Information Economics and Policy 69 (2024) 101111

Contents lists available at ScienceDirect

Information Economics and Policy

journal homepage: www.elsevier.com/locate/iep

Algorithms in the marketplace: An empirical analysis of automated pricing in e-commerce ☆

Philip Hanspach [a,*,1], Geza Sapi [b,2], Marcel Wieting [c,3]

[a] European University Institute, Florence, Italy
[b] European Commission DG COMP and Düsseldorf Institute for Competition Economics (DICE), Heinrich Heine University of Düsseldorf, Germany
[c] Charles River Associates and KU Leuven, Belgium

ARTICLE INFO

JEL classification:
D42
D82
L42

Keywords:
Algorithmic pricing
Artificial intelligence
Collusion

ABSTRACT

We analyze algorithmic pricing on the largest online marketplace in the Netherlands and Belgium. Based on two months of pricing data for around 2800 products, we find no significant correlation between the use of algorithms and an increase in prices of the Buy Box (the most prominently displayed offer for a product). We document that the presence of an algorithmic seller in monopoly markets goes hand-in-hand with lower prices. This effect is likely due to algorithms correcting excessively high human-set prices. We describe several characteristic algorithmic pricing patterns. While some of these pricing patterns are consistent with algorithmic collusion, such practice appears to be a fringe phenomenon. Overall, our findings call for careful policy with respect to pricing algorithms that remains alert to the possibility of algorithmic collusion but recognizes that pricing algorithms may benefit consumers.

1. Introduction

The advance of digitization, big data processing and analysis triggered new applications of *algorithmic pricing*, whereby sellers automate price-setting using sophisticated software tools. The increased prominence of algorithmic pricing in consumer-facing markets such as retail gasoline and e-commerce has recently attracted the attention of academics, practitioners and policy advocates. The main concern is that ever more intelligent algorithms may learn to (tacitly) collude, refrain from competing aggressively and keep prices high.

Despite the policy debate around algorithmic pricing, empirical research on the ability of algorithms to collude and sustain high prices is surprisingly scarce. We aim to fill this gap and investigate algorithmic pricing on *Bol.com*, the largest online marketplace in Belgium and the Netherlands. Based on two months of high-frequency pricing data for more than 2,800 popular products, we explore the potential of algorithmic retailers to successfully increase prices.

Our analysis of the Dutch market leader e-commerce platform *Bol.com* is likely to be relevant for other marketplaces as well. *Bol.com* is very similar to Amazon in format, functions, products and the availability of third-party re-pricer software. These marketplaces are particularly interesting environments to explore the effects of algorithmic pricing: They are consumer-facing, very supportive for dynamic pricing and are surrounded by a wide and active ecosystem of algorithmic pricing software providers, who often make little effort to hide their intent to raise prices and avoid competition (Section 2.1).

Our main aim is to contribute to our understanding of the effect of pricing algorithms on pricing patterns and price levels in online marketplaces with a focus on the role of market structure. While we uncover several pricing patterns that are consistent with collusion, overall we find no robust evidence for algorithmic pricing being associated with higher average Buy Box prices. This is so even in for products with few competing sellers and multiple algorithmic agents.

☆ We gratefully acknowledge financial support from the NET Institute, www.NETinst.org.
* Corresponding author.
E-mail addresses: philip.hanspach@eui.eu (P. Hanspach), geza.sapi@ec.europa.eu (G. Sapi), mwieting@crai.com (M. Wieting).
[1] Funding by the German Academic Exchange Service is gratefully acknowledged.
[2] The views expressed in this article are solely those of the author and may not, under any circumstances, be regarded as representing an official position of the organizations the author is affiliated with.
[3] The views expressed in this article are the views of the author only and may not be regarded as representing the views of Charles River Associates, its academic affiliates, its employees, or its clients.

https://doi.org/10.1016/j.infoecopol.2024.101111
Received 5 July 2023; Received in revised form 20 August 2024; Accepted 8 September 2024
Available online 18 September 2024
0167-6245/© 2024 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

We also qualitatively explore our dataset and structure the price data into several recurring patterns. Some of these patterns are consistent with algorithmic sellers tacitly colluding. Our aim is not to *prove* collusion. Instead, we aim to distill price patterns that can help competition authorities, firms, and researchers scan the horizon for potentially anti-competitive practices. A significant practical advantage of our approach is that it relies purely on publicly accessible data.

Thoroughly described algorithmic pricing patterns, such as those presented in this paper, are valuable for competition authorities, who may – once candidates are narrowed down – have powers to investigate the behavior further, with more apt methods than economic analysis.[4] For competition authorities, two key questions arise *before* considering an investigation: 1) Is there any potentially suspicious pricing behavior to be investigated (and what kind)? 2) Which sellers should the agency focus on in an investigation on a platform with an ocean of products? We provide detailed guidance on these questions.

Interestingly, we document a large price *reduction* of more than 20% due to algorithmic agents in monopoly products, compared to similar products sold by traditional sellers. This is a novel phenomenon that we explain by the improved ability of pricing algorithms to experiment and adjust prices separately for thousands of products, a task that is prohibitive for humans.

The paper is structured as follows. In Section 2, we present the online shopping platform *Bol.com* and the features of re-pricing software. Section 3 reviews the related literature on algorithmic pricing and collusion in off- and online markets. In Section 4.1 we describe the dataset and the underlying cloud scraping procedure. Section 4.2 provides descriptive statistics and Section 4.3 a graphical analysis of the main algorithmic pricing patterns. In Section 5 we conduct econometric analysis of algorithmic pricing. Section 6 discusses the policy implications of our findings.

2. Background

We start with an introduction of the marketplace platform *Bol.com*, its Buy Box (in Dutch, the *koopblok*), the sellers active on the platform, and algorithmic re-pricer services.

2.1. Bol.com and third-party sellers

Bol.com is the largest online store in the Netherlands offering products in categories such as books, music, computers, toys, baby, cosmetics, clothing, and DIY.[5] Bol's revenues in the Netherlands exceeded 1.6 billion euros in 2018, amounting to about five times the revenue Amazon achieved in the country (Statista, 2019). Since 2011, *Bol.com* admits third-party sellers and is itself acting both as seller as well as platform operator. In 2018 *Bol.com* hosted more than 20,000 third-party retailers who accounted for about 40% of the company's sales (EcommerceNews, 2018).[6]

The platform *Bol.com* charges third-party sellers a fixed fee per article sold as well as a percentage commission of the sales price. *Bol.com* is surprisingly opaque about the precise amounts, which seem to vary by product type and sellers must upload the article list to find out the exact fees payable per item (Bol.com, 2021a). Bol's commission is typically a fixed 12.4% on the net sales price excluding VAT, plus a fixed fee depending on product price. Some narrow categories have a different percentage commission, e.g. lamps (in Dutch: lichtbron) 5%, or E-bikes 9% but there seems to be no analog with, for example, Amazon's FBA program that charges separate fees depending on the use of ancillary platform services.[7]

2.2. The Buy Box

Similarly to other online marketplaces including Amazon, the product page on *Bol.com* contains a *Buy Box*: the promoted seller chosen automatically by the marketplace operator. The *Buy Box* seller is displayed very prominently filling the bulk of the product page (Figure 2). For a seller on a marketplace platform such as Bol or Amazon, winning the *Buy Box* is an important achievement, as it typically generates around 80–90% of sales (RepricerExpress, 2021).

This carries significant relevance from a competition perspective: One may think about the Buy Box as a *"winner-takes-it-all"*-feature which assigns all demand to whoever offers the lowest price. As the price is likely the key factor in the assignment algorithm to the Buy Box and revenue streams are largely dependent on owning the Buy Box, competition for the Buy Box mostly reduces to price competition (Musolff, 2021). This sets the stage for further analysis, in which we try to assess whether the presence of reprice engines is associated with higher prices.

While the ultimate algorithm to determine the Buy Box winner is secret, Bol lists some factors it takes into account for its choice (Bol.com, 2021b). These include primarily the product and shipping prices, delivery time, availability of the item in stock, and the *seller performance score*. The latter is a mix of seller rating and other key performance indicators, such as on-time delivery, telephone accessibility, completeness of product description, track-and-trace information, and seller cancellations (ChannelEngine, 2018).

2.3. Re-pricer software

The term *algorithmic pricing* is used interchangeably in different contexts. The literature distinguishes several types of pricing algorithms.[8] The simplest range from pure *pricing rules* that allow retailers to set different prices depending on various conditions. Software solutions are available to monitor these market conditions and turn them into inputs that the pricing algorithm can comprehend.

More complex algorithms may offer a higher level of autonomy to the re-pricer engine in setting prices: *adaptive learning* algorithms calculate the optimal price based on a set of input variables such as costs, inventories, or rivals' prices. These algorithms are *adaptive* because they autonomously experiment with prices, learn, and adapt to find the optimal values. Algorithms may offer *static* optimization, that does not consider long-term consequences of actions, such as often associated with retaliatory, collusive outcomes. *Dynamic optimization algorithms* provide the largest degree of autonomy, and take into account longer-term consequences of actions, such as retaliating and maintaining collusive outcomes.

On *Bol.com*, third-party sellers can programmatically manage their inventory and adjust the parameters of their offer (sales price, shipping fee, delivery time, availability) via Bol's APIs. Sellers on *Bol.com* can manually manage their prices, but it seems fair to say that *Bol.com* is designed to facilitate dynamic pricing. As manual pricing becomes complex with a larger inventory, sellers are often aided by external repricing software that combines inventory management with algorithmic pricing features.

