Multimarket Contact and Collusion in Online Retail

Poppius, Hampus

2020

Document Version:
Other version

Link to publication

Citation for published version (APA):
Poppius, H. (2020). *Multimarket Contact and Collusion in Online Retail*. (Working papers; No. 2020:5).

Total number of authors:
1

Working Paper 2020:5

Department of Economics
School of Economics and Management

Multimarket Contact and Collusion in Online Retail

Hampus Poppius

April 2020

Multimarket Contact and Collusion in Online Retail*

Hampus Poppius†

April 8, 2020

Abstract

When firms meet in multiple markets, they can leverage punishment ability in one market to sustain collusion in another. This is the first paper to test this theory for multiproduct retailers that sell consumer goods online. With data on the universe of consumer goods sold online in Sweden, I estimate that multimarket contact increases prices. To more closely investigate what drives the effect, I employ a machine-learning method to estimate effect heterogeneity. The main finding is that multimarket contact increases prices to a higher extent if there are fewer firms participating in the contact markets, which is one of the theoretical predictions. Previous studies focus on geographical markets, where firms provide a good or service in different locations. I instead define markets as different product markets, where each market is defined by the type of good. This is the first paper to study multimarket contact and collusion with this type of market definition. The effect is stronger than in previously studied settings.

JEL Classification: D22, D43, L41, L81.
Keywords: Tacit collusion, pricing, e-commerce, causal machine learning.

1 Introduction

Firms may refrain from fierce price competition if they expect reciprocity by competitors in additional markets. This notion was early expressed by Edwards (1955): "there is an incentive to live and let live, to cultivate a cooperative spirit, and to recognize priorities of interest in the hope of reciprocal recognition". The idea was later formalized by Bernheim & Whinston (1990), who show that multimarket contact (MMC)

*I am grateful for detailed comments by Hakan J. Holm and Petra Thiemann. This paper has also benefited from comments by Alessandro Martinello, Alexandros Rigos, Emil Palikot, Bruno Jullien, Patrick Rey, Daniel Ershov, Andrew Newman, Nicolaj Mühlbach, and seminar participants at the Competition Policy Workshop at Toulouse School of Economics, Lund University, and the Institute for Industrial Economics (IFN). I am also grateful for financial support from the Swedish Retail and Wholesale Council and the Tom Hedelius Foundation.

†Department of Economics, Lund University, Tycho Brahes väg 1, 220 07 Lund, Sweden. Institute of Retail Economics, Stockholm. E-mail: hampus.poppius@nek.lu.se

may indeed facilitate collusion. While empirical studies find support for the theoretical prediction, the focus has been exclusively on MMC across geographically separated markets. With digitization and low transportation costs, the geographical boundaries of markets become obsolete for many goods.

This paper studies MMC and collusion in online retailing, where geographical market definitions are obsolete. Instead, I define markets as distinct product markets, where each product market is a product category. Thus, two firms have MMC if they share multiple product categories. I use data on the universe of online retailers that sell consumer goods in Sweden to estimate the effect of MMC on price. To identify the effect of MMC on price, I compare time variation in MMC across firms. As Bernheim & Whinston (1990) predict, prices increase with MMC.

Bernheim & Whinston (1990) show that when firms meet in multiple markets, they may leverage slack enforcement power in one market to sustain collusion in another. Thus, for MMC to facilitate collusion, collusion must be stable in at least one market in isolation. Intuitively, when two firms meet in multiple markets and enjoy profits from collusive prices in one market, the fear of retaliation in that profitable market deters price cutting in the other markets.

In a given market, A, I measure the number of contact markets each retailer has with the competitors in that market. To identify the effect of MMC on prices, I estimate how the price changes over time with MMC. I compute my measure of MMC so that it is insensitive to entry and exit decisions in market A. Therefore, in the terminology of Schmitt (2018), the variation in MMC is driven solely by "out-of-market consolidation".

I find that an increase of one additional contact, on average across competitors, leads to a 0.2 percent increase in price. This estimate implies that an increase from the 25th to the 75th percentile of MMC corresponds to a price increase of 7.5 percent. For most estimates in the literature, the equivalent change in MMC corresponds to price increase of between zero and five percent. I decompose the change in MMC to show that the effect is driven by lost contact markets. Furthermore, the heterogeneity analysis shows that the effect is much larger when there are relatively few retailers in the contact markets. While the theory predicts the main result, the heterogeneity results strongly support that it is the mechanism in Bernheim & Whinston (1990) that drives the estimated effect.

There are two main contributions of this paper. First, I show that MMC increases prices in a setting very different from those earlier studied. Previous studies have found MMC to increase prices of airline tickets (Evans & Kessides, 1994; Singal, 1996; Miller, 2010; Bilotkach, 2011; Ciliberto & Williams, 2014), banking services (Pilloff, 1999; De Bonis & Ferrando, 2000; Coccorese & Pellecchia, 2009; Aysan, Disli & Schoors, 2013; Molnar, Violi & Zhou, 2013; Kasman & Kasman, 2016), cellular telephony (Busse, 2000; Parker & Röller, 1997), movie tickets (Feinberg, 2014), hotel nights (Fernandez & Marin, 1998), cement (Jans & Rosenbaum, 1997), hospitals (Schmitt, 2018), and radio broadcasting (Waldfogel & Wulf, 2006). In all these settings (except cement) firms provide one type of service in multiple geographical locations. MMC is thus equivalent to serving multiple locations concomitantly. Quite differently, the firms in this paper are online retailers that sell many different types of goods in one geographical market (Sweden). These firms instead meet in multiple product markets, and may thus punish a price cut on one product by cutting prices on other products. To the best of my knowledge, the literature has been completely silent about MMC in regard to this type of market definition.

One of the conditions for MMC to facilitate collusion is that markets must be different. Markets are more likely different when defined by different products, than merely the same good or service sold in different locations. In addition, cheap real-time monitoring of competitors' prices already facilitates collusion between online retailers. In fact, 53 percent of online retailers in the EU track competitors' online prices, and 67 percent of those that do use automatic software for it (European Commission, 2016). However, it is unclear from a theoretical stance how the relationship between MMC and prices for online retailers should compare to that observed in previous studies. Furthermore, geographical market definitions are obsolete for many products in the digital economy. Not only do online retailers reach distant locations due to low shipping costs, but many products are even provided digitally at zero shipping costs. Examples beyond online retail include cloud computing, software, digital applications, and e-books. Competing providers of such products have few geographical distinctions but may provide multiple services or goods concomitantly.

The second contribution is that I leverage the richness of the data to explore heterogeneity in the effect of MMC on prices in a principled way. A classical approach to effect heterogeneity is to test for the effect in different subgroups or interact the treatment variable with covariates. This approach, however, raises issues about multiple testing and spurious findings. To avoid these issues, I employ an algorithm that both searches for heterogeneity and tests it by recursively splitting the sample. The algorithm, provided by Athey, Tibshirani & Wager (2019), is called Generalized Random Forest (GRF) and it is a modification of the popular machine-learning algorithm Random Forest by Breiman (2001). The GRF is a completely data-driven method for finding effect heterogeneity when there are many potential subgroups to consider. As the Random Forest, the GRF performs very well in detecting and modeling non-linear relations and complex interactions. A key feature that Athey & Imbens (2016) call *honesty*, means that with one sample the algorithm searches for relevant subgroups and with a different sample it estimates the effects. The GRF predicts a conditional average partial effect of MMC on price for any combination of the heterogeneity variables (within the range of observed values) with pointwise confidence intervals.

The heterogeneity results strongly support one of the key mechanisms in Bernheim & Whinston (1990). Because collusion must be stable in isolation in the contact market for existence of any slack punishment ability, MMC may facilitate collusion if the contact market is more concentrated. This paper's results concur. Specifically, when there are relatively few retailers in the contact market, the effect of MMC on price is higher. Guided by the GRF predictions, I further bolster this result by standard conventional analysis.

The GRF predicts substantial heterogeneity in the effect of MMC on price. One third of the observations have predicted positive and significant effects that are more than twice the average effect. Most of the other observations have insignificant effects close to zero and only ten percent have significantly negative effects. The effect is larger if the retailer offers many other products within the product category and has a low number of contacts at the beginning of the time period.

The remainder of the paper is organized as follows. In Section 2 I describe the data, the market definition and the key measures. I describe the empirical strategy in Section 3. Section 4 presents the main results. Section 5 describes the estimation of heterogeneous effects and presents the corresponding results. Section 6 concludes.

