\max_{p_i} 2L\frac{v-p_i}{c}(p-m)$$

At the kink, the derivative of profit is negative

$$2L\frac{v - 2p_i + m}{c}|_{p=v-\frac{cz_i}{2}} = \frac{2L}{c}(v - 2v + cz_i + m) < 0 \text{ since } v > m$$

so the optimal price is the kink price $p_i = v - \frac{cz_i}{2} > p_m$. It decreases in z until we are back to the (local) monopoly price.

For any two z_A, z_B such that

$$29cz_B + 7cz_A < 18(v - m) < 29cz_A + 7cz_B$$

we have $p_A > p_m$ and $p_B < p_m$, so store A raises its price compared to before entry and store B decreases its price. The intuition behind this model is that when the entrant locates somewhat close to the store A ($z_A < z$), but not too close ($z_B > z$), some consumers will greatly prefer the entrant and switch. Meanwhile, the incumbent will keep the consumer who greatly prefer its own product. So the average distance the updated pool of customers has to travel to be at firm A is less than it used to be, which means firm A can raise its price.

4.3 Price effect

Figure 5 summarizes the effect entry has on prices in the market, depending on how differentiated the entrant is from the incumbent. When the products are not very differentiated (small z), firms are fierce competitors so entry leads prices to decrease. Intuitively, the market share effect dominates. As the distance between the two products increases, increasingly more customers will strongly prefer one product or the other, allowing firm to leverage the price-sensitivity effect. When the level of differentiation is higher than the monopoly region, $z > x_m$, the price-sensitivity effect dominates and entry causes prices to rise. However, the firms are still competing. When products are so differentiated that the firms are local monopolies, $z > \bar{z}$, the market share effect starts to kick in more, incentivizing firms to lower their price to capture more customers, until the firms are optimally behaving like monopolies, so that the price effect of entry is zero.

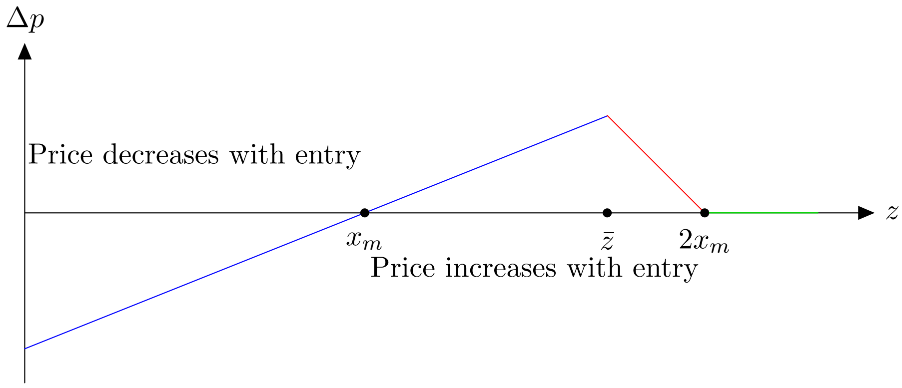


Figure 5: Effect of distance on Price effect of entry

The intuition here is that in general different stores serve low-income and high-income consumers, perhaps letting them compete at high, monopoly level prices because they are very differentiated. When Whole Foods comes in, it competes directly with the stores serving high-income consumers ($z < x_m$), leading to a decrease in price, and less intensely with the stores serving low-income consumers ($z_m < z < 2x_m$), leading to an increase in price in these stores. The mechanism that remains to be tested is the fact that Whole Foods is closer to the taste of rich consumers than poor consumers.

4.4 Welfare changes

The total consumer surplus from the two local monopolists is

$$CS_{2m} = 2\left(v - p_m - cE[x|x < x_m]\right)2x_m = \frac{1}{2}\frac{(v-m)^2}{c}$$

The total consumer surplus from the three firms is:

$$CS_{3f} = \sum_i (v - p_i - cE[x|x < x_i])2x_i + (v - p_i - cE[x|x < x_m(p_i)])x_m(p_i) + (v - p_W - cE[x|x < x_W(p_i)])x_W(p_i)$$

Overall, consumer surplus is almost always higher when there are more firms, even if the price increases for one incumbent firm. However, the welfare changes are not homogeneous. Figure 6 is an illustration of how customers located at different points on the circle (linearized for simplicity) around one incumbent and one entrant are affected by the entry of a firm.

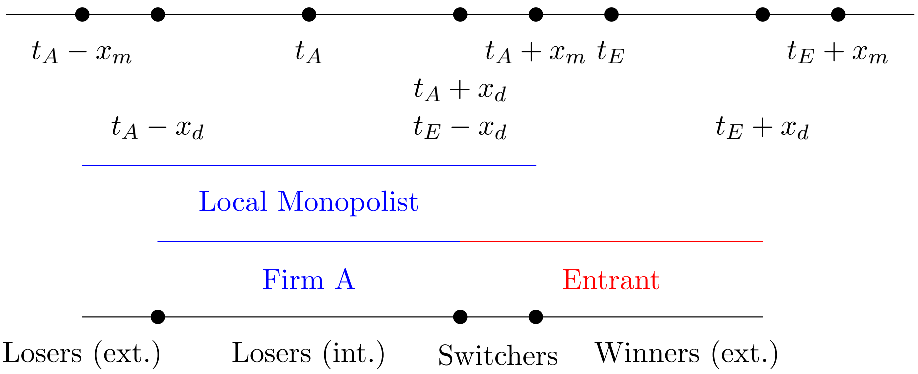


Figure 6: Distribution of welfare consequences when the price increases after entry

Note that first, the customers who were not served previously benefit because they are now happy to purchase a product on the market, where previously they might have to drive elsewhere to get their groceries, or eat out. Because more customers are served, I call this a gain at the extensive margin.

Second, the customers who decide to switch to the new product may gain or lose. This may seem surprising given these consumers' ideal tastes are closer to the entrant than to the incumbent. Indeed, when the duopoly price is lower than the monopoly price these customers gain. [11] However, this is not the case when switchers have to pay a higher price under duopoly than under monopoly.

Third, some customers lose welfare as they are buying the same good in both cases but pay a higher price. This is a loss of welfare at the intensive margin.