Re-pricer services such as ChannelEngine, EffectConnect, Channable, Vleks, Pricesearch.io and RepricerXL integrate with the *Bol.com* retailer API and automate the pricing process. They are able to update the price of a large number of items in near real-time and allow the seller to provide more or less guidance on setting these prices.[9] For example,

[4] In the mildest form, they can send information requests to firms, or upon sufficiently strong suspicion, organize inspections.

[5] *Bol.com* is also popular in the Dutch-speaking part of Belgium. We focus on the *Bol.com* platform as accessed from the Netherlands.

[6] Own sales constitute the bulk of the remaining 60%.

[7] See also Verkopen via bol.com [in Dutch].

[8] Klein (2021b) provides a detailed discussion of this topic.

[9] It seems that most re-pricer services – on Bol and Amazon likewise – update the prices every 20 minutes. See for example Chen et al. (2016) for Amazon and EffectConnect (2021) for Bol.

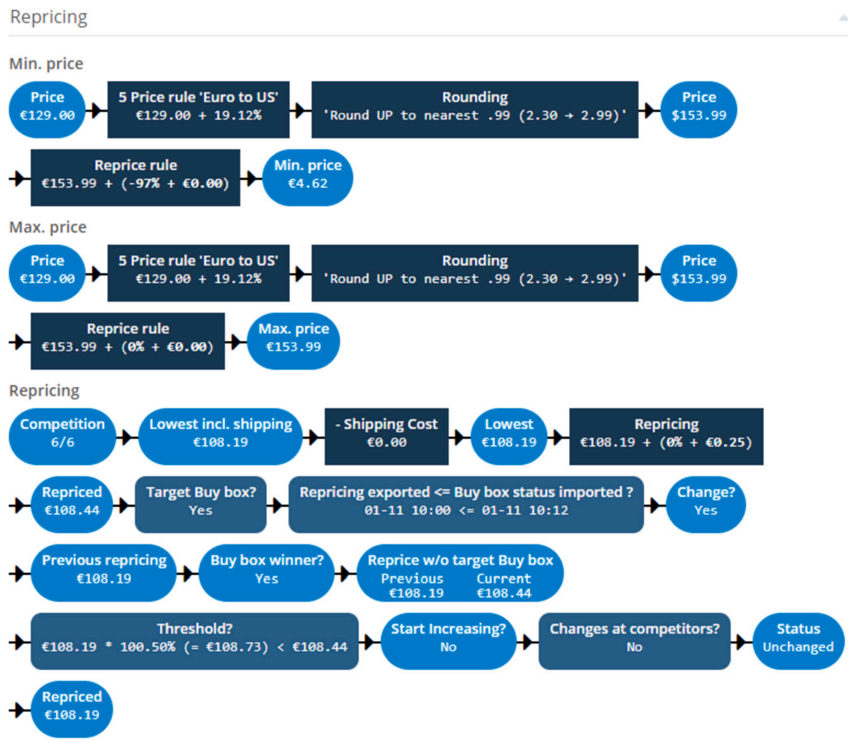


Fig. 1. ChannelEngine repricing rule illustration. Reproduced from https://help.channelengine.com/article/47-repricer, retrieved on the 17th of September 2021.

repricer.nl explains that the seller can set minimum and target prices, and choose which competitors to follow or ignore, but she can also leave the re-pricer freedom to adjust the prices (RepricerExpress, 2021).

ChannelEngine (2021) offers a detailed look under the hood of a re-pricer software.[10] It permits pricing rules based on, among other criteria, minimum, maximum, cost-plus and rival-plus type pricing. The seller can define scenarios and various triggers of new price rules. It also allows choosing the reference competitors, such as merchants using fulfillment services, Buy Box winners, sellers with a certain rating or manually picked rivals. An illustration of price rules can be seen in Figure 1. A sequence of "if-then" statements illustrate the simple way in which a re-pricer software takes account of rivals' prices and conditions such as having won the Buy Box.

In this paper, we find that algorithmic pricing in a competitive environment is not correlated with higher prices vis-à-vis consumers. While we do not claim to identify a causal effect of algorithms with prices (let alone *collusive* prices), this seems to be at odds with the statements made by re-pricer software vendors, who explicitly advertise their ability to raise prices and avoid competition, even using the economic textbook language of collusion.

Repricers are not shy about advertising their intent to raise prices. For example, SellerSnap – an Amazon re-pricer – warns its clients: *"Don't Be a Prisoner in Amazon Price Wars"*, explaining that *"your goal should be to get the Buy Box share you are entitled to while keeping the price high instead of racing to the bottom"* (IndustryNews, 2018). EffectConnect, a leading re-pricer for *Bol.com* recognizes its goal to raise prices: *"when your competitor increases the price, your price will go up along with that of your competitor"* (EffectConnect, 2021). Other re-price engines provide features to engage in price re-setting once a chosen minimum price has been reached in order to break a downward price correction spell (Musolff, 2021). Channable, another leading *Bol.com* re-pricer offers an entire menu block for *"Do not compete with"*, where the seller can configure the re-pricer to avoid price competition with rivals selected based on various criteria (Channable, 2019).

Despite the high level of automation these re-pricer algorithms allow, there are humans behind them. Autoridade da Concorrência (2019) reports exchanges in marketplace forums, where sellers using re-pricer software discuss competition, and make statements such as: *"The race to the bottom is a race that EVERYONE loses. STOP REPRICING YOUR STUFF INTO OBLIVION"*, *"MATCH the lowest person's price rather than attempting to undercut them. Undercutting is a win for no one other than the buyer"*, *"I see u* [sic!] *have a re-pricer on that undercuts the lowest FBA offer.* [...] *The result is a loss of profitability for everyone. Now your price is your choice and this message is in no way an attempt to fix pricing. You set your price to whatever you like but I just wanted to send you a message on what I observed on the listings you are on and share my thoughts with you."*[11]

Finally, the marketplace operator *Bol.com* openly declares its doubts about the lawfulness of some of its own policies. In particular, it notes about its price-transparency policy that *"providing this information might lead to price increases, possibly interfering with Dutch and Belgian competition law"* (Bol.com, 2021c): it is to be expected that such information is most valuable to algorithmic sellers, who frequently revise prices.[12] Overall, we believe Bol is an exciting environment to study pricing algorithms, with re-pricer software vendors openly advertising their ability to raise prices and the platform operator venturing into what it *itself* considers as the grey zone of the law.

3. Related literature

Our paper is closely related to the literature on the intersection between algorithmic pricing and collusion. Legal scholars and policymak-

[10] ChannelEngine integrates with, among other e-commerce platforms, *Bol.com*.

[11] To be clear: we do not suggest or claim any of these slogans prove collusion. However, they certainly do a good job of catching the attention of a very diverse audience interested in algorithmic pricing, including sellers, researchers, and policymakers.

[12] We suspect Bol may be concerned that its announced maximum prices may serve as focal points for collusion (Knittel and Stango, 2003).

ers recently expressed significant concerns about the potential of algorithmic pricing to facilitate collusive behavior. Capobianco and Gonzaga (2020), Ezrachi and Stucke (2016a,b, 2017), Mehra (2015) and Harrington (2018) discuss the competition policy implications of the question.[13]

In general, economic theory predicts three main avenues by which algorithms *may* facilitate collusion. First, increased transparency: automated, large-scale monitoring of rivals' actions may enable the quick detection of deviations from a collusive agreement. Transparency may therefore help sustain (tacit) collusion (Albæk et al., 1997; Albano et al., 2006).

Second, dynamic pricing increases the frequency of interaction: quick reaction to a deviation from collusive prices reduces the deviating firm's profit and therefore stabilizes collusion (Bigoni et al., 2019; Kühn and Tadelis, 2017; Brown and MacKay, 2020).

Third, rival firms delegating business decisions to common agencies, such as advertising bureaus or pricing software vendors, who act as the *hub* to facilitate coordination among the *spokes* in a *hub-and-spoke* scheme (Bernheim and Whinston, 1985; Decarolis and Rovigatti, 2019).

The theoretical literature linking pricing algorithms to collusion remains ambiguous about the ability of programmatic agents to collude. Calvano et al. (2020), Eschenbaum et al. (2022), Johnson et al. (2020) and Klein (2021a) show based on simulations in a repeated game framework that under certain conditions Q-Learning algorithms can converge to the collusive outcome, sustain supra-competitive prices and punish deviations. Other authors emphasize that the improved ability of algorithms to better predict demand and react to stochastic shocks destabilizes collusion as deviation becomes more profitable (Miklós-Thal and Tucker, 2019; O'Connor and Wilson, 2020). Asker et al. (2021) show how supracompetitive price outcomes can occur in a simple Bertrand game with reinforcement learning algorithms. Notably, the set of algorithms considered do not care about the future and only value current profits, essentially ruling out collusive equilibria in advance.

A handful of papers investigate how human and algorithmic agents interact in markets. Leisten (2021) develops a model of competition in which managers may override an automated pricing rule after the rule is chosen. Prices remain higher than a competitive benchmark, but collusion breaks down when managers must respond to a common demand or cost shock. Under such conditions, both the prediction-enhancing and the commitment-enhancing features of algorithms may serve to sustain supra-competitive prices.