2 Data, Market Definition, and Measures

I collect the data from Prisjakt.nu, a price comparison website. It is the leading price comparison website in Sweden and its aim is to list everything that is available for purchase online in Sweden. From this website, I collect the price set by each retailer on each product. In addition, I also collect the available information about the retailers and the products, e.g. the retailer's rating and the product category of the product. I collect the data at two points in time, the first set in September 2018 and the second set in September 2019.

The types of products span a wide range of product categories. The overarching level of categorization is the groups of categories, hereafter referred to as *groups*. The groups are *Audio & Video, Beauty & Health, Camera & Photo, Computers & Accessories, Fashion & Accessories, Games & Consoles, Home & Garden, Kids & Toys, Phones & GPS*, and *Sports & Outdoors*. Within each group, there are many specific product categories. The product categories define the finest level of categorization, and some examples are *Eyeliners, Motherboards, Shower Doors, Blenders*, and *Smartwatches*. The website sorts each product into exactly one product category to make it easy for users to find different versions of the type of product they want. A product category thus contains different variations of the same good, differentiated by e.g. brand, size, and color. In 2018, there are 951 product categories that contain on average 944 different products.

I use this fine level of product categorization as my market definition. Thus, I define a market as a product category. Because the products are sold online and shipped all over Sweden, there are no geographically distinct markets. The definition of markets as product markets defined by the product categories is both feasible and appealing. It is feasible because I adopt the existing distinction between groups of products that the website already made. To define each market by myself would not be feasible because of the vast number of products and it would require in-depth expertise about every product. This market definition is appealing because the website actually have the same goal with their categorization as I have with market definition. They want to keep substitute products in the same category, so that users can easily find different variations of the same good. If products that are not substitutes would be placed in the same category, it would make users' search slower. Similarly, if substitute products are placed in different categories, search is slower. Each product

Furthermore, this market definition is transparent. By adopting the product categories as market definition, I restrain my own discretion. Moreover, anyone can visit Prisjakt.nu[1] if interested in what products belong to which market.[2].

[1] https://pricespy.co.uk/ is the British version.

[2] A complete list of products and corresponding product categories at the time of data collection can be provided upon request

To give an overview of the data, Table 1 presents descriptive statistics of the sample. Columns 1 and 2 presents information for 2018 and 2019 separately, and column 3 presents information about the matched sample. To define the matched sample, let us first define a *listing*. Consider that each retailer sells a range of products. Furthermore, most products are sold by several different retailers. Denote each combination of a retailer and a product as a listing. Thus, a listing is the finest unit of observation in the data, and it is the level of analysis. Because I exploit variation over time while holding listing fixed effects constant, I limit the analysis to the listings that occur in both time periods. The matched sample thus consists of the listings that are repeated.

Panel A of Table 1 provides a description of the data structure. The top row shows that there are almost 900,000 products listed in 2018, but roughly 750,000 in 2019. There are about 570,000 products for which there is at least one repeated listing. The total number of retailers decreases slightly from 3,303 to 3,239. There are 2,511 retailers with repeated listings. Most of the retailers with no repeated listings either start or close operations between the two times of data collection.

Panel A also shows how the number of product categories vary over time. The number of categories increases from 951 to 979, but the number of categories in the matched sample is only 919. It is lower because some categories either dissolve completely, are divided into multiple new categories, or are merged with other categories.

The bottom row of the Panel A reveals that about half of the 2018 listings exist also in 2019. That is largely because many products are not offered by any retailer in 2019, and many retailers close down. The rest of the table describes some differences between the repeated listings and the non-repeated ones.

Panel B shows two statistics at the product level. First, the number of retailers slightly decrease over time. However, there are more retailers selling each product for which there exists repeated listings. That does not mean that there are on average 4.64 repeated listings per product. It means that for a product for which there exists at least one repeated listing, there are on average 4.64 retailers at a given point in time.

The average price per product is on average 5779 SEK in 2018.[3] The distribution of prices is however very skewed, with 84 percent of the prices in 2018 below 5000. I show the distribution of prices in more detail below. The increase in average prices reflects the increasing availability of costlier products online. As shown later, repeated listings barely increase in price. The products for which there are repeated listings, however, are somewhat more costly than other products.

Panel C presents some statistics at the retailer level. The average number of products per retailer declines over time. The decline is unsurprising as we see less products in total in 2019 while the number of retailers is roughly the same. The retailers that appear in the matched sample have almost as many products on average as the 2018 sample. The retailers in the matched sample participate in almost 25 different product categories, on average, slightly more than the average in either 2018 or 2019. We return to the time variation in the number of categories per retailer in the matched sample further below.

Panel C shows descriptive statistics by product category. The number of products declines over time. The categories in the matched sample decline somewhat less

[3] 10 SEK $\approx$ 1 USD

Table 1: General descriptive statistics

	2018	2019	Matched
	Panel A		
#Products	898,048	759,178	569,121
#Retailers	3,303	3,239	2,511
#Categories	951	979	919
#Listings	3,557,546	2,816,314	1,767,269
	Panel B		
By product	2018	2019	Matched
#Retailers	3.96	3.71	4.64
	(5.55)	(4.89)	(5.95)
Mean price in SEK	5778.79	6243.20	6651.29
(10 SEK ≈ 1 USD)	(56503.60)	(57506.57)	(65127.92)
	Panel C		
By Retailer	2018	2019	Matched
#Products	1077.07	869.50	1037.46
	(4523.16)	(3763.66)	(4137.12)
#Categories	23.10	21.94	24.63
	(47.37)	(44.7)	(47.74)
	Panel D		
By Category	2018	2019	Matched
#Products	944.34	780.13	885.69
	(2928.40)	(2416.66)	(2698.20)
#Retailers	80.24	72.58	78.24
	(68.69)	(63.09)	(65.40)

Descriptive statistics at different levels of the data. Panel A presents counts. Panels B, C, and D presents means with standard deviation in parentheses. The "matched" column includes only products/retailers/categories that have at least one repeated listing. The statistics however include information also about non-repeated listings. For example, the top row in Panel B computes that the average number of retailers offering each of the products that occur in the matched sample is 4.62. It does not mean that there are on average 4.62 repeated listings for these products. Some of the 4.62 retailers may offer the product only in one of the years. Because the number of retailers offering the product can differ depending on the year, the number presented is the average of both time periods.

than the average category. I show further below that in 2018, the categories in the matched sample are slightly larger than the average category. The number of retailers per category is roughly 80 in 2018 but less than 73 in 2019.

Figure 1 presents the distribution of average price per product. The distribution but only for those products that have at least one repeated listing appears in front. For visualization purpose, I disregard the 11 percent of products that have an average price above 8,000 SEK. Clearly, the mean price presented in Table 1 is driven by products of very high price that are not even shown in this figure. The figure shows that most products have prices below 1000 SEK. The mode is 99 SEK. The products that have at least one repeated listing, and thus are subject to analysis, account for roughly the same proportion of all products throughout the distribution.

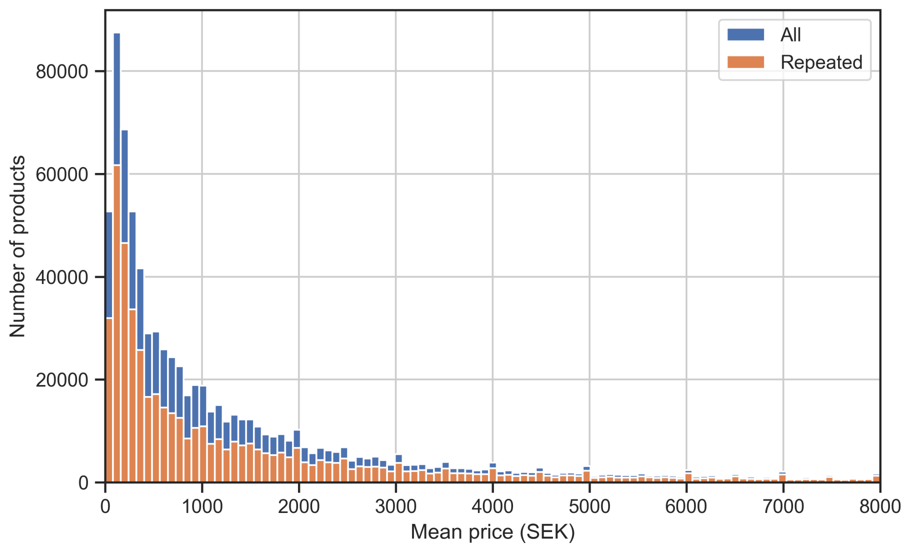

Figure 1: **Distribution of prices in 2018.** Distribution of average price per product in 2018 for all products and for products with at least one repeated listing. 10 SEK ≈ 1 USD. 11 percent of all products have average prices above 8,000 SEK and does not appear in the histogram.