Last, in this model some customers lose welfare as they no longer find it in their interest to purchase the good now that its price has increased. This is a loss of welfare at the extensive

[11] Suppose an individual located x away from the incumbent switches. It means that $v - p_e - c|z - x| > v - p_A - cx \Leftrightarrow c|z - x| < cx + p_e - p_A$, they prefer the entrant's good conditional on price. Visually they are located closer to the entrant than to the incumbent. How does his utility after entry compare to his prior utility? $v - p_e - c(z - x) - (v - p_m - cx) = p_m - p_e - cz$ This change is ambiguous, and is negative when the duopoly price is higher than the monopoly price.

margin.

5 Discussion

The main empirical result of the paper is that the entry of a Whole Foods store into a new zipcode causes consumer goods to increase, especially for lower-income households. Above, I propose a theoretical mechanism that has been identified in the literature and that relies on product differentiation and entry of firms. In this section, I discuss three pieces of empirical evidence that support this theory and help to put aside alternative explanations. First, I show that even in the incumbent stores, households sort themselves into different types of grocery stores, corroborating the hypothesis of differentiation correlated with income. Second, I show that the mechanism is unlikely to be a displacement of businesses. Third, I show that the mechanism is unlikely to be quality upgrading.

5.1 Differentiation

The mechanism suggested above relies on product differentiation. A large literature has shown that lower-income and higher-income households have quite different tastes, for example Allcott et al. (2019). Because of the anonymity required by the Nielsen data, I cannot show the difference between Whole Foods' assortment and the other stores' assortment. Here, I propose to look at the (anonymous) chains of stores most favored by lower-income households compared to higher-income households, and vice versa, in my dataset. I start by defining share of expenditure by retailer and group (bottom half vs upper half of the income distribution) over the entire dataset. I then select the retailers that have reached a share above 5%. I define a retailer as catering to lower-income households if this group favors this retailer much more than others.

	Log price			
	(1)	(2)	(3)	(4)
Low-income Chain	-0.055	-0.045		
	(0.014)	(0.007)		
Hich-income Chain			0.951	0.207
			(0.015)	(0.055)
Zipcode by year FEs	Yes	Yes	Yes	Yes
Zipcode by product by month	Yes	Yes	Yes	Yes
Zipcode by barcode by month	No	Yes	No	Yes
Number of zipcodes	233	233	233	233
R2	0.60	0.90	0.63	0.90
N	4499642	2101400	4499642	2101400

Table 3: Differences between chains, by Income Group

Notes: Table reports the coefficients from regressing log price on a dummy identifying a chain that is preferred by low-income households (Columns 1-2) or high-income households (Columns 1-3). Standard errors are clustered at the zipcode level and are reported in parenthesis.

Table 3 shows that the chain favored by low-income households sells products that are 5.5% less expensive on average. This could be due to product assortment or to a competitive approach. Controlling for barcodes, stores from the chain favored by low-income households still charge 4.5% lower prices on average, which means that both mechanisms are at play. This is even more obvious when looking at the chain favored by high-income households: without controlling for barcodes, prices are almost twice higher. However when controlling for barcodes prices are "only" 20% higher, suggesting that a lot of the overall price difference is due to higher quality in that chain. These descriptive results support the idea that the stores that low-income households and high-income households go to are very different, which can explain why the entry of Whole Foods would have a different impact on prices paid by low-income households and prices paid by high-income households.

5.2 Alternative story: displacement

The mechanism suggested above relies on an increase of the number of firms operating in the markets, or zipcodes. An alternative story for the result I obtain is that high-end stores, such as Whole Foods, simply displace lower-end stores which may used to cater to lower-income population by providing cheap goods. It should be noted that the main specification of the empirical analysis controls for barcode by store fixed effects, which means that for low-income households, prices of the same goods bought in the same stores increase after the entry of Whole Foods relative to after. Still, a displacement effect could be happening simultaneously, and is consistent with the narrative related by people advocating for barriers to gentrification. I therefore propose to study the evolution of the number of stores in the zipcodes as an outcome. If I find that the number of stores stays constant, it would mean that there is a replacement story going on. I use the Census's Zipcode Business Patterns to do this. Unfortunately, these data are at the yearly level, so the estimation is much less precise than the ones for prices.

$$N_{mt} = \sum_{j=-1}^{3} \beta_j \mathbb{I}(\text{Years Since Entry}_{mt} = j) + \delta_m + \eta_t + \epsilon_{mt}$$

The result for grocery stores is presented in Figure 7. Being three years after the entry of Whole Foods has a positive and statistically significant effect on the number of grocery stores (excluding convenience stores) operating in the zipcode. The effect is not one, which suggests there might be some exit going on, however qualitatively I view this as evidence of entry. Additionally, Figure 15 shows that there is a precisely estimated zero effect on the number of competing store formats such as mass merchandise stores and convenience stores. Taken together, these results contribute to setting aside the displacement hypothesis.

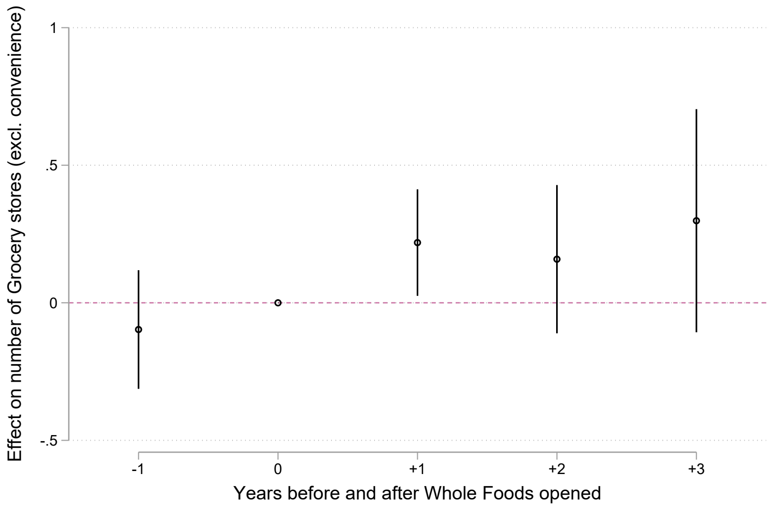


Figure 7: Effect of Entry on Number of Grocery stores in Zipcode

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of number of stores on 5 annual treatment effects, in addition to zipcode fixed effects and month fixed effects. The reference category is the year the Whole Foods store opened. Standard errors are clustered at the zipcode level.