In the same strand, Normann et al. (2021) compare tacit collusion incentives in a laboratory setting when only humans interact to the case of one firm in the market delegating its decisions to an algorithm. The authors find that in three-firm markets the presence of one algorithmic player makes collusion more likely, but this effect wears off with four firms competing. Somewhat surprisingly, the algorithmic player earns lower profits than the rivals. Werner (2021) provides experimental evidence in a setting of human sellers relying on algorithms. Since most firms rely on algorithms, oligopoly markets seem to be especially prone to collusion for three-firm markets.

While most commentators appear to be wary of algorithms eventually facilitating collusion, there are also critical views in policy circles, arguing that the idea of algorithms forming cartels may be speculative: the argument is that mindless algorithms can never achieve a *"meeting of minds"*, which is the legal standard for collusion (Colombo, 2018).

In a similar vein, some economists emphasize the inability of algorithms to sustain collusion without explicit communication. Referring mainly to experimental economics literature, Kühn and Tadelis (2017) and Schwalbe (2018) argue that, much like humans, self-learning algorithms would do poorly in coordinating actions to achieve a desirable collusive outcome, at least in the absence of explicit communication.

Empirical literature analyzing algorithmic pricing in real markets is scarce and we aim to contribute to this strand. In a recent paper, Assad et al. (2020) study Germany's retail gasoline market where algorithmic-pricing software became widely available around mid-2017. The authors find that the adoption of algorithmic pricing software increases margins significantly, especially in duopoly markets where both rivals move to algorithmic pricing. The magnitude of price increase is consistent with other recent papers studying collusion in retail gasoline markets (Clark and Houde, 2013, 2014; Byrne and De Roos, 2019). In a similar context, Luco (2019) provides empirical evidence on the presence of anti-competitive effects of the introduction of a price disclosure policy in the gasoline market – in particular where consumers' search intensity is low. Our main result is different – lower prices in single-seller markets when the seller is algorithmic. However, there are several explanations for this related to the differences between the market context of sellers on online marketplaces and gas station managers, notably the large product portfolios of the largest online sellers. We discuss this in greater detail following the results in Section 5.1.

Our article is closely related to research on algorithmic pricing and competitive strategies in electronic marketplaces. Chen et al. (2016) studies the behavior of algorithmic sellers on Amazon. The authors develop a methodology for identifying algorithmic sellers and find that compared to non-algorithmic competitors, these win the Buy Box more often, are active in the marketplace for significantly longer, (surprisingly) tend to specialize on *fewer* products, and acquire a larger number of positive feedback, suggesting that they also sell more. Our paper extends this line of research by focusing on the interaction of market structure and prices, allowing us to draw conclusions about the effect of algorithms on prices in different market settings. We draw inspiration from Chen et al. (2016) to identify algorithmic sellers and extend this research by focusing on screening for collusion and the ability of algorithms to sustain higher prices.

Zhu and Liu (2018) analyze the patterns of Amazon's entries into its third-party sellers' product spaces. The authors find that Amazon is more likely to enter as the seller for more popular products with higher seller ratings. Jiang et al. (2011) provide descriptive evidence for Amazon specializing in high-demand products and leaving the sale of a long-tail of products for third parties. Musolff (2021) provides causal evidence on the effect of algorithmic sellers on price competition exploiting data directly from re-pricers and Amazon. He finds that the presence of algorithmic sellers initially decreases prices but introduces re-setting strategies similar to Maskin-Tirole's Edgeworth cycles which aim at *avoiding* fierce price competition à la Bertrand.

Results differ according to the characteristics of products and market structure. In the market for over-the-counter medicine, Brown and MacKay (2020) show how asymmetries in pricing technology may translate into asymmetries in prices. Firms with more frequent price changes are associated with lower prices. If price setting frequency can be chosen, between-seller asymmetry in price setting frequency is the supra-competitive equilibrium associated with overall higher profits. Even price strategies that *do not* appear collusive at face value such as commitments on linear functions of rivals' prices (without reward-punishment schemes) may lead to such supra-competitive outcomes. In online grocery retail, Aparicio et al. (2023) find strategies characterized by firms matching their competitors prices, as well as by frequent, tiny price adjustments. Our own exercise of describing observed pricing patterns (see Section 4.3) finds examples both of uncorrelated price settings, possibly explained by random experimentation, as well as tiny, frequent updates.

To our knowledge, our paper is the first to empirically investigate the propensity of algorithms to raise prices on a popular *European* B2C e-commerce platform. Earlier research on pricing algorithms focused predominantly on petrol markets as well as on Amazon. We are first to consider how dynamic pricing may facilitate collusion and allow sustaining elevated prices on a local giant, European online marketplace.

[13] Despite the attention from competition authorities (Konkurransetilsynet.no, 2021; Autoridade da Concorrência, 2019; GOV.UK, 2021), to our knowledge so far no agency led a case involving autonomous algorithmic collusion. See Ritter (2017) for a list of antitrust cases where algorithms played a significant role in some form.

Table 1
Summary statistics.

Product-seller-level	Mean	Standard deviation	Minimum	Median	Maximum
Buy Box price in EUR	52.45	191.47	0.00	20.99	6199
Price in EUR	47.64	88.39	0.49	24.95	5499.00
Seller rating	8.8	0.45	1.0	8.9	10.0
Seller delivery time in days	2.86	3.22	1	1	136.5
Seller shipping fee in EUR	0.2	0.71	0	0	3
Product-level					
Total Buy Box price in EUR	59.01	215.60	0.00	22.44	6199
Seller rating	8.8	0.46	3.3	8.9	10.0
Seller delivery time in days	3.38	3.84	1	1	59.5

Buy Box price in EUR: Price of the Buy Box product excluding delivery cost.
Total Buy Box price in EUR: Price of the Buy Box product including delivery cost.
Seller rating: Numerical rating (0–10). **Seller delivery time**: Delivery time for the product (in days).
Seller shipping fee: Shipping fee (in Euro).

Fig. 2. The *product* page.

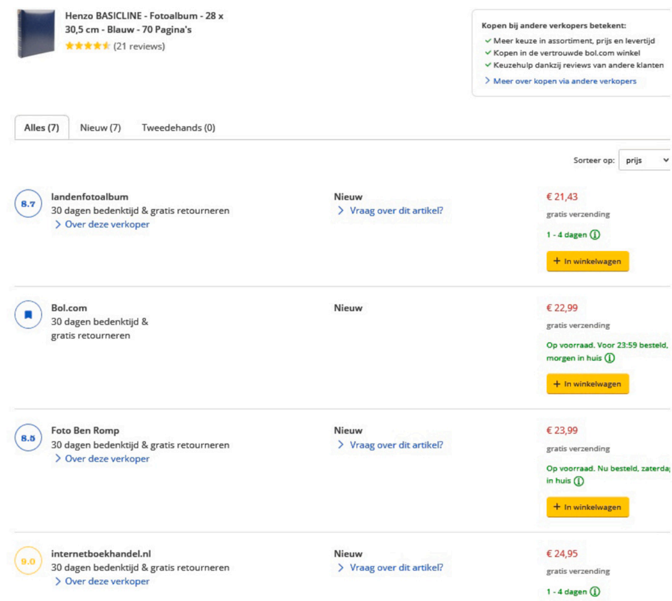


Fig. 3. The *compare all sellers* page.

We are also the first to document potential efficiencies associated with pricing algorithms.

4. Descriptive analysis

4.1. Data

The data used in this article was obtained by scraping the *Bol.com* website in two rounds. The main crawl was conducted between the 26th of December 2018 and the 25th of January 2019.[14] Starting after Christmas in December 2018, we scraped the top 500 pages of bestselling products on *Bol.com* once. Each page contains 24 products, yielding a total list of 12,000 products. From this list, we eliminated products that were not available for sale (*out of stock*) between the 28th and 30th of December 2018. The final sample covers 2,846 products that were available for sale in a stable manner over three consecutive days. For each of these products we scraped the *product* page (Figure 2) and the *compare all sellers* page (Figure 3).

The *product* page prominently shows the Buy Box price and featured seller next to an image of the product. In this example, the Buy Box price is 22.99 and Bol.com is the featured seller (*“verkoop door bol.com”*). It is further indicated that the same product is also offered by six other sellers.

The *compare all sellers* page can be reached by clicking on a link from the *product page*. It is a paginated list of all sellers offering the product at a given time. We focus on the first page of the list with the top ten sellers of new items and exclude second-hand offers. We extract the seller rating (0–10), all prices, shipping costs, and expected delivery times.

Table 1 presents the main descriptive statistics at the product-seller-level, where each seller’s offer for each listing, every time it is crawled, is a separate observation and is used for the graphical analyses, as well as the product-level, where these observations have been summarized so that each listing, every time it is crawled, represents one observation and which is used for the regression analyses.

We used a cloud-based web-scraping service to conduct the crawls and increased the computing resources during a second crawl discussed in the robustness section. The URLs scraped are identical between the two crawls with the exemption of products that disappeared from the platform by the second crawl.