I compute the measure of MMC in accordance with the empirical literature, but at a finer level. First, the number of contacts is defined for each retailer pair. For each retailer pair, at time t, the number of contacts is the number of categories in which they both sell at least one product. Table 2 presents the number of contacts between six example retailers. To avoid a dominant part of the shown elements to be zero, I randomly draw from retailers that participate in at least 20 categories. These retailers are thus relatively large or differentiated, but they serve only the purpose of illustrating the number of contacts. The matrix is only a small part of the full matrix that includes all retailers. As shown, each retailer indexes one row, and one column. Each element in the matrix contain the number of categories that row and column retailers participate in concomitantly, in year 2018. A corresponding matrix exists for the number of contacts in 2019. Because the diagonal elements has the same retailer as row and column index, these elements are the total number of categories that the retailer participate in.

Table 2: Contact matrix

	Naturkompaniet	Fitnessbutiken	Northsport	Nordiska Galleriet	My RFM Brands	Eye Of Beauty
Naturkompaniet	83	13	30	2	4	3
Fitnessbutiken	13	40	8	0	0	0
Northsport	30	8	40	1	0	0
Nordiska Galleriet	2	0	1	42	0	0
My RFM Brands	4	0	0	0	37	26
Eye Of Beauty	3	0	0	0	26	30

Examples of retailer pairs' number of contacts in 2018. Diagonal elements give the total number of categories that the retailer participates in. Off-diagonal elements give the number of contacts, i.e. the number of categories that both retailers participate in.

Naturkompaniet, as an example, is a retailer that offers products in many (83) different categories. Most of these are in the Sports & Outdoors group. Fitnessbutiken and Northsport sell products related to sports activities, and thus share some categories (13 and 30) with Naturkompaniet. The number of contacts for Naturkompaniet and Fitnessbutiken is thus 13. Nordiska Galleriet sell mainly interior design products that belong to the Home & Garden group. My RFM Brands and Eye Of Beauty mainly sell cosmetics and beauty products that belong to the Beauty & Health group. Therefore, Nordiska Galleriet barely has any contact categories with any of the other five retailers shown in the matrix. My RFM Brands and Eye Of Beauty have contact with each other in most of the categories they serve, but almost no contacts with the other retailers.

Second, to describe retailer i's degree of MMC in a given category, I take the average number of contacts across all other retailers in that category. Thus, retailer i's average MMC when selling product p in category c, at time t, is given by,

$$\text{avg\_MMC}_{ict} = \frac{1}{(N_c - 1)} \sum_{j \neq i}^{N} 1\{j \in \mathcal{C}_{ct} \forall t = 1, 2\} * \#\text{contacts}_{ijt}$$

where j indexes other retailers than i. N_c is the number of competing retailers in category c, and N is the total number of retailers in the sample. $\#\text{contacts}_{ijt}$ is the number of contacts for the two retailers i and j at time t, i.e. element ij of the contact matrix for year t. $1\{j \in \mathcal{C}_{ct} \forall t = 1, 2\}$ is an indicator function that equals one if j participates in category c in both time periods, and zero otherwise. The measure is thus defined at the retailer-by-category level in each time period.

To understand how the avg_MMC change over time, consider the following example. Suppose that Naturkompaniet, Fitnessbutiken, and Northsport are the only retailers that participate in category c in both time periods. In 2018, the avg_MMC for Naturkompaniet in category c is thus $(13 + 30)/2 = 21.5$. Suppose that during the following year, Fitnessbutiken exits four of the categories it shared with Naturkompaniet, and that Northsport enters three categories that Naturkompaniet also participates in. Hence, in 2019, the new avg_MMC for Naturkompaniet in category c is $(9 + 33)/2 = 21$. It thus decreases by 0.5.

It is important that the avg_MMC is an average only over those competing retailers that participate in the category in both time periods. This limitation is important for two reasons.

First, the theory of MMC and collusion regards repeated interaction for an infinite (or indefinite) time horizon. Thus, if the retailers do not expect to interact for many time periods, avg_MMC should have no effect. Second, if I would include all retailers in category c to compute avg_MMC, the change in this measure would partly be driven by firms entering and exiting category c. Because entries and exits in category c may have other effects on the price retailer i sets in category c, inclusion of all retailers could bias the estimated effect on price. Schmitt (2018) describes this point in more detail.

Most papers in the literature study the average of the avg_MMC across all firms in a market. They thus have a measure that varies at the level of market by time period. However, to exploit the granularity of the data, I analyze at the finest level. The main measure of MMC defined at the retailer-by-category level, or in general terms firm-by-market level. This level of analysis ensures that it is indeed the retailers that experience an increase in MMC that increases their prices. The benefit of aggregating to the market level is to draw conclusion for aggregate market behavior. Because I only have two time periods such aggregate analysis is very noisy, despite the large number of markets.[4]

The avg_MMC thus describes a retailer's average number of contacts with its long term competitors in the category of observation. The unit of observation is however even finer, as retailers typically sell multiple products within the same category. Because avg_MMC is defined at the retailer-by-category level, Table 3 provides some descriptive statistics at that level, and at the category level.

Panel A of Table 3 presents descriptive statistics at the retailer-by-category level in 2018. On average, the avg_MMC is 38.4. A closer look at the distribution shows that most observations have lower avg_MMC, with a median at 25.7. The number of listings shows that retailers have on average almost 50 listed products in a category they participate in. However, this distribution is very skewed, and the median is only 7. The average number of repeated listings within a category is 35.1 per retailer.

Panel B of Table 3 presents some descriptive statistics at the category level for the categories that occur in both time periods. In 2018, there are on average 82 retailers per category. The number of retailers vary considerably, and most categories have less than 82. The categories have on average nine retailers in 2019 that did not participate in that category a year before. There are however 15 retailers that participate in the category, on average, in 2018 but not in 2019.

The category sizes, in terms of products, range widely. The average of 960 products per category is driven by some categories that have many products. The median is 260, while many categories have less than 100 products. The average number of brands is 33, but here as well, the median is much lower than the average. Thus, there

[4] At the aggregate level, I cannot reject the null hypothesis of no effect.

Table 3: Descriptive statistics - retailers and categories

Panel A				
By retailer×category	Mean	p25	Median	p75
avg_MMC	38.4	11.9	25.7	49.6
#listings	49.9	2	7	26
#repeated listings	35.1	2	5	17
Panel B				
By category	Mean	p25	Median	p75
#retailers	81.9	36	65	110
#entries	9.4	2	6	13
#exits	15.4	6	12	20
#products	959.7	81	259.5	850
#brands	33.3	7	17	39

All rows but #entries and #exits describe data in 2018. Includes retailer×category combinations that occur in both 2018 and 2019.

Table 4: Variation in price and multimarket contact

Price	Mean	p25	Median	p75
Δprice	130	0	0	45
Δln(price)	.006	0	0	.053
Δprice>0: 37.1%			Δprice<0: 24.3%	
Multimarket contact	Mean	p25	Median	p75
Δavg_MMC	.775	−.535	.355	2.519
Mean #new contacts	6.55	1.286	5.255	10.228
Mean #lost contacts	5.775	1.157	3.534	9.000
#observations: 1,767,269				

Includes all repeated listings.

are typically multiple products per brand in a category.

Table 4 describes the variation in prices and avg_MMC. Prices are fairly sticky. The average price change per listing is 130 SEK ($\approx$ 13 USD), which is an average price change of 0.6 percent. 37 percent of listings increase in price, and 24 percent decrease. Thus, approximately 39 percent of listings have exactly the same price in September 2019 as in September 2018.[5]

The avg_MMC increases on average by 0.775. However, a decomposition of the change in avg_MMC into new and lost contacts shows that while the average number of contacts change only little, there are on average around six new and lost contacts.