5.3 Alternative story: quality improvement

An alternative explanation for my main result could be that consumers across the income distribution start buying higher-quality goods as Whole Foods opens a new store in their neighborhood. This might be due to several things, including exposure to types of goods that consumers didn't know before, and quality-upgrading competition. It is important to note that the specification I show throughout the main text and the robustness checks is at the barcode by store level, which means that the price increases I estimate are for the same products bought in the same stores. An increase in the average quality of products bought might also be happening simultaneously, which would potentially mitigate the welfare concerns. This seems unlikely, as Allcott et al. (2019) show that the entry of a new store is typically not enough to change the type of good people buy. They structural model allows them to quantify the role of supply-side differences as explaining only about 10% of the nutrition inequality across the income distribution, while the remaining 90% are explained by preferences. Nevertheless, I address this question by estimating a different specification on the same sample as in the core of the analysis. In this alternative specification, instead of controlling for barcode by store fixed effects, I control for product module by store fixed effects. In the Nielsen dataset, there are 1310 product module codes, so they are relatively well defined products, for example, "canned lima beans" or "hair coloring - Women's". However, there are on average 4,138 barcodes per product module (median: 798). There remains a potential variation in price through quality differentiation, whether horizontal (permanent or temporary hair coloring, color etc.) or vertical (whether it is free of harmful additives, such as with the USDA organic seal). In the sample, there exists substantial variation in price within product module code: while product module code fixed effects explain 36% of the variation in price, barcode fixed effects explain 54% of the variation in price. The results of this alternative specification is shown in Figures 16 and 17. There is no detectable effect on the price paid for goods once I remove the barcode by store fixed effects. Given that there is a positive effect on price for continuing products, this suggests that if anything entry has a negative effect on the quality of products chosen, so that when not controlling for barcodes the effects are muted. Although it is hard to quantify, this implies that the effect of entry on consumer welfare might be smaller than under the hypothesis of constant quality.

6 Conclusion

Policymakers worry about gentrification. An oft-cited positive development is the increased entry of businesses. Since these new businesses *a priori* target the new residents, as opposed to the historical and lower-income residents of these neighborhoods, it is important to study the consequences these new businesses have on consumer welfare, particularly that of lower-income households.

This paper shows that between 2004 and 2017, the entry of a Whole Foods store, a harbinger of gentrificiation, into a new zipcode caused prices of non-durable consumer packaged goods to increase on average, an effect mostly driven by a statistically significant increase in prices paid by lower-income households of 4% three years after entry. This result is robust to a range of robustness checks, in particular I show that prices are mostly changing in independent stores or stores belonging to small, local chains as opposed to national chains that may practice uniform pricing and might therefore be less able to react to local conditions.

I propose a mechanism of Bertrand competition with differentiated products, showing that if the products offered by the competitors are neither too similar nor too dissimilar, the entry of a firm may cause prices to rise. This rare effect is generated by the fact that consumers split across the different products, and each remaining firm is faced with a set of consumers with a higher revealed preference for its product, and lower price-sensitivity on average. On the other hand it cannot use a low price to attract consumers who strongly prefer the alternative anymore.

While policymakers may hardly discourage entry, policies that encourage a wide assortment of products in stores, including products that cater to lower-income households, may dampen the negative consequences of the business side of gentrification.

References

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell, "Food deserts and the causes of nutritional inequality," *The Quarterly Journal of Economics*, 2019, *134* (4), 1793–1844.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro, "Retail globalization and household welfare: Evidence from mexico," *Journal of Political Economy*, 2018, *126* (1), 1–73.

Atkinson, Rowland and Gary Bridge, *Gentrification in a global context*, Routledge, 2004.

Basker, Emek, "Job creation or destruction? Labor market effects of Wal-Mart expansion," *Review of Economics and Statistics*, 2005, *87* (1), 174–183.

Chen, Yongmin and Michael H Riordan, "Price-increasing competition," *The RAND Journal of Economics*, 2008, *39* (4), 1042–1058.

Deck, Cary and Jingping Gu, "Price increasing competition? Experimental evidence," *Journal of Economic Behavior & Organization*, 2012, *84* (3), 730–740.

Delgadillo, Víctor, "Selective modernization of Mexico City and its historic center. Gentrification without displacement?," *Urban Geography*, 2016, *37* (8), 1154–1174.

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

Ershov, Daniel, "Competing with superstars in the mobile app market," *Available at SSRN 3265662*, 2018.

Glaeser, Edward L, Hyunjin Kim, and Michael Luca, "Measuring gentrification: Using yelp data to quantify neighborhood change," Technical Report, National Bureau of Economic Research 2018.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst, "Endogenous gentrification and housing price dynamics," *Journal of Public Economics*, 2013, *100*, 45–60.

Hausman, Jerry and Ephraim Leibtag, "Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart," *Journal of Applied Econometrics*, 2007, *22* (7), 1157–1177.

Holmes, Thomas J, "The diffusion of Wal-Mart and economies of density," *Econometrica*, 2011, *79* (1), 253–302.

Iacovone, Leonardo, Beata Javorcik, Wolfgang Keller, and James Tybout, "Supplier responses to Walmart's invasion in Mexico," *Journal of International Economics*, 2015, *95* (1), 1–15.

Javorcik, Beata S and Yue Li, "Do the biggest aisles serve a brighter future? Global retail chains and their implications for Romania," *Journal of International Economics*, 2013, *90* (2), 348–363.

Jia, Panle, "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Perloff, Jeffrey M, Valerie Y Suslow, and Paul J Seguin, "Higher prices from entry: Pricing of brand-name drugs," *U of California, Berkeley, Competition Policy Working Paper No. CPC99-03*, 1995.

Salop, Steven C, "Monopolistic competition with outside goods," *The Bell Journal of Economics*, 1979, pp. 141–156.

Ward, Michael B, Jay P Shimshack, Jeffrey M Perloff, and J Michael Harris, "Effects of the private-label invasion in food industries," *American Journal of Agricultural Economics*, 2002, *84* (4), 961–973.

Zukin, Sharon, Valerie Trujillo, Peter Frase, Danielle Jackson, Tim Recuber, and Abraham Walker, "New retail capital and neighborhood change: Boutiques and gentrification in New York City," *City & Community*, 2009, *8* (1), 47–64.