Seller rating is the only variable where some data interpolation was necessary. Sellers in our sample can be unrated because they are “*new*” or “*unproven*,” due to possible crawling errors, and also because Bol does

[14] As a robustness check, we did a second crawl between the 18th of February 2020 and the 20th of April 2020 and covered the same list of products as the first crawl. The second crawl skipped the first week of April 2020 due to a subscription issue with our cloud-based scraping service provider. We discuss the second crawl below in the robustness section.

not seem to display the rating of sellers in the Buy Box at all times (for example – but not only – when the seller is Bol).[15]

In particular, for unrated sellers we use the average seller rating in the detailed (third-level) product category.[16] We decided on this approach by asking ourselves what a buyer would most likely assume about the *quality* of a seller with no rating. We believe assuming an average quality is realistic, and this quality may differ by product category such as Health or Toys.[17]

In our remaining analysis, we will assess competitive conditions by product. Doing so risks ignoring potential substitution patterns among closely related products. We therefore perform robustness checks in Section 5.2, identifying closely related products based on the similarity of product descriptions and the most detailed product categorization available. We find that the results change little overall.[18]

4.2. Identifying algorithmic sellers

We do not directly observe which sellers rely on automatic pricing tools.[19] We therefore apply *heuristics* that allows us to identify algorithmic sellers with a high probability based on their observed behavior.

To do so, we draw inspiration from Chen et al. (2016) to define sellers as algorithmic. Our thinking was guided by the following observations for choosing criteria to label sellers as algorithmic:

- *Algorithmic sellers change their prices often*: we can define a seller as algorithmic if it performed a certain number of price changes within a given time period (e.g. a crawl, a week, or a month).
- *Algorithmic sellers' prices correlate with other benchmarks* (e.g. lowest price, second lowest price, *Bol.com*'s price, any competitor's price): we can define a seller as algorithmic if its prices show a sufficiently strong correlation with one or more of such benchmarks.

We conducted extensive analysis with different individual criteria and combinations of criteria to identify algorithmic sellers. We concluded that the *total number of price changes* over a crawl is the most reliable approach to select algorithmic sellers, and is superior to other criteria for the following reasons:

A high number of price changes by a seller for a product is a reasonable indicator of the seller using pricing algorithms. Since our data very likely covers only a small sample of the seller's product range, a high number of price changes for a product likely implies many more price changes in the seller's full product portfolio, for which automated pricing tools are very likely needed.

For example, for some products we observe sellers with hundreds of price changes. We are confident that this seller uses an automated pricing engine, since, as we discuss below in more detail, she could not manually set such prices on possibly hundreds of her products outside our crawl.

Looking at the time elapsed between price changes could in theory be an indicator used to identify algorithmic sellers, but in practice, this measure is also inevitably affected by the crawl frequency: the latter is by nature somewhat uneven over the period of data collection, due to latency and the varying availability of cloud computing resources by time.

Correlations with other price series is another potential marker, and it was also the chosen approach in Chen et al. (2016). For our purposes, this heuristic has the drawback that we may fail to spot algorithmic sellers who do not adopt a price-correlation strategy. The potential error by such *false negatives* is rather large in our view: For example, as we will explain in Section 4.3, some sellers seem to randomly experiment with prices, that do not show any obvious correlation with other series. These sellers are algorithmic, but the price correlation criterion would not label them as such. A further drawback is that for one-seller products this measure cannot be defined. Monopoly markets constitute an interesting benchmark and deserve attention on their own, and we report novel findings for one-seller products.[20]

The single criterion with the relative number of price changes (2 standard deviations above product category mean, or "2 SD") is our most preferred screen to identify algorithmic sellers, for multiple reasons. First, this criterion strikes a reasonable balance between type-1 and 2 errors: Visually inspecting the prices of those product-seller pairs that are deemed algorithmic by this criterion confirms that the classification is correct, hence there is little risk of flagging sellers as algorithmic that should not be regarded as such.

Second, the relative number-of-price-changes criterion is the most general with respect to the pricing strategies algorithmic sellers may employ, since it does not require a particular correlation with other prices. It is also available for monopoly products (unlike price-correlations), a benchmark we wish to exploit in our analysis.

We explain below the detailed implications of various definitions of algorithmic sellers and provide robustness checks using criteria that take the number of changes as a baseline and combine these with various price correlations in Section 5.2.

4.2.1. Total number of price changes

We first filter algorithmic seller accounts by looking at the distribution of the total number of price changes across the crawl period. The left panel in Figure A.2 presents the CDF over the total number of price changes. It indicates that most product-seller pairs experience few price changes, if at all. For a small share of product-seller pairs, we observe hundreds of price changes in the long tail of the distribution. Frequent price changes are very likely the result of algorithmic price engines at work. We normalize the distribution of price changes per product and consider algorithmic pricing if it exceeds the mean by 2 (3) standard deviations in the overall normalized distribution shown in the right panel of Figure A.2.[21]

Figure 4 shows the distribution of algorithmic and non-algorithmic sellers over the number of sellers.[22] We observe a downward-sloping curve with the bulk of algorithmic sellers on products with few sellers. For instance, for duopolies, in over 20% of the observations one seller

[15] Sellers relatively new to the market at the time of our crawl may not have yet had the time to obtain ratings. Bol's policy to display ratings is not fully transparent. Some sellers appear to be already several months on the marketplace with hundreds of products, yet Bol may display no rating.

[16] This is the lowest level of category aggregation in our data, including for example "*watches*", "*cook books*", and "*SSDs*." In case there are insufficient ratings in a sub-category, we take the average rating of the higher-level category.

[17] We have conducted a number of robustness checks on the rating variable and found it had a minor impact on the coefficients. In particular, we repeated the main analysis with three different options, with our preferred option being that of imputing missing values with the category average of the same seller (option 1), using the minimum rating of a seller within the category (option 2), and the referee's proposed solution of using a dummy variable for missing ratings (option 3), to consider sellers with missing rating as a category onto itself. The results are consistent, details are available upon request from the authors. Berentsen et al. (2019) provide a recent review of the economic literature on unrated sellers in platform markets.

[18] This finding carries important practical relevance: Identifying related products that may be potential substitutes is tedious, a challenging task. It relates to the question of recommending related products, on which major e-commerce platforms employ several data engineers.

[19] This is a practical problem a competition authority or firm would typically face in the horizon-scanning phase for potentially anti-competitive conduct. In that stage, public data is particularly valuable as sending information requests and organizing unannounced inspections are often premature.

[20] We do observe sellers that operate alone and set prices automatically.

[21] The number of affected sellers by applied criterion can be found in Table 5 in Section 5.2.

[22] For this figure we rely on the number of price changes as the underlying criterion for algorithmic pricing at the seller level.

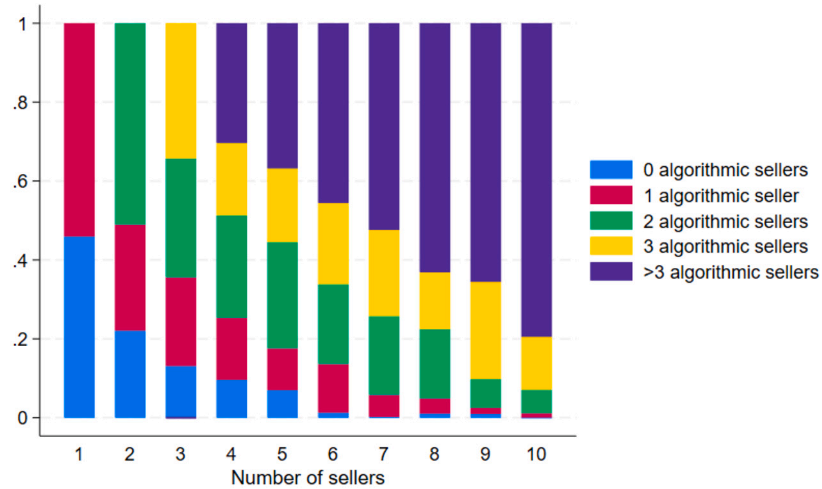


Fig. 4. Share of algorithmic sellers on all sellers.

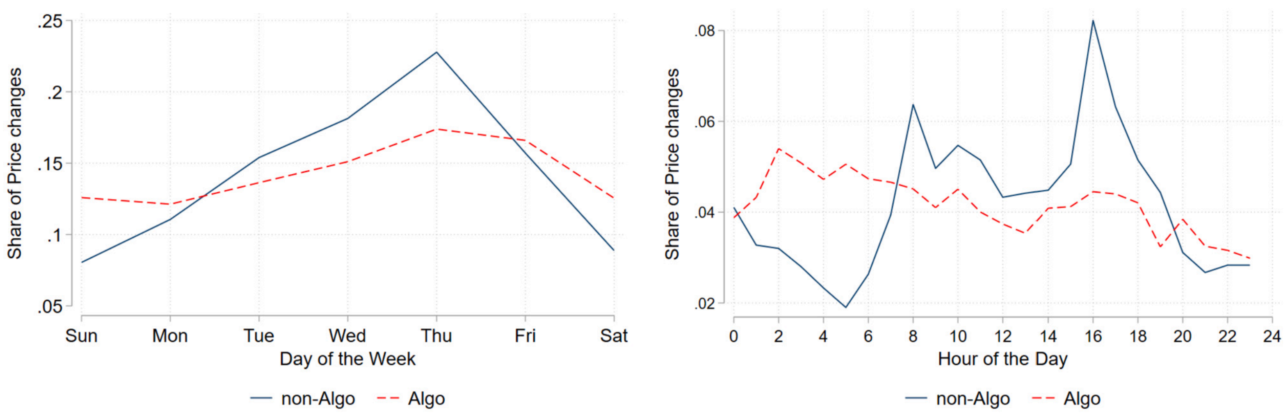


Fig. 5. Share of Price Changes by Weekday and Hour of the Day (Central European Time) in all for Algorithmic versus Non-Algorithmic Price Changes. Product-Seller-pairs are flagged algorithmic if we document more than 20 price changes.

is algorithmic. In about 50% of the observations, both sellers are algorithmic.