3 Empirical Strategy

There are reasons to estimate the relationship between avg_MMC and price carefully. Most obviously, large retailers offer products in many different categories. It is also plausible that large retailers might price differently than small retailers. If large retailer consistently charge higher prices, e.g. because of reputation or market power, avg_MMC would correlate positively with prices because of retailers' sizes. Because I observe the retailers' sizes, this is not a problem. There are, however, unobserved differences between retailers which might correlate with both avg_MMC and prices. In addition, there are many unobserved characteristics of the products and categories that could bias the relationship between avg_MMC and price as well.

The standard approach in the literature is to take the average of avg_MMC across all firms in a market at each time period, apply market Fixed Effects (FE) and thus control for all time-invariant characteristics of the markets. However, most papers measures avg_MMC using all firms, even firms that enter or exit. Schmitt (2018) emphasizes that because such measure of MMC varies with the composition of firms in the market, it is probably endogenous. If a large retailer that participates in many markets enters, avg_MMC most likely increases. Such entry may however affect prices for other reasons than MMC. Similarly to Schmitt (2018), the measure in this paper varies only with "out-of-market consolidation".

We could also expect retailers set higher prices for products of their specialization. Retailers may be better at providing services or have better reputation regarding some products, which allows them to charge higher prices. Such specialization would likely correspond to offering similar products in other categories in the same group. Because other retailers with similar specialization most likely participate in many of those similar categories, retailers have higher avg_MMC in the categories of their specialization. To solve this problem, I extend the standard FE approach and apply listing FE. I thus hold all time-invariant retailer-by-product characteristics constant. Retailer specialization is therefore not a concern, unless it varies over time.

Some characteristics that do vary over time could still cause bias. Retailers could re-focus their specialization, or generally expand or retract from some markets. Therefore, I specify an empirical model that exploits within-listing variation over time, while also controls for time-varying characteristics.

[5]The inflation rate in Sweden for the time period between data collection was 1.5 percent.

Because I have only two time periods, the FE specification is equivalent to taking first differences within each listing. I then estimate the following linear equation:

$$\Delta lnP_{ip} = \alpha + \beta\Delta\text{avg\_MMC}_{ic} + \gamma\Delta\mathbf{X}_{ipc} + e_{ipc} \tag{1}$$

where i indexes the retailer, p the product, and c the product category. P_{ip} is the price for product p set by i. Δ denotes that each variable is the first difference within listing over time. $\mathbf{X}_{ipc}$ is a vector of control variables that vary over time. This specification thus accounts for all potential bias that stems from retailer, product, or category characteristics that are fixed over time. In addition, it controls for time-varying variables that poses additional threats.

The time-varying variables included in $\mathbf{X}_{ipc}$ are retailer i's number of products, i's number of accepted payment methods, i's "share" of category c (i.e. $\#products_{ic}$ divided by $\#products_c$), the number of retailers in category c, the number of products in c, c's popularity[6], the number of retailers selling product p, i's number of entries (new categories), i's number of exits (dropped categories), i's number of entries in categories that belong to the same group as c, and i's number of exits from categories in the same group as c.

Because I include retailer i's number of entry and exit in other categories, general expansion or retraction poses no risk to identification. Furthermore, any relevant change in specialization would increase or decrease retailer i's number of categories in the same group. The number of entries and exits that are in the same group as category c thus eliminate such risk as well. Another reason for entries and exits in similar categories could affect pricing in category c is complementarities between products. In some situations, selling complementary products may enable the retailer to charge a higher price than it could otherwise. In online retail, however, it is very easy for customers to multi-stop shop, i.e. purchase different products from different retailers. While the control variables exits/entries in the same group accounts for such alternative explanation, the results also indicate that such entries and exits do not affect the price.

The identifying variation comes from decision to entry and exit categories where competitors also participate. One alternative explanation for an estimated effect of avg_MMC on price could be that retailers respond to positive cost shocks by entering new markets to recover lost profits. To explore this explanation I decompose the change in MMC into two parts - new contacts and lost contacts. To understand more clearly, consider that for retailer i in category c, in 2019 there are some new contact categories that i did not have in 2018. For each competitor j, the number of such contact categories is the number of new contacts. I then compute the average over all competitors j that participate in category c in both time periods. I compute the equivalent average number of lost contacts. The net difference between new and lost contacts is the change in avg_MMC.

The effect is driven by lost contacts. Therefore, this alternative explanation would need retailers to exit other product categories in response to lower costs. Furthermore, the retailers need to exit specifically those product categories where other retailers selling the affected good also participate. It is not clear how to rationalize such exit behav-

[6]the average rank of the ten most popular products in the category, determined by user activity on the website

ior. A retailer would be forced to leave one category because another becomes more profitable only if there are substantial capacity constraints. Additional pages on an online store are not subject to such constraint and inventory typically need not to be large. Therefore, such alternative explanation should not be of significant concern.

Because the effect is driven by exits, I discuss that case in particular. Retailers exit product markets when they are no longer profitable. That may for example be because entering retailers compete away profits or because of decreasing demand. For the exact same reasons, contact categories that facilitated collusion before, might not anymore. Thus, exit decisions are plausibly driven by the same factors as those that determine whether MMC can sustain collusion. Therefore, retailers do not randomly exit markets that they could leverage to sustain collusion. Something changed in those markets that likely destabilized collusion and even made some retailers exit.

It is not clear why entry and exit in other categories should affect pricing in category c, apart from strategic interaction with competitors. Some papers in the literature use instrumental variables (IV) estimation to eliminate bias remaining even after the fixed effects. Ciliberto & Williams (2014) rightfully acknowledge that if any time-varying characteristic determines both the measure of MMC and prices, the FE estimate is still biased. They do not, however, explain what such time-varying characteristic might be. Their IV estimate is much higher than their FE estimate, which suggest that the bias the FE cannot eliminate is negative. Alternatively, the IV estimates a local effect which is different from the average effect estimated by the FE because of treatment heterogeneity.

4 Main Results

Table 5 presents the main results. Column 1 shows the estimated coefficient of avg_MMC without controls. When I include the controls and estimate the main specification in equation (1) the coefficient moderates somewhat, as shown in column 2. The estimated effect of avg_MMC on price is thus 0.002. It means that if avg_MMC increases by one, price increases by 0.2 percent. This effect might seem rather small, but it is actually quite significant in relation to the literature. Ciliberto & Williams (2014) finds that a change in MMC that corresponds to moving from the twenty-fifth to the seventy-fifth percentile, leads to a percentage increase in price of between 1 and 8.5 percent depending on the specification. The equivalent comparison in Evans & Kessides (1994) is 5 percent, and in Schmitt (2018) it is no more than 1 percent. In this setting, an increase in avg_MMC from the twenty-fifth to the seventy-firth percentile leads to a price increase of 7.5 percent. The effect for online retailers is thus relatively strong.

The estimates in column 3 show that the effect of avg_MMC is driven by the lost contacts. Both coefficients have the expected sign, but only the coefficient for lost contacts is statistically significant. It means that the estimated effect of avg_MMC is mainly driven by retailers that lower the price when they lose contact categories. This finding contradicts the alternative explanation that firms enter new product categories in response to positive cost shocks. If that was a significant driver of the estimated coefficient of avg_MMC on price, the effect of new contacts should be much larger

Table 5: Results

	Δln(price)		
	(1)	(2)	(3)
Δavg_MMC	.0027***	.0020**	
	(.0006)	(.0009)	
#New contacts			.0011
			(.0009)
#Lost contacts			$-.0027^{***}$
			(.0009)
Controls	NO	YES	YES
Observations	1,767,269	1,767,269	1,767,269
R^2	.002	.008	.008

Clustered standard errors in parentheses (on retailer×category). *** p<0.01, ** p<0.05, * p<0.1. Controls include retailer i's number of products, i's number of accepted payment methods, i's "share" of category c, the number of retailers in category c, the number of products in c, c's "popularity", the number of retailers selling product p, i's number of entries (new categories), i's number of exits (dropped categories), i's number of entries in categories in the same group as c, i's number of exits from categories in the same group as c.

than that of lost contacts. It is not.

The lost contacts may drive the effect because collusion makes firms inefficient and less innovative. Given that a contact that facilitates collusion must be collusive itself, the pivotal lost contacts must have been collusive. However, if the collusive profits attract entrants that compete fiercely, collusion destabilizes. If collusion made the incumbents inefficient, sudden fierce competition may force them to leave the category altogether.