7 Additional Figures

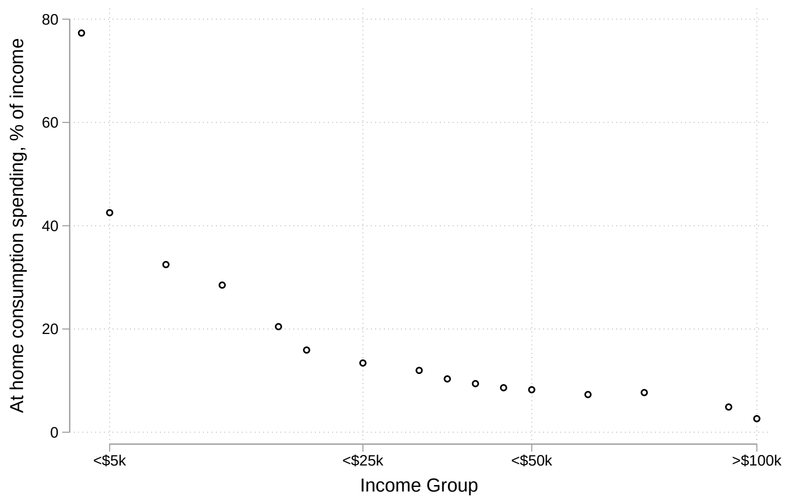


Figure 8: Share of retail expenditure on income, by income group

Notes: Figure shows the share of expenditure on income for each of the income bins defined by Nielsen, where income is defined as the upper end of the bin and expenditure is defined as the sum of all spending reported to through the Homescan tool.

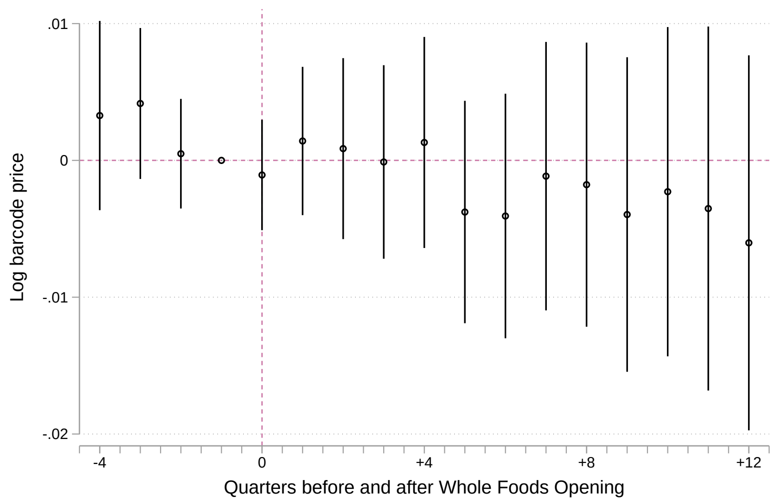
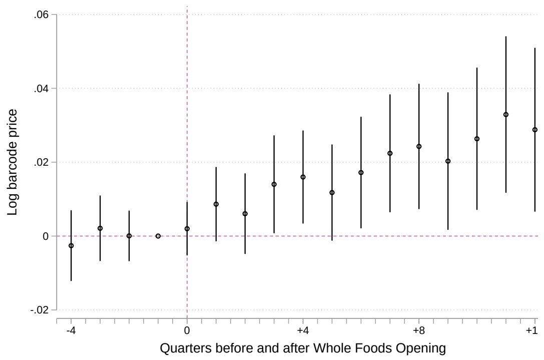
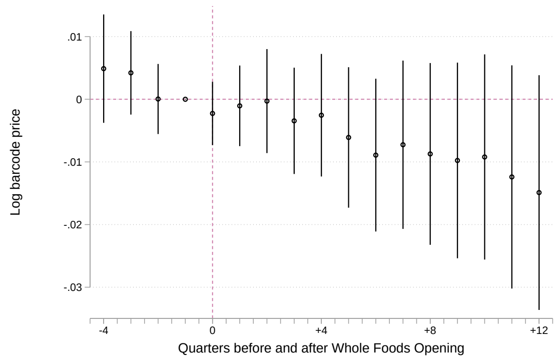


Figure 9: All households, only RMS stores

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of log prices on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects and month fixed effects, separately for households in the bottom half of the income distribution (top figure) and for households in the top half of the income distribution (bottom figure). Here, I only keep stores that participate or have participated in the Nielsen RMS program, and are therefore identified individually. The reference category for each graph is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 4 for details.



(a) Below median income, only RMS stores



(b) Above median income, only RMS stores

Figure 10: Effect of Entry on Prices of goods purchased from RMS stores, Depending on Income Group

Notes: Figure shows the coefficients and 95% confidence intervals obtained from two regressions of log prices on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects and month fixed effects, run separately for households in the bottom half of the income distribution (top figure) and for households in the top half of the income distribution (bottom figure). Here, I only keep stores that participate or have participated in the Nielsen RMS program, and are therefore identified individually. The reference category for each graph is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 4 for details.