A sanity check based on a frequency criterion is the time when we observe price changes. Figure 5 shows that the human price changes are most common during weekdays and during working hours for Europeans, who we suppose to be the majority of sellers on Bol.com. Price changes that we attribute to algorithmic sellers based on frequent price changes tend to be spread out more evenly throughout the week and the hours of the day.

4.2.2. *Price correlations*

Sellers using automated re-pricers may set prices relative to others on the platform. As explained in Section 2.3, re-pricer tools offer the functionality to peg prices to rivals, the Buy Box price, and other references. Following Chen et al. (2016) we use the price of any competitor $k \neq j$ on the same product i. We calculate Spearman's rank correlation coefficients for all sellers $j \neq k$ for product i, ρ_{jik}.

For each seller-product pair, we use the maximum entry jk of the correlation matrix as our measure of algorithmic pricing. We refine our criterion on the normalized total number of price changes with a cut-off of .7 in terms of the price correlation with a competitor on the same product.[23]

In the remainder of this article, we present the main results obtained by flagging sellers as algorithmic using the number of total price changes heuristic.[24] In Section 5.2 we provide robustness checks combined with other heuristics to identify algorithmic sellers.

Table 2 provides summary statistics on the relevant indicators for non-algorithmic and algorithmic sellers. To make the comparison more informative, prices are residualized against product-fixed-effects. Algorithmic sellers set slightly lower prices and Buy Box prices. Furthermore, algorithmic sellers *win* the Buy Box more often than traditional sellers. Lastly, we see that algorithmic sellers manage substantially more products than non-algorithmic sellers.

Table 2
Buy Box price by subsamples.

	Algo = 0		Algo = 1		Difference
	mean	sd	mean	sd	p-value
BuyBox Price	0.03	12.91	-0.03	6.60	(0.00)***
Price	0.39	6.86	-0.38	6.59	(0.00)***
BuyBox share	0.18	0.39	0.40	0.49	(0.00)***
Rating	8.77	0.52	8.80	0.31	(0.00)***
Total Price Changes	1.16	6.32	7.46	27.85	(0.00)***
Nr. of Products per Seller	23.43	35.67	771.09	826.43	(0.00)***
Price change in % (stdd.)	0.03	0.31	-0.03	0.30	(0.00)***

(Buy Box) Price and Price changes are residualized by Product Fixed Effects.

[23] In Section 5.2 we perform robustness checks varying this cut-off.

[24] Precisely, two standard deviations above the normalized distribution of price changes.

Table 3
Frequency of algorithmic price patterns in sub-sample where at least one seller has more than 20 price changes.

Algo Pattern	Frequency (in %)
Jitter	52
Alternate	20
Feathers and Rockets	11
Random Jumps	11
Balloons and Rocks	6

4.3. Algorithmic pricing patterns

Which price patterns emerge in markets with algorithmic sellers? The question is relevant for researchers and policymakers alike. Understanding the resulting pricing patterns helps researchers link the observed data to theoretical models of dynamic pricing. Practitioners, such as competition authorities scanning the horizon for anti-competitive behavior need to be able to identify pricing patterns that may indicate collusion, and do so in a data-sparse manner. A contribution of our research is that we provide graphical illustration of pricing patterns that emerge in real data. We graphically analyze high-frequency price data to identify patterns that indicate algorithmic pricing (see also Connor, 2007; Zitzewitz, 2012).

To investigate the issue, we select those products where we previously identified algorithmic sellers to be present and plot the prices of all sellers as well as the Buy Box. First of all, and in line with the findings by Brown and MacKay (2020), we note large heterogeneity in price-setting behavior between sellers in terms of observed patterns and price-setting frequency.

We categorize five prominent and recurrent pricing patterns that are characteristic of algorithmic sellers. We do not apply quantitative criteria to distinguish these patterns and instead rely on our own intuition to classify them. The resulting price pattern categories may therefore even overlap in some cases. However, most of the time they are rather clearly distinguishable to the human observer. While our categories are not exhaustive, we are convinced that they cover the most persistent pricing behavior the human eye can detect in our data. We discuss the most interesting price patterns in more detail below.[25]

1. **Price jitter up** and **down**: There is a rapid, transitory increase (*jitter up*) or decrease (*jitter down*) in the price of the seller. The jitter is usually in place only for a very short period of time (Figure 6a).
2. **Rockets and feathers**: The price shoots up rapidly, and then gradually and slowly decreases, often reaching the starting point (Figure 6b).
3. **Balloons and rocks**: The price increases slowly and gradually up to a point, where it collapses and falls rapidly, often reaching the starting point (Figure 7a).
4. **Alternating price**: The price jumps up or down for a longer but transitory period between two prices, after which it returns close to the earlier level (Figure 7b).
5. **Random jumps**: The price changes frequently in a seemingly random manner (Figure 8).

The patterns *jitter up* and *down* are rather popular. We observe either of these in about half of the 300 products selected for inspection, with the jitter pointing up and down in about an equal number of cases (Table 3). While upward jitters are produced by a large number of sellers, interestingly, jitters pointing down are very typical to Bol. We see a downward jitter in 75 seller-product pairs, out of which 56 times the seller is Bol.

Price jitters are also documented by Chen et al. (2016) on Amazon, who conclude *"the very rapid price 'jitters' are likely caused by transient inconsistencies in Amazon's infrastructure, rather than actual price changes by sellers."*

We find it unconvincing that on *bol.com* these jitters would be caused by a malfunction, for many reasons. First, it would be unlikely that the same glitch slipped in both on Amazon and Bol, independently. Second, we observed some of the affected products over time and never encountered any inconsistencies.[26] Third, downward jitters are mostly (but not only) performed by Bol as the seller. It seems unlikely that different sellers would be affected differently by a platform-wide malfunction. Fourth, jitters persist in both of our crawls, with more than a year having elapsed between them. It seems unlikely that a technical error would not have been eliminated over this time. Fifth, we observe products where the jitters lead to an apparent reaction by other actors, such as a change in the Boy Box seller. This testifies that other players and the platform operator also perceived these rapid price changes. It therefore appears likely that the jitters on *Bol.com* are the result of actual pricing behavior.

A possible collusive explanation for the upward price jitters may be signaling to competitors the intent to raise prices. Byrne and De Roos (2019) document that petrol stations in Australia used price jumps *"to signal their intentions, and to create a mutual understanding of a coordinated pricing strategy among rivals."* A potentially collusive explanation of *downward* jitters may relate to signaling a firm's ability to reduce prices and punish deviating rivals if needed. We would expect the seller with the lowest marginal cost to be the most likely to engage in this kind of signaling since she is the most able to drive prices down. Consistently with this, as explained above, we observe downward jitters predominantly when the seller is Bol. It is beyond doubt that Bol is a fiercely competitive seller, one we would typically expect to have the lowest marginal costs.[27]

A further concern with downward jitters is that even absent collusion, these large transitory downward price jumps are also direct proof that the seller is pricing above cost, and does so most of the time.[28]

The pattern *rockets and feathers* is observed in about 11% of the 300 products considered. Motivated mainly by Edgeworth cycles, this pattern is very often associated with collusion.[29] In their seminal work, Maskin and Tirole (1988) provide a dynamic competitive model that gives rise to an equilibrium with Edgeworth cycles, and add that their *"model can be viewed as a theory of tacit collusion."* While the *rockets and feathers* pricing pattern did trigger cartel investigations (e.g. Byrne, 2012), it is debated to what extent this phenomenon emerging in a *non-cooperative equilibrium* can and should be regarded as collusive.

Rockets and feathers is also the pricing pattern that emerges in simulation studies of collusion by Q-Learning algorithms (Calvano et al., 2020 and Klein, 2021a). It has also been shown to be characteristic to collusion in gasoline markets (Eckert, 2013 and Byrne and De Roos, 2019). Furthermore, most non-collusive explanations of *rockets and feathers* pricing we are aware of are based on consumer search and unexpected cost changes (Yang and Ye, 2008; Cabral and Fishman, 2012 and Tappata, 2009).[30]

[25] We observe very similar pricing patterns during a second crawl more than a year later, discussed in the robustness section and the appendix.