5 Heterogeneity

5.1 Number of firms in the contact market

A key result in Bernheim & Whinston (1990) is that not any contact facilitates collusion. There must be differences between the firms or the markets for MMC to facilitate collusion. The first type of difference that they show enables MMC to facilitate collusion is the number of firms in the contact market. For example, if two firms meet both in market B, where there are few firms, and in market A, where there are more firms, they may transfer ability to collude from market B to market A. Given that collusion is stable in market B, when considered in isolation, the threat of additional retaliation in that market may sustain collusion in market A. Therefore, the effect of MMC on prices should be higher when the contact markets have relatively few retailers.

A traditional approach to estimate such effect heterogeneity is to interact Δavg_MMC with a heterogeneity variable that describes how many retailers participate in the contact markets. However, the effect might vary non-linearly with the heterogeneity variable and the heterogeneity is thus poorly represented by the interaction term. A common alternative is to split the sample into subsamples based on the heterogeneity variable, and present estimates for the different subsamples. While the subsample approach may reveal non-linearities, it is susceptible to multiple testing issues and spurious results due to arbitrarily chosen splits.

Therefore, I employ a recently developed technique called Generalized Random Forest (GRF) by Athey et al. (2019). This technique is an extension of the causal forest (Wager & Athey, 2018), and builds on the causal tree in Athey & Imbens (2016). I describe the algorithm in detail in the appendix, and only briefly outline the procedure here.

The GRF splits the sample into subgroups in which it estimates the partial effect of avg_MMC on prices with equation 1. It determines the subgroups in a principled way that maximizes effect heterogeneity across the subgroups. To avoid spurious findings, it uses distinct samples to determine the subgroups and to estimate the effects. Athey & Imbens (2016) call this feature "honesty". Splitting the sample into subgroups once is called growing a tree, because this procedure is a modification of the Random Forest algorithm in Breiman (2001). Each tree defines a set of subgroups, also called leaves, by sample splits based on a set of heterogeneity variables Z. Each tree gives predicts a specific $\hat{\beta}$ for each leaf, and that prediction applies for every observation that falls into that leaf. Bootstrap sampling and randomly selected subsets from Z make the trees different from each other. With many trees, the average prediction, for a given set of $z_1, z_2, ..., z_p \in Z$ provides an asymptotically unbiased prediction of the conditional partial effect.

The GRF is a powerful tool, mainly because it allows the researcher to search over a high-dimensional space of heterogeneity variables, while it still allows for statistical inference. More specifically, there might be many variables that could determine effect heterogeneity, especially considering all possible interactions and non-linear transformations of the heterogeneity variables. The GRF is designed to search over a wide range of non-linear combination of many heterogeneity variables to identify the variables and interactions that are predictive of estimated effects. Because it uses separate samples to identify relevant variables and estimate the effects, it still allows for statistical inference. In this section, I use the GRF in combination with traditional subsample estimates to draw conclusions about the number of retailers in the contact markets. In the next section, I use the algorithm to explore additional important heterogeneity.

Over time, each retailer pair may gain and lose contact markets. For each retailer by category, I calculate the average number of retailers that participate in the new/lost contact markets across all retailers in that category. For example, consider retailer A that sells products in the product category USB-sticks. In this category, suppose A has 49 competitors so that there are 50 retailers in total in 2018. Next year, suppose A has on average one new contact market and zero lost contact markets. The new contacts could facilitate collusion in the market for USB-sticks, but only if there is slack punishment ability in the new contact markets. According to Bernheim & Whinston (1990), that is more likely if the contact markets contain fewer firms. Therefore, I measure the difference between the number of retailers in the product category (USB-sticks in the example) and the average number of retailers in the new contact categories:

$$\overline{\Delta N}_{new} = \frac{1}{K} \sum_{k=1}^{K} (N_c - N_k)$$

where N_c is the number of retailers in the product category of the observation, N_k is the number of retailers in the new contact category k, and K is the number of new contact categories. The measure thus tells how many more retailers participate in the product category than in the categories that are new contact categories. For computation of this measure, I only use data from 2018. I calculate the equivalent for lost contacts.

In the example, suppose that for the retailer that sells USB-sticks, the new contact categories contain on average 40 retailers. Because there are 50 retailers that sells USB-sticks, the average difference is 10. Reversely, if the contact categories contain more retailers than the category of the observation, the difference is negative. We expect that the contacts are more likely to facilitate collusion if the difference in number of retailers is positive. The contact categories are then less concentrated and more likely to have slack punishment ability.

To test whether the difference in number of retailers determine effect heterogeneity of avg_MMC on price, I fit a honest GRF with 1000 trees on the full sample. I set minimum leaf size to 200 and the number of variables to consider for each split (m) to ten. The total number of heterogeneity variables is 30, including $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$. I list all the heterogeneity variables with explanations in the appendix. For a set of $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values, I predict out-of-bag (OOB)[7] the $\hat{\beta}$ for each observation that would have been with that set of $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values. I then take the average prediction for that set of $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values. I repeat the prediction procedure for a range of $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values for comparison.

Figure 2 visualizes the average prediction for different combinations of $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values within the 10th and 90th percentile of the variables. The predictions increase with the difference, which means that the contact categories need to contain considerably fewer retailers for the effect size to be large.

The advantage of this type of predictions, i.e. OOB predictions on the full sample, is that I can present the average prediction for the full sample. These predictions thus correspond to what would happen on average in a representative sample that suddenly gained or lost contacts, given the number of retailers in the contact categories. However, it is computationally expensive to predict effects for all observations. It is computationally expensive even though the forest is relatively small (1000 trees) and the predictions use on average only half of the trees because of OOB predictions. Furthermore, Athey et al. (2019) states that relatively small forests provide accurate predictions but that much more trees are needed for valid confidence intervals. With computational restraints, there is thus the following trade-off. Either I predict for a large and representative sample and discuss average predictions (as in Figure 2), or I predict valid confidence intervals but for much fewer observations. I choose to predict average effects for the full sample, thus providing representative predicted effects. I complement the representative estimates with traditional subgroup estimation of equation 1.

[7] Predicts for each observation l using only the trees in which l was neither used for fitting nor estimation.

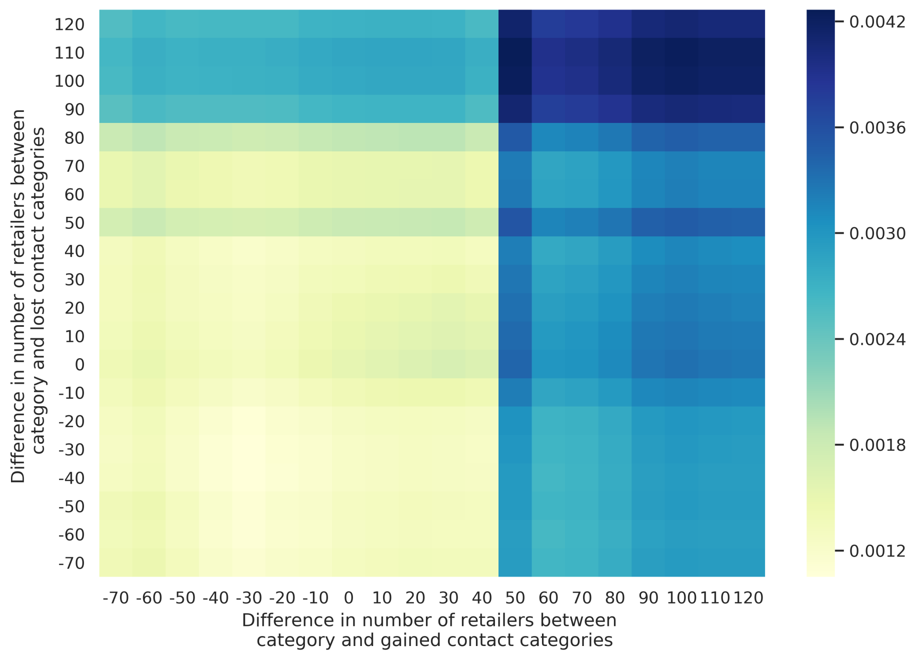

Figure 2: **Difference between the number of retailers in the category and in the contact categories.** Average out-of-bag predictions of the conditional partial effect for different sets of values of the difference in the number of retailers in the category and the number of retailers in the new/lost contact categories. The honest GRF has 1000 trees, minimum leaf size is 200, and the number of variables considered at each split is ten. Total number of heterogeneity variables included is 30. Darker colors mean higher effect.