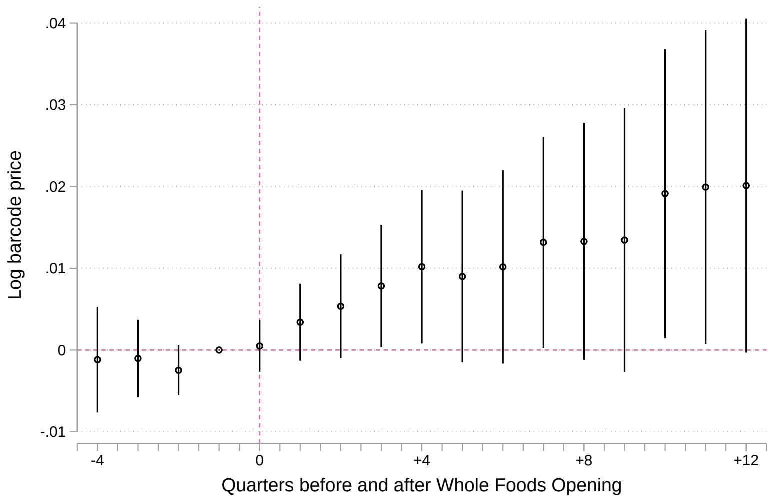
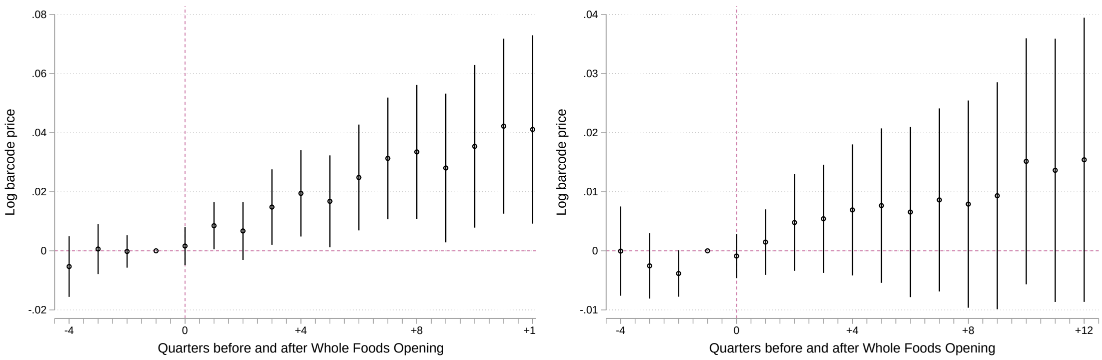


Figure 11: All households, Controls

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of log prices on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects, state-by-month fixed effects, zipcode size-by-month fixed effects and store type-by-product group-by-month fixed effects. Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 5 for details.



(a) Below median income (b) Above median income

Figure 12: Effect of Entry on Prices, Depending on Income Group

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of log prices on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects, state-by-month fixed effects, zipcode size-by-month fixed effects and store type-by-product group-by-month fixed effects. Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 5 for details.

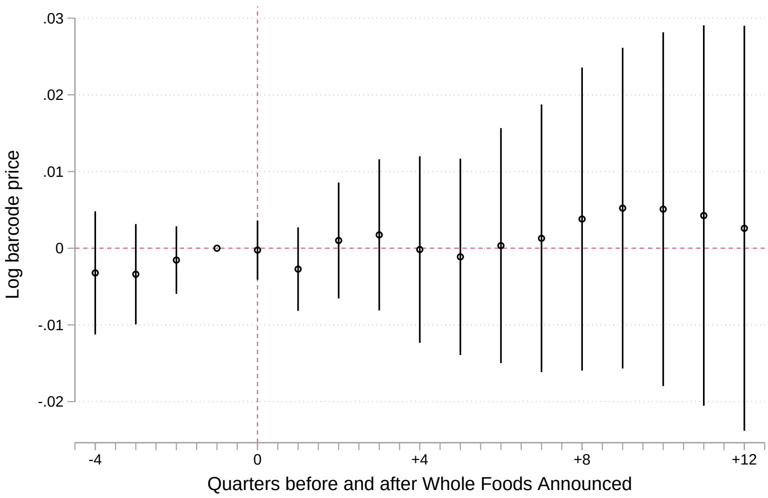
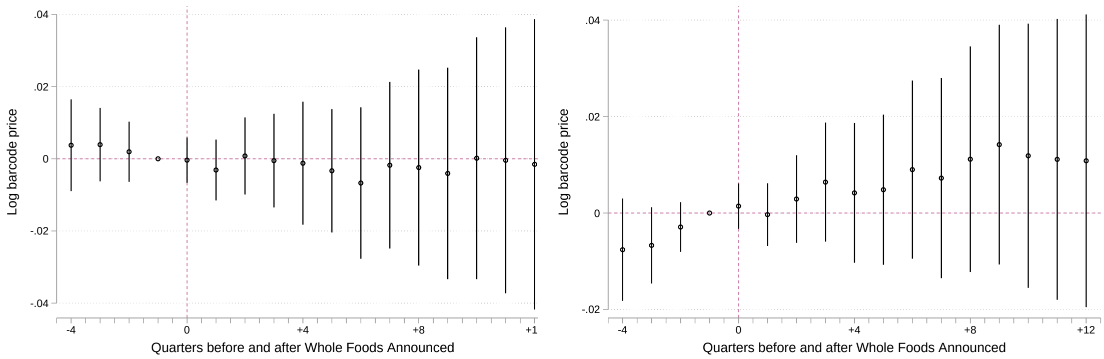


Figure 13: Effect of Announcement on Prices of goods purchased

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of log prices on 16 quarterly “false” treatment effects which are actually announcement effects, in addition to barcode-by-store fixed effects and month fixed effects. Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 6 for details.

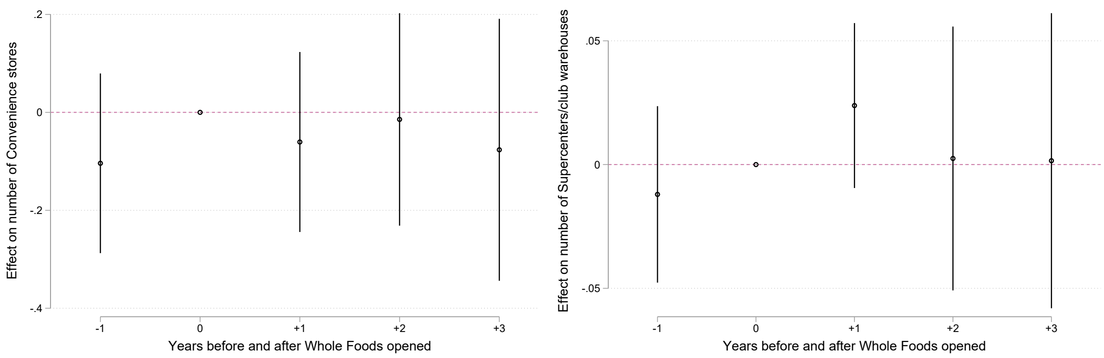


(a) Below median income

(b) Above median income

Figure 14: Effect of Announcement on Prices, Depending on Income Group

Notes: Figure shows the coefficients and 95% confidence intervals obtained from two regressions of log prices on 16 quarterly "false" treatment effects which are actually announcement effects, in addition to barcode-by-store fixed effects and month fixed effects, run separately for households at the bottom of the income distribution (top figure) and at the top of the income distribution (bottom figure). Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 6 for details.