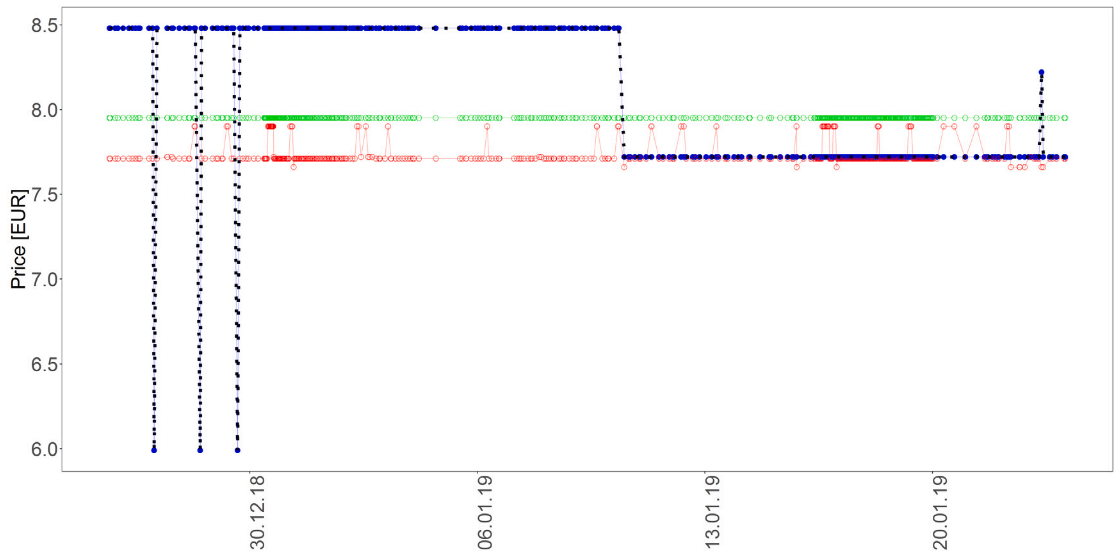
[26] For instance Chen et al. (2016) report the shopping basket not working during their crawl. We shopped for the crawled products several times while our crawler ran and never encountered problems with the basket.

[27] Our results in Section 5 show that Bol being present as seller has a very large effect on market outcomes, typically driving prices down.

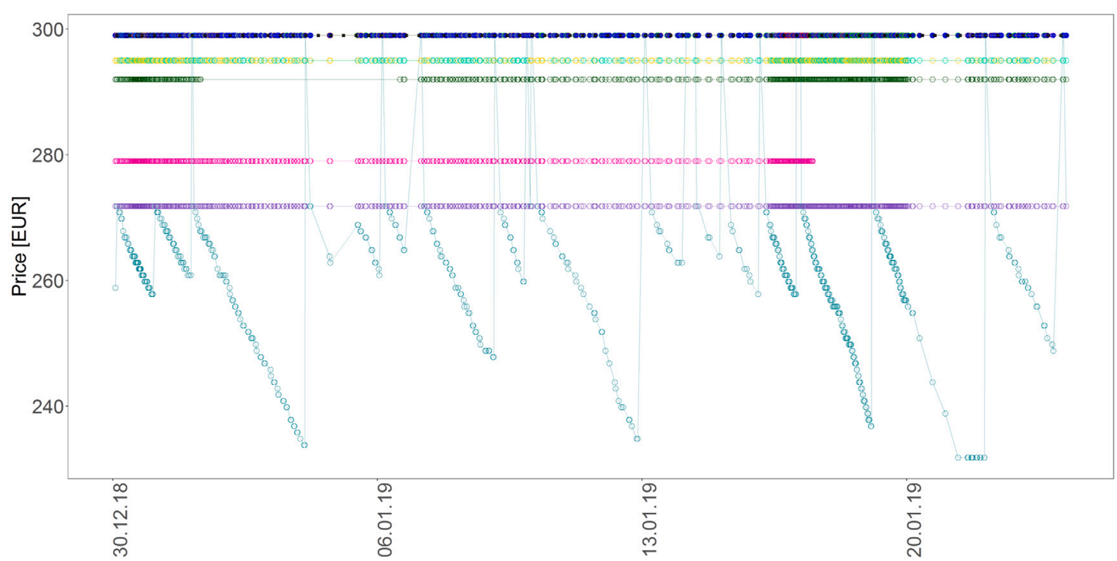
[28] This is because for rational firms even the bottom of the downward jitter is above costs, so the fact that these price reductions are very rapid implies that most of the time the firm prices relatively far above costs.

[29] Borenstein et al. (1997) argue that retailers may prefer not to reduce prices in response to negative cost shocks and prefer to use previous prices as focal points for coordination.

[30] For example, in Tappata (2009) marginal costs change over time which influences how consumers search. When marginal costs are high, consumers expect



(a) Jitter up and down



(b) Rockets and feathers

Fig. 6. Sample price patterns. Black dots denote the Buy Box seller.

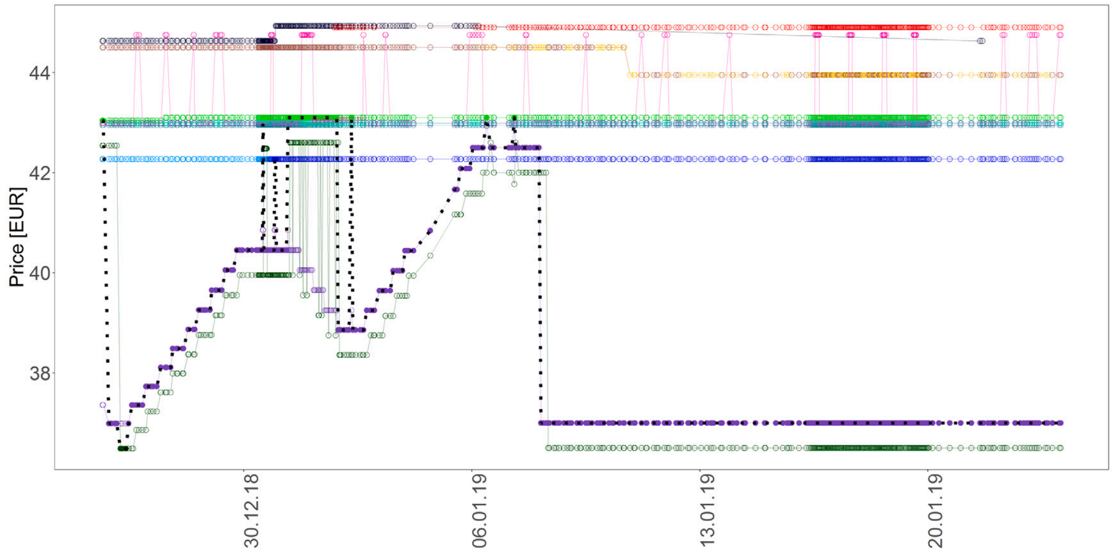
Finally, Figure 8 shows an example of a price series (the violet one towards the center) that exhibits very frequent jumps that are striking but follow no clear pattern (the red time series beneath also *jitters* downwards twice). These pricing patterns are different from the pricing patterns observed in non-algorithmic markets. Typical pricing patterns for the products that we identify as non-algorithmic are much more static and typically only change in irregular intervals less than once per day.

Examples of such non-algorithmic pricing patterns for 6 products at 3 different price points are shown in Figures 9 to 11. In Figure 9, a single seller adjusts a price upwards after being constant for over 3 weeks, then lowers it slightly below the previous level less than a week after the increase. In Figures 10 and 11 we also see prices mostly staying constant over several weeks with neither seller frequently switching prices or closely following competitors' prices.

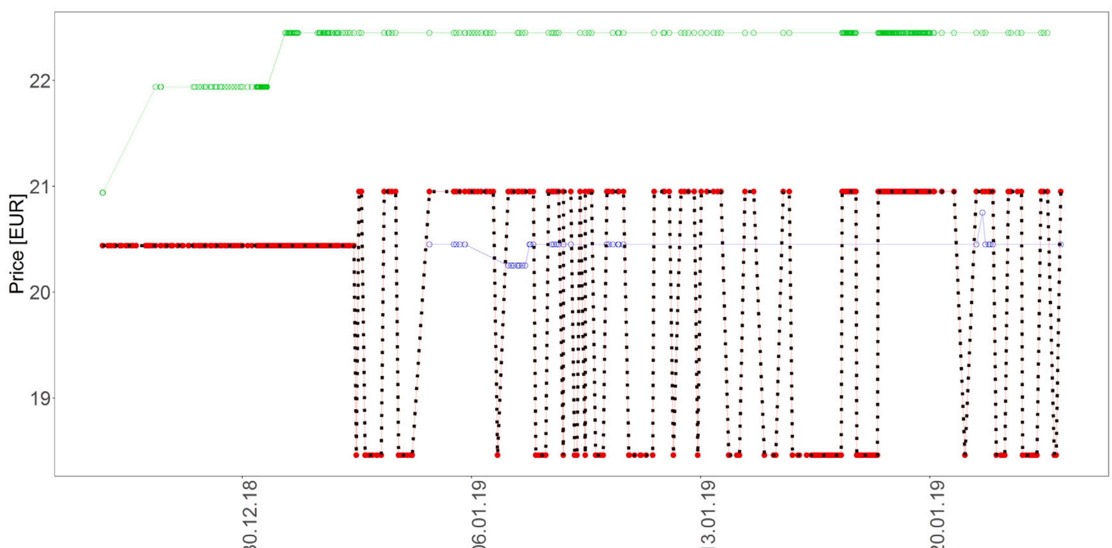
We believe these models do not apply well to an online marketplace such as *Bol.com*. We observe several price changes within a short period. These changes appear unlikely to be driven by unexpected cost shocks since marginal costs are hardly changing within a day, or even during the few weeks of our sample. We are not sure whether tacit collusion is the main reason behind the *Rockets and feathers pattern*.[31] Given the prevalence of the pattern and the fact that the price jumps tend to be significant, we however believe this pattern is a candidate for any screen for tacit collusion.

little price dispersion and search less. If marginal costs unexpectedly drop, firms have little incentive to lower prices because consumers are not searching much (*feathers*). On the other hand, if marginal costs are low consumers expect large price dispersion and intensify search: then firms' response to a positive cost shock is to raise prices significantly (*rockets*).