Table 6 presents such complementary regression results. The first column estimates equation 1 on the full sample, i.e. the same as column 2 in Table 5. The second column estimates the same equation but only for the observations with $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values above zero. Columns 3 and 4 further increase the threshold of these values, meaning that contact categories contain relatively few retailers. The estimate increase with the threshold, just as the GRF predictions. This type of analysis combines the GRF with traditional sample splitting, and it is reassuring that they agree more or less completely.

Table 6: Heterogeneity results

	Δln(price)			
	(1)	(2)	(3)	(4)
Δavg_MMC	.0020**	.0026**	.0035**	.0042**
	(.0009)	(.0011)	(.0014)	(.0019)
Observations	1,767,269	807,530	382,971	277,022
R^2	.002	.010	.013	.013
Mean difference in N	Any	>0	>50	>75

Clustered standard errors in parentheses (on retailer×category). *** p<0.01, ** p<0.05, * p<0.1. Mean difference in N denotes the mean difference in number of retailers in the category between the category of the listing and the new or lost contact categories. The condition is for both new and lost contact categories. Full set of controls in all columns.

5.2 Exploration of heterogeneous effects

As described, the GRF allows a completely data-driven and theory-agnostic search for heterogeneous effects, without invalidating statistical inference. Because the data covers a very wide range of product types and firms, it is likely that the effect of MMC on prices is heterogeneous. In this section, I use the GRF to explore heterogeneity as follows. First, I predict the individual effect, i.e. for each observation, with valid pointwise confidence intervals.[8] These predictions give the full distribution of predicted effects of MMC on price. Second, I let details of the fitted GRF suggest what heterogeneity variables are the most important. Last, test for what levels of the most important variables is the effect different from zero.

To estimate valid confidence intervals for each observation in the sample, the forest must be large. I therefore estimate a GRF with 20000 trees. For out-of-sample prediction for each observation, I apply the following cross-fitting procedure. First, I randomly split the sample into ten folds. Then, I fit a GRF on the first fold, and predict the individual effect with confidence interval for the observations in the other nine folds. I repeat the fitting and estimation for each fold, so that each observation is used

[8]For each observation, the pointwise confidence interval may test the null hypothesis that the effect is zero. However, the difference between observations cannot be tested.

to fit exactly one GRF, and for prediction by the nine other GRFs. Each GRF is honest and has minimum leaf size of 500, $m = 10$, and 20000 trees.

For the out-of-sample individual predictions, each observations is predicted by the nine GRFs for which the observation was not used for fitting. For each prediction, the GRF also estimates the variance. For each observation, I average the nine predictions and variances. I use the average variances to compute 95 percent confidence intervals around the average predicted effect.

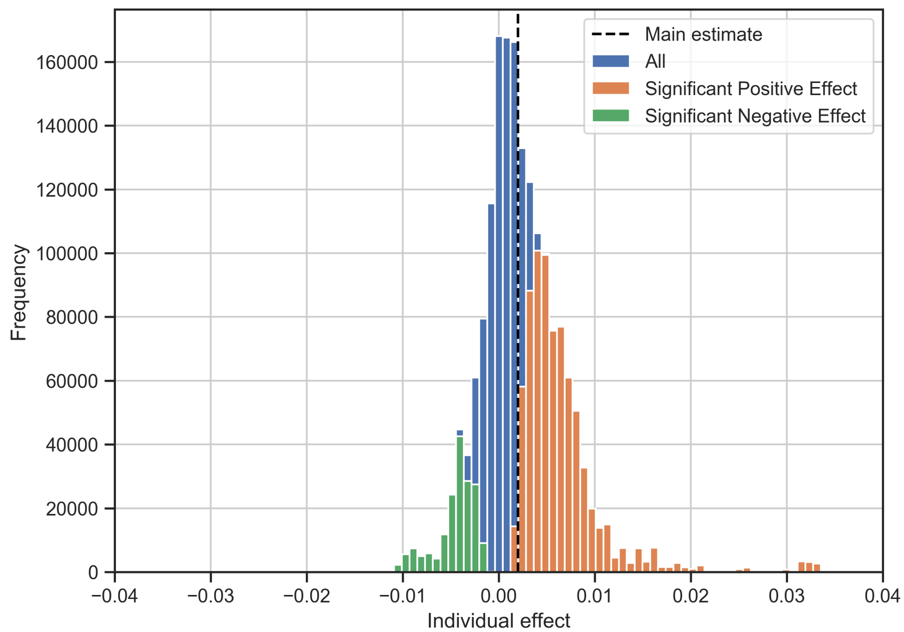

Figure 3: **Predicted conditional partial effects.** Histogram of average GRF predictions over the nine folds that the observation does not belong to. Each GRF is honest, fitted on a distinct ten percent of the sample, has 20000 trees, minimum leaf size of 500, and the number of variables considered at each split is ten. The histogram distinguishes observations that have significantly positive or negative effects as well as effects that are not significantly different from zero. The dashed line marks the average effect estimated by equation 1 on the full sample.

Figure 3 presents a histogram of the predicted effects at the individual listing level. Many of the listings have individual effects between zero and the average effect of 0.002. However, many listings have significantly positive effects much higher than the average effect. Notably, many predicted effects are several times higher than the average effect. Thus, while most listings have a small or zero (or even negative) effect of avg_MMC on price, some have a quite substantial effect. This result is obviously in line with the prediction of Bernheim & Whinston (1990) that only in some circumstances may MMC facilitate collusion.

The GRF provides information about the importance of the different heterogeneity variables. Similarly to standard Random Forest packages, the variable importances (typically called feature importances) depend on how often the variable is chosen for a split. A variable that is chosen more often thus has higher importance measure. Because the importance measure could be calculated in many different ways, we should not consider the measure as a result to any test. However, it provides a good indication of which variables are worth exploring.

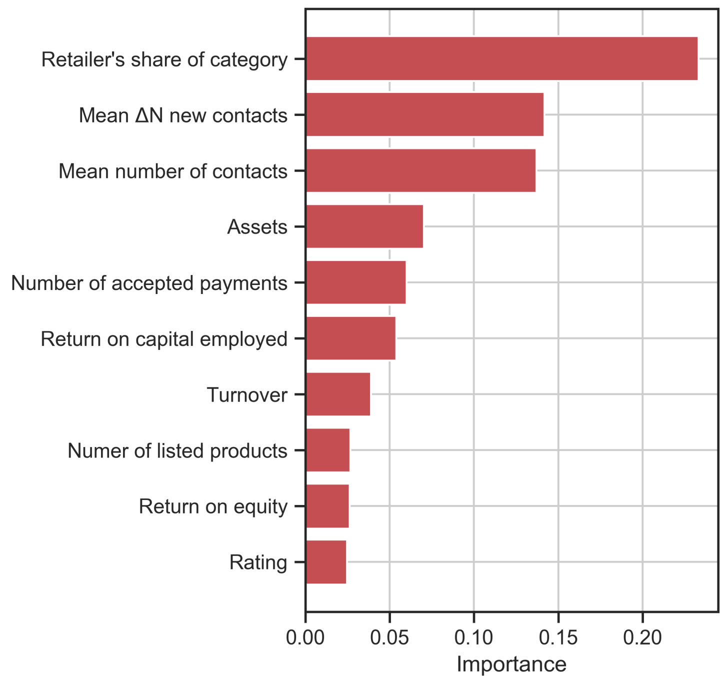

Figure 4: **Variable importances**. Importance of variables in predicting the effect of avg_MMC on ln(price). Determined by how often a variable is chosen for a split.

Figure 4 shows the importance measures of the ten variables with the highest measures. The importances are concentrated to just a few variables. The top variable is the retailer's "share" of category c, i.e. the number of products the retailer sells in the category divided by the total number of products that exist in the category. The second most predictive variable is already analyzed in Figure 2. The third variable is the mean number of contacts (in 2018). The rest of the variables are of much lower importance, and therefore I do not explore them further. We can, however, see that they are all about the retailers and not about the products or the categories.

Among the three most important variables, two are yet to explore. To estimate how the effect size vary with a variable, I simulate a listing with each heterogeneity variable set to the mean value. I then set the variable of interest to a specific value, and make one prediction (with variance estimate) with each of the ten large GRFs. Finally, I calculate the average prediction and variance estimate. I repeat the prediction for a range of values for the variable of interest to compare how the effect varies with that variable. The GRF in section 5.1 predicts the *average effect* for a specific set of $\overline{\Delta N}_{new}$ and $\overline{\Delta N}_{lost}$ values. In this section, the GRFs predict the *individual effect* for a specific value of the variable of interest when all other variables are at the mean.