(a) Convenience Stores

(b) Club stores, warehouses and supercenters

Figure 15: Effect of Entry on number of establishments

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of number of stores on 5 annual treatment effects, in addition to zipcode fixed effects and month fixed effects. The top figure shows the results for the number of convenience stores (NAICS 445120) and the bottom figure shows the result for club stores, warehouses and supercenters (NAICS 452910). The reference category is the year the Whole Foods store opened. Standard errors are clustered at the zipcode level.

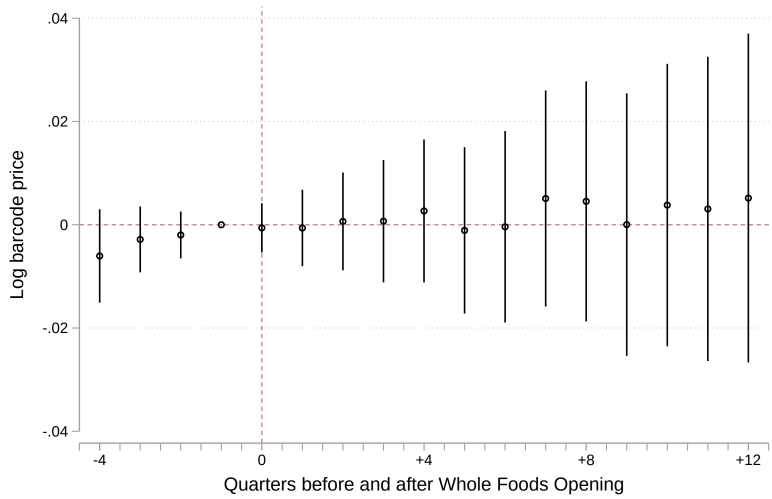
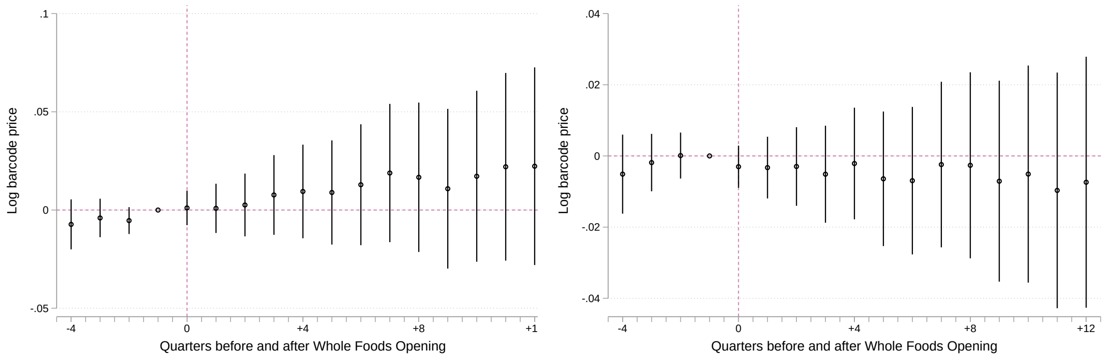


Figure 16: Effect of Opening on Prices of goods purchased

Notes: Figure shows the coefficients and 95% confidence intervals obtained from a regression of log prices on 16 quarterly treatment effects, in addition to product module-by-store fixed effects and month fixed effects. Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 7 for details.



(a) Below median income

(b) Above median income

Figure 17: Effect of Opening on Prices of goods purchased depending on Income Group

Notes: Figure shows the coefficients and 95% confidence intervals obtained from two regressions of log prices on 16 quarterly treatment effects, in addition to product module-by-store fixed effects and month fixed effects, ran separtely for low-income households (top figure) and high-income households (bottom figure). Here, stores are defined by the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the store code attributed by Nielsen for those stores that participate in the RMS program. The reference category is the price one quarter before the Whole Foods store opening. Standard errors are clustered at the zipcode level. See the paper and Table 7 for details.

8 Additional Tables

	Log price					
Households	All	$\leq \$50k$	$> \$50k$	All	$\leq \$50k$	$> \$50k$
	(1)	(2)	(3)	(4)	(5)	(6)
Four Quarters Before	-0.002	-0.005	-0.000	0.003	-0.003	0.005
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
Three Quarters Before	-0.000	-0.000	-0.001	0.004	0.002	0.004
	(0.002)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
Two Quarters Before	-0.001	-0.001	-0.002	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
One Quarter Before	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Quarter of Entry	0.000	0.002	-0.000	-0.001	0.002	-0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
One Quarter After	0.004	0.006	0.003	0.001	0.009	-0.001
	(0.002)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
Two Quarters After	0.005	0.005	0.005	0.001	0.006	-0.000
	(0.004)	(0.005)	(0.005)	(0.003)	(0.006)	(0.004)
Three Quarters After	0.007	0.013	0.005	-0.000	0.014	-0.003
	(0.004)	(0.007)	(0.005)	(0.004)	(0.007)	(0.004)
Four Quarters After	0.010	0.017	0.008	0.001	0.016	-0.003
	(0.005)	(0.007)	(0.006)	(0.004)	(0.006)	(0.005)
Five Quarters After	0.008	0.014	0.007	-0.004	0.012	-0.006
	(0.006)	(0.008)	(0.007)	(0.004)	(0.007)	(0.006)
Six Quarters After	0.009	0.022	0.006	-0.004	0.018	-0.009
	(0.006)	(0.010)	(0.008)	(0.005)	(0.008)	(0.006)
Seven Quarters After	0.012	0.026	0.009	-0.001	0.023	-0.007
	(0.007)	(0.011)	(0.009)	(0.005)	(0.008)	(0.007)
Eight Quarters After	0.013	0.030	0.009	-0.002	0.025	-0.009
	(0.008)	(0.012)	(0.010)	(0.005)	(0.009)	(0.007)
Nine Quarters After	0.013	0.027	0.009	-0.004	0.022	-0.010
	(0.009)	(0.013)	(0.011)	(0.006)	(0.010)	(0.008)
Ten Quarters After	0.017	0.033	0.013	-0.002	0.028	-0.009
	(0.010)	(0.015)	(0.012)	(0.006)	(0.010)	(0.008)
Eleven Quarters After	0.018	0.039	0.013	-0.003	0.035	-0.012
	(0.010)	(0.016)	(0.012)	(0.007)	(0.011)	(0.009)
Twelve Quarters After	0.018	0.038	0.012	-0.005	0.032	-0.015
	(0.011)	(0.018)	(0.013)	(0.007)	(0.012)	(0.010)
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Retailer-zip-barcode FEs	Yes	Yes	Yes	No	No	No
Store-barcode FE	No	No	No	Yes	Yes	Yes
N	3105372	959659	2055517	2322396	698893	1550254
R2	0.92	0.93	0.92	0.90	0.91	0.90
Store-barcode cells	673997	215494	455809	507622	158702	346414
N zipcodes	233	211	225	233	211	225