[31] It is certainly not the only reason, as we observe this pattern by single sellers as well.



(a) Balloons and rocks



(b) Alternating Prices

Fig. 7. Sample price patterns – continued. Black dots denote the Buy Box seller. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

We observe a *balloons and rocks* pricing pattern in 6% of the 300 products inspected in detail. We are only aware of this pricing pattern being previously reported in energy markets (Douglas, 2010 and Bremmer and Kesselring, 2016). To our knowledge it has never been described as conduct that would harm consumers, nor has it been discussed in the context of algorithmic pricing. This is somewhat surprising, first because it resembles the theoretical pattern of collusive phases, followed by undercutting and gradual forgiveness which resemble the algorithmic punishment-and-forgiveness strategy described by Calvano et al. (2020) and second because the *balloons and rocks* pricing pattern can be the outcome of very simple algorithmic rules. We explain this based on Figure 7a.

In Figure 7a two sellers display *balloons and rocks* pricing: *Dark Purple* and *Dark Green*. This can be the result of two banally simple "*if-then*" pricing rules meeting. The Dark Green seller always undercuts the Buy Box seller by a fixed amount. If Dark Green wins the Buy Box (it never does in Figure 7a), she leaves the price unchanged.

The Dark Purple seller is slightly more complex as it acts as the price leader, and is experimenting: under normal circumstances (when the environment is favorable) it always increases the price by a small fixed amount as long as it holds the Buy Box. Dark Purple reduces the price period by period if she is not the Buy Box seller. This explains the pricing of Dark Purple until the price drop (i.e. the *"rocks"* event) on the 8th of January 2019, but not the drop itself.[32]

[32] We investigated the price drop event on the 8th of January 2019 in detail. The most plausible explanation for *Dark Purple*'s large price reduction is an exogenous event: One day before the price drop the market saw a new entrant so that the number of competing firms increased from nine to ten. Perhaps more importantly, immediately before the price drop, the main competitor, *Dark Green*, increased its delivery time from five to six days. This led to an instant punishment in the ranking on the *compare all sellers* page, where *Dark Green* moved

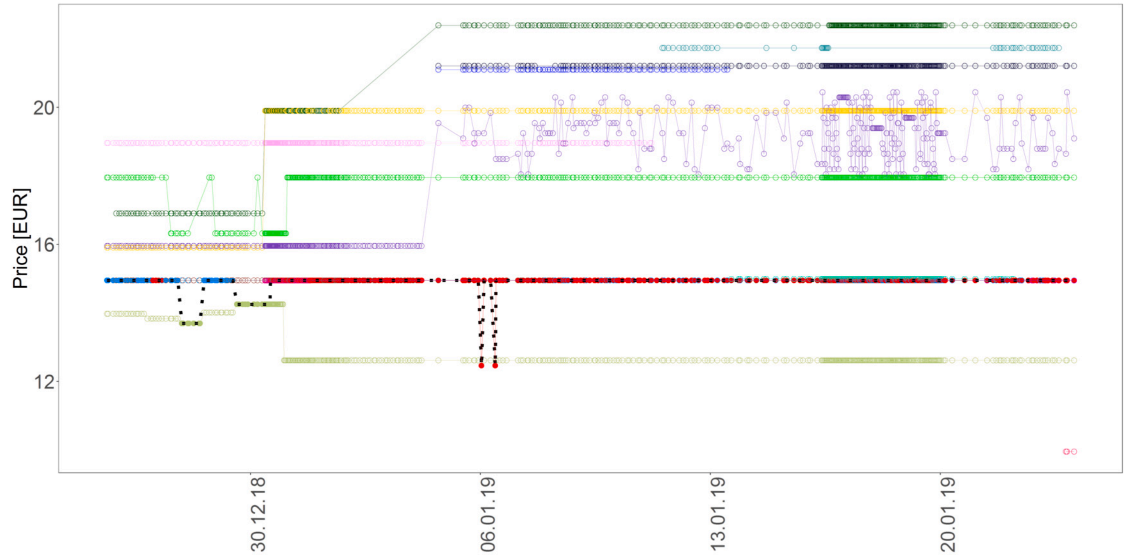


Fig. 8. Sample price pattern – Random Jumps.

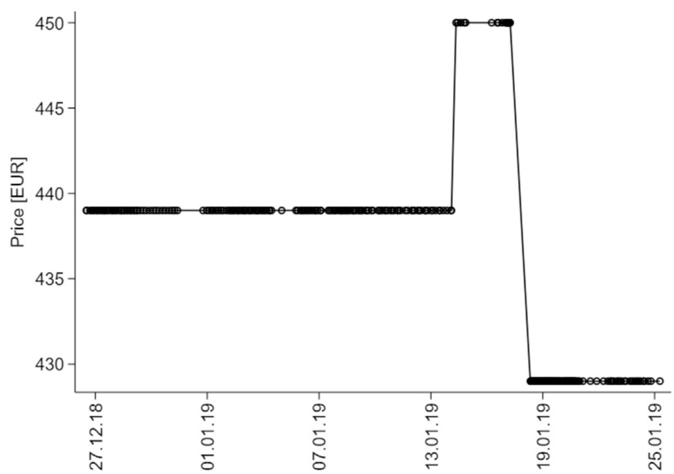


Fig. 9. Sample price pattern – non-algorithmic prices example 1.

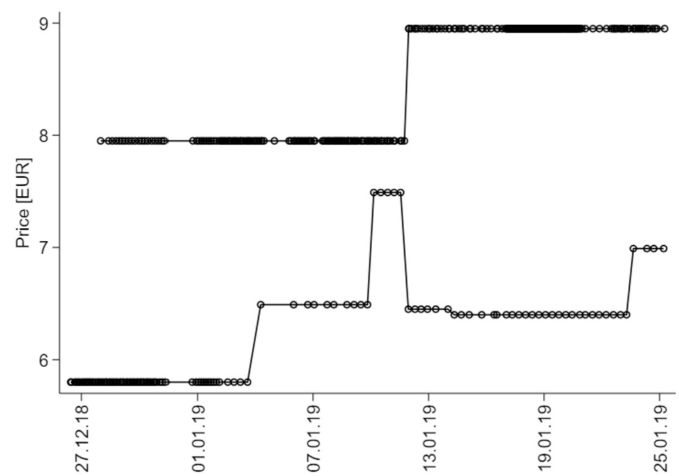


Fig. 11. Sample price pattern – non-algorithmic prices example 3.

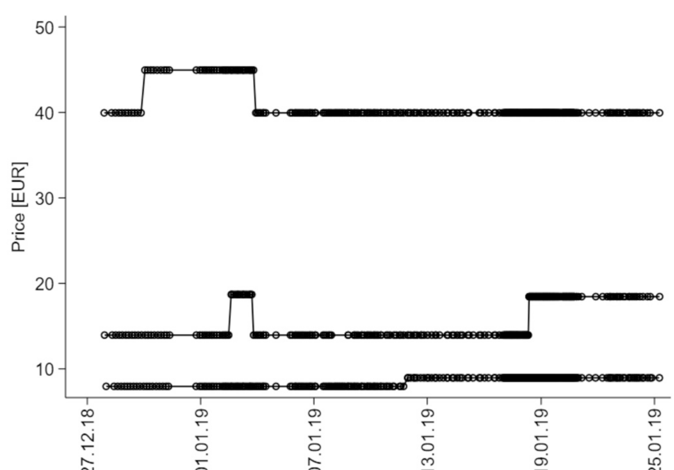


Fig. 10. Sample price pattern – non-algorithmic prices example 2.

from the first to seventh position. This in turn led to a complete reshuffle of sellers on the *compare all sellers* page.

In particular, the high-price but also high-rating sellers *Pink*, *Red*, *Brown* and *Cyan* moved to a more prominent display rank from the bottom of the list. The *Dark Purple* seller was low-price but also had a low (below-median) seller rating. The up-ranking of high-rated sellers changed the competitive environment from one where prices mattered for the ranking in the seller comparison to one where rating was more rewarded. Under these conditions, *Dark Purple* stopped experimenting and set a low fixed price to compensate for its relatively low rating.

In summary, in Figure 7a *balloons and rocks* pricing is a combination of price-experimenting (*balloons*) and an exogenous event that changes competition fundamentally, away from price towards the quality (rating) dimension. The *Dark Purple* seller rapidly needs low prices to compensate buyers for its quality handicap (*rocks*). The *balloons and rocks* pattern of *Dark Green* is fully explained by her strategy to always follow the Buy Box seller and undercut it by a small amount.

Off-the-shelf re-pricer solutions appear perfectly capable of defining scenarios like those in our explanations of the *balloons and rocks* pattern in Figure 7a, including the environment changing due to new entry and referencing rivals with certain characteristics, such as rating and delivery time.[33]

[33] See ChannelEngine (2021) and Figure 1 above.