Figure 5 visualize how the predicted effect of avg_MMC on price varies over different levels of the retailer's share of the category. This variable equals one if the retailer sells all the products in the category, and close to zero if it only sells a few. One could think that dominant firms in the category sell many products, while less dominant retailers only sell a few. Given that I do not observe any quantities, true market shares are unknown. The GRF result in Figure 5 shows that retailers that sell most of the products in the category have much higher individual effects of avg_MMC. While an observation with mean values for all variables might not be representative, the observed difference is huge. Listings by retailers that sell at least 60 percent of the products in the category enjoy far larger effects than the rest.

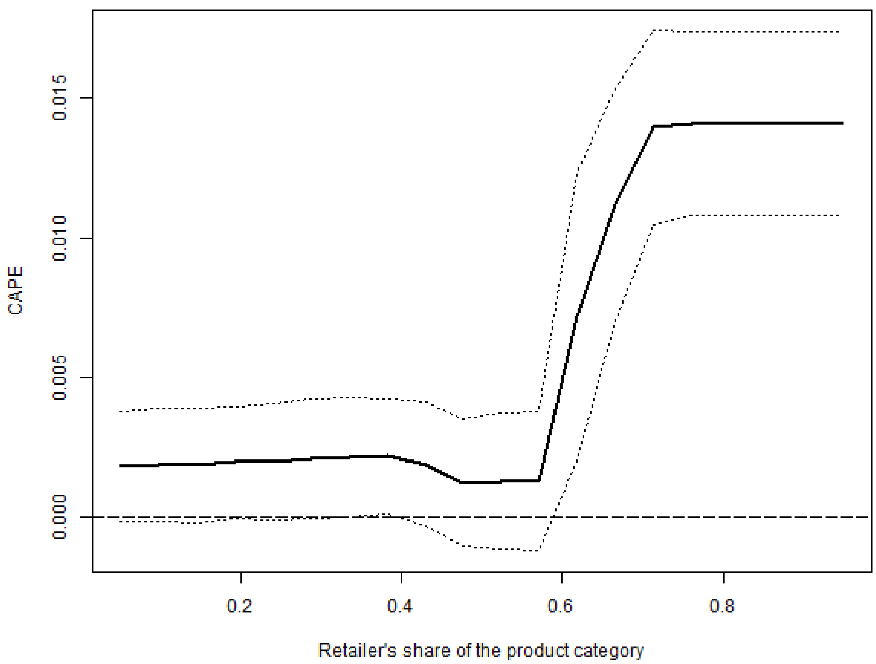

Figure 5: **Retailer's "share" of the category**. Predictions of the conditional average partial effect for different levels of the retailer's share of the category (i.e. $\#products_{ic}/\#products_c$). The GRF is honest, has 20000 trees, minimum leaf size is 500, and the number of variables considered at each split is ten. The other 29 variables are set to their mean value. Dotted lines show 95 percent pointwise confidence intervals.

Figure 6 shows how the effect varies with the number of initial contacts. This figure shows that if the retailer has many contacts, the effect of an additional contact is lower. This result should not be surprising given the mechanism at work. In Bernheim & Whinston (1990), there are two firms and two markets. In the empirical setting of this paper, however, there are many firms and many markets. When a pair of retailers get an additional contact market, it may facilitate collusion in market c if there is slack punishment power in the contact market and this slack is transferred to market c. However, if there are already many contact markets, the slack punishment power could be transferred to any of the existing contact markets other than c. Therefore, if there are many other markets, it is less likely that the additional contact facilitates collusion in market c.

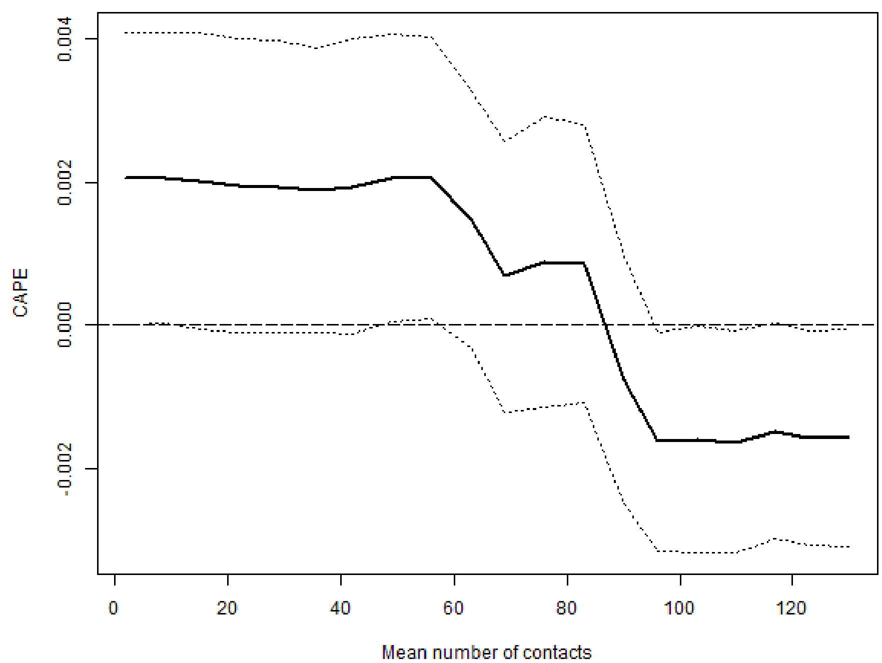

Figure 6: **Mean number of contacts**. Predictions of the conditional average partial effect for different levels of the retailer's average number of contacts with other retailers in the category in 2018. The GRF is honest, has 20000 trees, minimum leaf size is 500, and the number of variables considered at each split is ten. The other 29 variables are set to their mean value. Dotted lines show 95 percent pointwise confidence intervals.

6 Conclusion

This paper is the first to show that multimarket contacts increases prices for multiproduct online retailers. Moreover, this is the first paper to consider multimarket contact across markets separated by product types rather than geography. Despite promises of perfect competition when consumer can easily search prices, I show that the online retailers leverage multimarket contact to sustain collusion even more than the literature has shown for, e.g., airlines, hotels, banks, and radio broadcasters.

The main result is that multimarket contact increases prices, normally interpreted as facilitation of tacit collusion. Specifically, a change in multimarket contact from the 25th to the 75th percentile increases the price by 7.5 percent. By decomposition of the MMC measure, I show that this effect is driven by the lost contacts. This result excludes some alternative explanations regarding entry/exit responses to cost shocks.

The generalized random forest enables me to carefully explore and test for heterogeneous effects. I show that the effect is higher when the contact markets contain fewer retailers. This result in strongly in line with the first setting in which Bernheim & Whinston (1990) show that multimarket contact may facilitate collusion. Additional heterogeneity exploration shows that retailers that cover a wide range of the different products in the product market and have fewer contacts to begin with can more easily leverage multimarket contact to sustain higher prices.

The main policy implication is that competition authorities need be cautious when assessing online markets during merger control. Market shares of merging firms in the markets where they both participate may suggest no competition concerns, but says nothing about multimarket contact and tacit collusion. In fact, when the merging firms operate completely different markets, the risk of multimarket contact is the highest. The merged entity's wider coverage of product markets increases its multimarket contact with other firms and thereby facilitates collusion. The digital economy thus faces an increased risk of tacit collusion, but also new tools to combat it. If authorities collect data from the web, they can easily measure multimarket contact, track prices, and identify markets susceptible to collusion.

References

Athey, S. & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Aysan, A. F., Disli, M., & Schoors, K. (2013). Bank competition and outreach: Evidence from turkey. *Emerging Markets Finance and Trade*, *49*(sup5), 7–30.

Bernheim, B. D. & Whinston, M. D. (1990). Multimarket contact and collusive behavior. *The RAND Journal of Economics*, 1–26.

Bilotkach, V. (2011). Multimarket contact and intensity of competition: evidence from an airline merger. *Review of Industrial Organization*, *38*(1), 95–115.

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

Busse, M. R. (2000). Multimarket contact and price coordination in the cellular telephone industry. *Journal of Economics & Management Strategy*, *9*(3), 287–320.