Table 4: Effect of Entry on prices (main results)

Notes: Table reports regressions of log price on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects and month fixed effects. Columns 1-3 define the stores as the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the Nielsen store code when available (see Figures 2 and 3). Columns 4-6 show the regressions results only keeping stores that have a Nielsen store code (see Figures 9 and 10). Columns 1 and 4 show the regression results using all households in the panel. Columns 2 and 5 show the regression results using only households in the bottom half of the income distribution (defined as less than $50,000). Columns 3 and 6 show the regression results using only households in the top half of the income distribution. Standard errors are clustered at the zipcode level and are reported in parenthesis.

	Log price					
Households	All	$\leq \$50k$	$> \$50k$	All	$\leq \$50k$	$> \$50k$
	(1)	(2)	(3)	(4)	(5)	(6)
Four Quarters Before	-0.001	-0.005	-0.000	0.004	-0.003	0.006
	(0.003)	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)
Three Quarters Before	-0.001	0.001	-0.003	0.005	0.005	0.004
	(0.002)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
Two Quarters Before	-0.002	-0.000	-0.004	-0.000	0.001	-0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
One Quarter Before	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Quarter of Entry	0.000	0.002	-0.001	-0.002	0.000	-0.004
	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
One Quarter After	0.003	0.009	0.001	0.001	0.010	-0.003
	(0.002)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
Two Quarters After	0.005	0.007	0.005	0.002	0.008	-0.000
	(0.003)	(0.005)	(0.004)	(0.003)	(0.006)	(0.004)
Three Quarters After	0.008	0.015	0.005	0.002	0.017	-0.003
	(0.004)	(0.006)	(0.005)	(0.004)	(0.007)	(0.004)
Four Quarters After	0.010	0.019	0.007	0.002	0.020	-0.005
	(0.005)	(0.007)	(0.006)	(0.004)	(0.007)	(0.005)
Five Quarters After	0.009	0.017	0.008	-0.002	0.015	-0.007
	(0.005)	(0.008)	(0.007)	(0.004)	(0.007)	(0.006)
Six Quarters After	0.010	0.025	0.007	-0.002	0.020	-0.009
	(0.006)	(0.009)	(0.007)	(0.005)	(0.009)	(0.006)
Seven Quarters After	0.013	0.031	0.009	-0.000	0.028	-0.009
	(0.007)	(0.010)	(0.008)	(0.005)	(0.009)	(0.007)
Eight Quarters After	0.013	0.033	0.008	-0.001	0.030	-0.011
	(0.007)	(0.011)	(0.009)	(0.006)	(0.010)	(0.008)
Nine Quarters After	0.013	0.028	0.009	-0.002	0.025	-0.010
	(0.008)	(0.013)	(0.010)	(0.006)	(0.010)	(0.008)
Ten Quarters After	0.019	0.035	0.015	0.002	0.032	-0.007
	(0.009)	(0.014)	(0.011)	(0.007)	(0.011)	(0.009)
Eleven Quarters After	0.020	0.042	0.014	0.000	0.040	-0.011
	(0.010)	(0.015)	(0.011)	(0.007)	(0.012)	(0.010)
Twelve Quarters After	0.020	0.041	0.015	-0.001	0.037	-0.011
	(0.010)	(0.016)	(0.012)	(0.008)	(0.013)	(0.010)
Store type-group-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode size-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Retailer-zipcode-barcode FEs	Yes	Yes	Yes	No	No	No
Store-barcode FE	No	No	No	Yes	Yes	Yes
N	3102732	955704	2052462	2318840	693983	1546151
R2	0.92	0.93	0.92	0.90	0.92	0.90
Store-barcode cells	673121	214267	454794	506476	157170	345065
N zipcodes	233	211	225	233	211	225

Table 5: Effect of Entry on prices (additional controls)

Notes: Table reports regressions of log price on 16 quarterly treatment effects, in addition to barcode-by-store fixed effects, state-by-month fixed effects, zipcode size-by-month fixed effects and store type-by-product group-by-month fixed effects. Columns 1-3 define the stores as the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the Nielsen store code when available (see Figures 11 and 12). Columns 4-6 show the regressions results only keeping stores that have a Nielsen store code. Columns 1 and 4 show the regression results using all households in the panel. Columns 2 and 5 show the regression results using only households in the bottom half of the income distribution (defined as less than \$50,000). Columns 3 and 6 show the regression results using only households in the top half of the income distribution. Standard errors are clustered at the zipcode level and are reported in parenthesis.

Households	All (1)	< \$50k (2)	≥ \$50k (3)	All (4)	< \$50k (5)	≥ \$50k (6)
	Log price					
Four Quarters Before	-0.003	0.004	-0.008	-0.005	-0.005	-0.006
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)
Three Quarters Before	-0.003	0.004	-0.007	-0.006	-0.002	-0.007
	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)
Two Quarters Before	-0.002	0.002	-0.003	-0.002	-0.003	-0.001
	(0.002)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
One Quarter Before	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Quarter of Announcement	-0.000	-0.000	0.001	-0.001	0.003	-0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
One Quarter After	-0.003	-0.003	-0.000	-0.002	0.003	-0.002
	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)
Two Quarters After	0.001	0.001	0.003	0.004	0.007	0.002
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Three Quarters After	0.002	-0.001	0.006	0.004	0.007	0.004
	(0.005)	(0.007)	(0.006)	(0.004)	(0.007)	(0.006)
Four Quarters After	-0.000	-0.001	0.004	0.001	0.008	-0.001
	(0.006)	(0.009)	(0.007)	(0.005)	(0.009)	(0.007)
Five Quarters After	-0.001	-0.003	0.005	-0.001	0.006	-0.002
	(0.006)	(0.009)	(0.008)	(0.005)	(0.009)	(0.008)
Six Quarters After	0.000	-0.007	0.009	0.002	0.005	0.003
	(0.008)	(0.011)	(0.009)	(0.006)	(0.011)	(0.009)
Seven Quarters After	0.001	-0.002	0.007	-0.001	0.007	-0.003
	(0.009)	(0.012)	(0.011)	(0.007)	(0.012)	(0.010)
Eight Quarters After	0.004	-0.002	0.011	0.003	0.009	0.002
	(0.010)	(0.014)	(0.012)	(0.008)	(0.014)	(0.011)
Nine Quarters After	0.005	-0.004	0.014	0.005	0.011	0.004
	(0.011)	(0.015)	(0.013)	(0.008)	(0.015)	(0.012)
Ten Quarters After	0.005	0.000	0.012	0.005	0.017	0.002
	(0.012)	(0.017)	(0.014)	(0.008)	(0.018)	(0.012)
Eleven Quarters After	0.004	-0.000	0.011	0.002	0.018	-0.003
	(0.013)	(0.019)	(0.015)	(0.009)	(0.020)	(0.014)
Twelve Quarters After	0.003	-0.002	0.011	0.002	0.021	-0.004
	(0.013)	(0.020)	(0.015)	(0.010)	(0.021)	(0.014)
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Retailer-zipcode-barcode FEs	Yes	Yes	Yes	No	No	No
Store-barcode FE	No	No	No	Yes	Yes	Yes
N	2506189	792956	1635723	1910174	591344	1254819
R2	0.91	0.92	0.92	0.89	0.90	0.90
Store-barcode cells	553594	180869	369904	430400	137839	289855
N zipcodes	180	162	177	180	162	177

Table 6: Effect of announcement on entry (falsification test)

Notes: Table reports regressions of log price on 16 quarterly announcement effects (see the paper for details), in addition to barcode-by-store fixed effects and month fixed effects. Columns 1-3 define the stores as the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the Nielsen store code when available. (see Figures 13 and 14) Columns 4-6 show the regressions results only keeping stores that have a Nielsen store code. Columns 1 and 4 show the regression results using all households in the panel. Columns 2 and 5 show the regression results using only households in the bottom half of the income distribution (defined as less than \$50,000). Columns 3 and 6 show the regression results using only households in the top half of the income distribution. Standard errors are clustered at the zipcode level and are reported in parenthesis.

Households	All	< $50k$	≥ $50k$	All	< $50k$	≥ $50k$
	Log price					
	(1)	(2)	(3)	(4)	(5)	(6)
Four Quarters Before	-0.006	-0.007	-0.005	-0.003	-0.008	0.000
	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.007)
Three Quarters Before	-0.003	-0.004	-0.002	0.001	-0.003	0.003
	(0.003)	(0.005)	(0.004)	(0.004)	(0.006)	(0.005)
Two Quarters Before	-0.002	-0.005	0.000	-0.001	-0.007	0.003
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
One Quarter Before	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Quarter of Entry	-0.001	0.001	-0.003	-0.001	0.005	-0.005
	(0.002)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
One Quarter After	-0.001	0.001	-0.003	-0.002	0.006	-0.006
	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)	(0.005)
Two Quarters After	0.001	0.003	-0.003	-0.003	0.005	-0.009
	(0.005)	(0.008)	(0.006)	(0.005)	(0.008)	(0.006)
Three Quarters After	0.001	0.008	-0.005	-0.004	0.011	-0.013
	(0.006)	(0.010)	(0.007)	(0.006)	(0.009)	(0.007)
Four Quarters After	0.003	0.009	-0.002	-0.004	0.012	-0.012
	(0.007)	(0.012)	(0.008)	(0.007)	(0.010)	(0.008)
Five Quarters After	-0.001	0.009	-0.006	-0.010	0.012	-0.020
	(0.008)	(0.013)	(0.010)	(0.007)	(0.010)	(0.009)
Six Quarters After	-0.000	0.013	-0.007	-0.011	0.017	-0.022
	(0.009)	(0.016)	(0.011)	(0.008)	(0.012)	(0.010)
Seven Quarters After	0.005	0.019	-0.002	-0.005	0.026	-0.019
	(0.011)	(0.018)	(0.012)	(0.009)	(0.013)	(0.012)
Eight Quarters After	0.005	0.017	-0.003	-0.007	0.019	-0.020
	(0.012)	(0.019)	(0.013)	(0.010)	(0.014)	(0.012)
Nine Quarters After	0.000	0.011	-0.007	-0.013	0.013	-0.025
	(0.013)	(0.021)	(0.014)	(0.011)	(0.015)	(0.013)
Ten Quarters After	0.004	0.017	-0.005	-0.012	0.019	-0.027
	(0.014)	(0.022)	(0.015)	(0.012)	(0.016)	(0.015)
Eleven Quarters After	0.003	0.022	-0.010	-0.015	0.027	-0.036
	(0.015)	(0.024)	(0.017)	(0.013)	(0.017)	(0.016)
Twelve Quarters After	0.005	0.022	-0.007	-0.013	0.025	-0.034
	(0.016)	(0.026)	(0.018)	(0.014)	(0.018)	(0.017)
Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Retailer-zipcode-product FEs	Yes	Yes	Yes	No	No	No
Store-product FE	No	No	No	Yes	Yes	Yes
N	4453052	1394848	2978743	3342293	1026238	2258884
R2	0.74	0.77	0.75	0.64	0.70	0.66
Store-barcode cells	2021985	650781	1379277	1527698	486117	1055184
N zipcodes	233	211	225	233	211	225

Table 7: Effect of entry on prices (testing for the quality mechanism)

Notes: Table reports regressions of log price on 16 quarterly treatment effects, in addition to product-by-store fixed effects (see paper for details) and month fixed effects. Columns 1-3 define the stores as the unique combination of the retailer code attributed by Nielsen, the zipcode of the resident and the Nielsen store code when available (see Figures 16 and 17). Columns 4-6 show the regressions results only keeping stores that have a Nielsen store code. Columns 1 and 4 show the regression results using all households in the panel. Columns 2 and 5 show the regression results using only households in the bottom half of the income distribution (defined as less than $50,000). Columns 3 and 6 show the regression results using only households in the top half of the income distribution. Standard errors are clustered at the zipcode level and are reported in parenthesis.