Ciliberto, F. & Williams, J. W. (2014). Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry. *The RAND Journal of Economics*, *45*(4), 764–791.

Coccorese, P. & Pellecchia, A. (2009). Multimarket contact and profitability in banking: evidence from italy. *Journal of Financial Services Research*, *35*(3), 245–271.

De Bonis, R. & Ferrando, A. (2000). The italian banking structure in the 1990s: testing the multimarket contact hypothesis. *Economic Notes*, *29*(2), 215–241.

Edwards, C. D. (1955). Conglomerate bigness as a source of power. In *Business concentration and price policy* (pp. 331–359). Princeton University Press.

European Commission (2016). Preliminary report on the e-commerce sector inquiry. *Commission Staff Working Document*.

Evans, W. N. & Kessides, I. N. (1994). Living by the "golden rule": Multimarket contact in the us airline industry. *The Quarterly Journal of Economics*, *109*(2), 341–366.

Feinberg, R. M. (2014). Price effects of multimarket contact among movie chains in small us metropolitan areas. *Economics Letters*, *123*(1), 6–8.

Fernandez, N. & Marin, P. L. (1998). Market power and multimarket contact: Some evidence from the spanish hotel industry. *The Journal of Industrial Economics*, *46*(3), 301–315.

Jans, I. & Rosenbaum, D. I. (1997). Multimarket contact and pricing: Evidence from the us cement industry. *International Journal of Industrial Organization*, *15*(3), 391–412.

Kasman, S. & Kasman, A. (2016). Multimarket contact, market power and financial stability in the turkish banking industry. *Empirical Economics*, *50*(2), 361–382.

Miller, A. R. (2010). Did the airline tariff publishing case reduce collusion? *The Journal of Law and Economics*, *53*(3), 569–586.

Molnar, J., Violi, R., & Zhou, X. (2013). Multimarket contact in italian retail banking: Competition and welfare. *International Journal of Industrial Organization*, *31*(5), 368–381.

Parker, P. M. & Röller, L.-H. (1997). Collusive conduct in duopolies: multimarket contact and cross-ownership in the mobile telephone industry. *The RAND Journal of Economics*, 304–322.

Pilloff, S. J. (1999). Multimarket contact in banking. *Review of Industrial Organization*, *14*(2), 163–182.

Schmitt, M. (2018). Multimarket contact in the hospital industry. *American Economic Journal: Economic Policy*, *10*(3), 361–87.

Singal, V. (1996). Airline mergers and multimarket contact. *Managerial and Decision Economics*, *17*(6), 559–574.

Wager, S. & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Waldfogel, J. & Wulf, J. (2006). Measuring the effect of multimarket contact on competition: Evidence from mergers following radio broadcast ownership deregulation. *The BE Journal of Economic Analysis & Policy*, *5*(1).

Appendix

Generalized Random Forest for heterogeneous effects

The Generalized Random Forest (GRF) in Athey et al. (2019) is a modification of the Random Forest (RF) by Breiman (2001). The RF is one of the main algorithms in the supervised machine learning literature, developed as a tool to predict an outcome y given a set of characteristics X. The GRF modifies the RF algorithm to predict heterogeneous effects of some treatment on an outcome. Initially, Wager & Athey (2018) developed the Causal Forest that predicts heterogeneous effects of a binary treatment. The GRF is an extension that makes it possible to estimate heterogeneity in the partial effect of a continuous variable. The GRF algorithm that I use works as follows.

As in a random forest, the GRF grows many individual trees. The aim of each tree is to generate a function that maps any combination of heterogeneity variables $z_1, z_2, ..., z_p \in Z$ to a conditional partial effect $\hat{\beta} \mid Z = z$. The $\hat{\beta}$ is that in equation 1. Each tree uses only a randomly drawn half of the sample, denoted by $S \in \mathcal{S}$. S is then further split randomly into S_1 and S_2 of equal sizes. The fitting sample, S_1 is used to determine the structure of the tree. First, all observations in S_1 are split into two nodes, based on a value of one of the heterogeneity variables. To select the splitting value, the algorithm searches over m different variables in Z and for each potential splitting value, it estimates equation 1 in each of the two resulting nodes. It chooses splitting value that maximizes heterogeneity in $\hat{\beta}$ while penalizing poor precision in the estimate. The chosen split results in two new nodes in which the algorithm again searches over a set of variables to make a new split.

The set of variables that the algorithm searches over could include all the variables in Z, but it typically does not. The number of variables, m, that the algorithm is allowed to use is a hyperparameter set by the researcher. At each splitting point, m variables are randomly drawn from Z.[9] The random draw of a subset of heterogeneity variables de-correlates the trees, which improves prediction when averaging over many trees.

The tree is grown by further splitting until reaching a minimum number of observations in each node. The resulting nodes are called the end nodes, or leaves. Then, the prediction of the conditional partial effect in each leaf is the $\hat{\beta}$ from estimating equation 1 with the observations in S_2 that falls into the leaf. Thus, the GRF only uses S_1 to determine the splits, and only S_2 to estimate the effects. This sample splitting within each tree, called honesty, avoids spurious findings (Athey & Imbens, 2016).

To implement the described procedure, two steps are altered for computational feasibility. First, the GRF does not actually estimate equation 1 in the nodes for each potential split. Instead, it estimates the equation in the parent node once, and then uses a gradient based approximation in the nodes at each potential split. I further facilitate the estimation in each parent node by orthogonalization of ln(price) and MMC so that the estimation uses only residuals of ln(price) and MMC (in first differences) from linear regressions of the variable on the controls in equation 1. The orthogonalization is called "local centering" in Athey et al. (2019).

Second, it does not estimate the effect in the leaves using the S_2 observations. For each observation l and for each tree, it assigns equal weights to the S_2 observations that falls into the same leaf as l and zero weight to other observations. Then, it averages the weights over all trees so that each observation l has a set of weights α_l, which includes one weight for each observation included in the fitting. Finally, the algorithm estimates equation 1 for each observation l with weights α_l. Thus, instead of averaging over estimated effects, it averages weights to estimate the effect once per observation. Analogously, it is a nearest neighbor weighting estimator, where "nearest" means near in the variables that determine effect heterogeneity.

[9] In the standard RF, m variables are drawn. In the GRF, however, at each split point, the number of variables actually considered for the split is $\min\{\max\{\text{Poisson}(m), 1\}, p\}$, where p is the number of variables in Z. This selection of the number of variables to consider is to satisfy a condition called minimum split probability. See Athey et al. (2019) for more details.

Heterogeneity variables

Variable name	*Description*
Assets	Value of the retailer's total assets
avg_MMC$_{2018}$	The retailer's average number of contacts with competitors in the category in 2018
Brand count	Number of products from the listing's product brand
Brand's share of category	Number of products in the category that are from the listing's product brand divided by the total number of products in the category
Brand's share of retailer	Number of products in the retailer's product range that are from the same brand as the listing
Brand also retailer	Indicator: 1 if brand active as retailer, 0 otherwise
Category size	Number of products in the product category
Category's percentile rank	Category's percentile rank based on the average rank of the ten lowest ranked products in the category
Change in rank	The category's change in percentile rank from 2018 to 2019
Hybrid retailer	The retailer has at least one brick-and-mortar (physical) store
Mean change in rank of lost contacts	Self explanatory
Mean change in rank of new contacts	Self explanatory
$\overline{\Delta N}_{lost}$	Difference between the number of retailers in the product category and the average number of retailers in the lost contact categories
$\overline{\Delta N}_{new}$	Difference between the number of retailers in the product category and the average number of retailers in the new contact categories
Number of brands in the category	Self explanatory
Number of employees	Retailer's number of employees
Number of listed products	Retailer's number of products listed at the price comparison website
Number of payment methods	Retailer's number of accepted payment methods
Number of retailers offering the product	Self explanatory
Number of retailers in the category	The number of retailers that offer at least one product in the listing's product category

Product's percentile rank within category	Self explanatory
Profit margin	Retailer's aggregate profit margin
Profits	Retailer's total profits
Retailer's number of categories	The number of categories in which the retailer offers at least one product
Retailer's rating	Rating on the price comparison website (scale 0-10)
Retailer's share of the category	Retailer's number of products in the category divided by the total number of products in the category
Retailers per brand in category	Number of retailers in the category divided by the number of brands in the category
Return on capital employed	Self explanatory
Return on equity	Self explanatory
Turnover	Self explanatory