The combination of these Algorithms is double-harmful to consumers: first, during *Dark Purple*'s experiment phase (the *balloons* period) prices are excessively high. Second, *Dark Green*'s follower strategy completely disqualifies her as a competitor, exerting as good as no pressure on *Dark Purple*, who usually wins the Buy Box, while Dark Green never does.[34]

A more quantitative approach to analyzing pricing patterns in algorithmic vs. non-algorithmic markets can focus on the analysis of statistical moments including skewness and kurtosis.[35] We summarize these moments by product category, looking at the difference between Buy Box prices and mean Buy Box prices for each product. The results are reported in Table A.9 and discussed in the Appendix. In our case, the comparison of algorithmic and non-algorithmic markets with respect to skewness and kurtosis does not show a clear picture. The skewness and kurtosis are sometimes higher and sometimes lower in algorithmic markets. The limited availability of data, especially the lack of demand data that prevents a classic market definition exercise, is a hindrance. Ideally, one would analyze how price residuals changed in an individual product market. Still, we believe that the statistical properties of algorithmic pricing patterns deserve further examination in a more focused study.

We now move on to the detailed econometric analysis of how algorithmic pricing affects Buy Box prices and sellers' prices.

5. Econometric analysis

Having established the prevalence of algorithmic pricing on *Bol.com*, in this Section, we examine the effect of algorithmic pricing on market outcomes in further detail. In the first sub-section, we investigate the effects of algorithmic sellers on Buy Box prices. We then perform a battery of robustness checks on our findings.

5.1. Algorithmic pricing and the Buy Box price

How do algorithmic agents affect the price of the Buy Box? The question is relevant in a screening exercise for algorithmic pricing. It provides useful guidance about whether policymakers should be concerned about algorithmic pricing in the first place. If algorithmic sellers are predominantly associated with low prices, there is little reason for regulatory attention. If however, algorithms go hand-in-hand with increased prices, attention may be warranted.

The Buy Box is without doubt the most valuable bounty for which firms on *Bol.com* compete. In broad terms, it appears to us that the market environment on a platform such as *Bol.com* can be looked at as price competition with a slightly heterogeneous, nearly homogeneous product. There are few markets where products would be as standardized as on *Bol.com*, with nearly a dozen sellers bidding for a product as specific as an *"Oral B Pro 2 – 2900 electric toothbrush, in double-pack."*[36] Nevertheless, sellers are somewhat differentiated, for example in terms of their rating and delivery time.

In such an environment, we would expect prices to rapidly converge towards marginal costs as the number of sellers increases. The price of a seller is an important factor based on which Bol awards the Buy Box, albeit not the only one.

Overall, the theory provides little certainty about whether algorithms can sustain higher prices or drive up competition. We aim to explore this question empirically. To do so, we study the effect of the presence of algorithmic sellers on the Buy Box price. We regress the following specification where the dependent variable $Log(BboxPrice)$ is the log-transformed Buy Box price of product i in period t.

$$Log(BboxPrice_{it}) = \beta_0 + \beta_1 Bol_{it} + \beta_2 N.Algo_{it} + \mathbf{X} + \mu_{id} \times \lambda_d + \epsilon_{it} \quad (1)$$

Bol_{it} indicates whether the platform operator competes as a seller for the respective product. Most importantly, $N.Algo_{it}$ is a set of dummy variables counting the number of algorithmic sellers on product i at time t with respect to the criteria and cut-offs described in Section 4.2. We further control for the Buy Box seller rating scaled between 0 and 10 as well as the Buy Box delivery time.

To control for seller heterogeneity and to increase the efficiency of our estimation, we include a vector of control variables $\mathbf{X}$. These include *Deliverytime*, the delivery time for a product in days (which was 3 days on average, see Table 6), and *Rating*, the seller rating on a scale from 1–10.

The key variable in our regression is $N.Algo_{it}$. The coefficient on, for example, $N.Algo = 1$ measures the average difference in log-price between the baseline category (no algorithmic sellers) and observations for a particular product at a particular time with one algorithmic seller. We observe within-day variation in the number of (algorithmic) sellers due to seller entry and exit. Sellers may enter or exit for a variety of reasons: a seller might run out of stock, Bol decides not to show the seller because her price falls outside the permitted range of prices for that product, or the seller decides (not) to offer a certain product during the day. However, as we cannot verify the reasons for seller entry and exit, and in particular whether entry and exit are exogenous to algorithm adoption, we refrain from interpreting this coefficient as a causal effect of the algorithm on prices.

The use of re-pricer software and hence algorithmic pricing is endogenous in two ways. First, a seller needs third-party software to engage in algorithmic pricing and may decide to acquire such software or not. Second, if a seller has the necessary software, she may still decide product by product to engage in algorithmic pricing (or instead, for example, leave prices constant). If the algorithm adoption and/or use decisions are correlated with product-time specific unobservables, OLS estimates may be attenuated. A common approach to mitigate biases in OLS estimates from endogeneity is the inclusion of fixed-effects (Angrist and Pischke, 2009, section "Individual fixed-effects"). This is the approach we follow in the paper. We use product-day-specific fixed effects.

The literature has dealt with the endogeneity of algorithmic pricing in different ways. Chen et al. (2016) leave causality largely aside and do not explicitly address endogeneity. In their analysis of algorithmic pricing in gas markets, Assad et al. (2020) use as instruments for adopting algorithmic pricing the share of other algorithmic stations of the same brand, as well as broadband availability around the station. Unfortunately, neither of these instruments can be directly applied to our setting.[37] We perform a robustness check using the number of algorithmic sellers on other products in the same product category as an instrument. The results are consistent with our main regressions, albeit coefficients are insignificant and the instruments weak.[38] We do not

[34] The behavior of *Dark Purple* and *Dark Green* sellers in Figure 7a is consistent with market sharing, whereby *Dark Green* would sell only to the most price-conscious, savvy users who click on the *"compare all sellers"* page and do not rely solely on the Buy Box.

[35] Automating the detection of geometric patterns in the time series of product prices is a significant practical challenge, especially with a high number of monitored products. A recent and comprehensive review of methods to detect geometric patterns in time-series data is provided in Shirato et al. (2023) and Andrienko et al. (2020).

[36] https://www.Bol.com/nl/nl/prijsoverzicht/oral-b-pro-2-2900-elektrische-tandenborstel-duopack/9200000117664163/, retrieved on the 17th of September 2021.

[37] A gas station belongs to a specific brand, and gas-brands manage many stations who individually decide about adopting algorithmic pricing software or not. Our outcome variable is the Buy-Box price, which varies by product-level. We have relatively few products in our sample offered by the same seller, and the same product may be offered by *multiple* sellers. These sellers may or may not use algorithms to set their prices.

[38] Detailed results are available upon request.

Table 4
Buy Box price by subsamples (Crawl 1).

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
Bol	-0.126***	-0.225***	-0.149***	-0.088***
	(-3.02)	(-5.31)	(-6.49)	(-5.40)
N.Algo = 1	-0.238***	-0.016	0.001	-0.006
	(-3.63)	(-0.71)	(0.15)	(-1.10)
N.Algo = 2			-0.002	-0.012
			(-0.18)	(-0.69)
Rating	0.073	0.038**	0.014	0.033***
	(0.55)	(2.14)	(1.25)	(2.97)
Deliverytime	0.003*	-0.001	-0.013***	-0.016***
	(1.85)	(-1.51)	(-4.54)	(-2.86)
Constant	2.626**	3.194***	3.313***	3.166***
	(2.24)	(20.58)	(31.85)	(31.82)
N	166352	156319	316877	98387
R2	1.000	1.000	1.000	1.000
ProductxDate FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

consider broadband availability as useful instrument in the context of e-commerce in Belgium and the Netherlands, where – unlike in Germany for Assad et al. (2020) – fast internet is widely available and both countries are extremely densely populated and highly digitized.[39]

Our empirical approach amounts to restricting variation in the data at the level of the same product within a day. The inclusion of fixed effects helps alleviate endogeneity concerns due to unobserved demand shocks. By including product-date fixed effects, we hope to largely eliminate time-specific demand changes. The rich set of fixed effects constitutes our main strategy to mitigate endogeneity concerns due to unobserved product characteristics, and algorithmic sellers potentially self-selecting into certain products. We do not aim to interpret our results in a causal manner. The main aim of our analysis is to provide methods to screen for potential algorithmic collusion, and raise awareness about the potential of this occurring. A correlational interpretation of results is sufficient for this objective.

To establish how the effect of the presence of algorithmic pricing on the Buy Box price varies by the degree of competition on a certain product, we estimate Equation (1) by looking at sub-samples cut at the number of sellers present.

5.1.1. Summary of results

We find no strong evidence that the presence of algorithmic sellers would increase prices, nor do we find robust evidence that in settings where all sellers use algorithms, higher average prices would prevail. If anything, there are settings in which the presence of pricing algorithms leads to lower average prices, particularly for products with one seller.

These results do not rule out algorithmic collusion. However, they cautiously suggest that if collusion between firms using pricing algorithms takes place, it is likely a fringe phenomenon rather than the average, at least on the online platform we have analyzed. The result that markets with a single monopoly seller go hand-in-hand with lower prices draws attention to the possibility that pricing algorithms may give rise to efficiencies.

The results hold across a large number of alternative criteria to identify algorithmic sellers and sub-samples cut at different numbers of sellers present. We present the regression results for our preferred criterion for algorithmic sellers in Table 4 and plot error bars around the main results in Figure 12.

There are few sub-samples in which for certain criteria we observe significantly higher average prices associated with algorithmic sellers, but these are rare and do not appear robust. These results also tend to be weakly statistically significant, which we do not lend much credence to in light of our large sample size. We discuss the results across the range of criteria and sub-samples below.

While our results suggest little concern with algorithmic pricing, we caution against interpreting *the absence of evidence* as *evidence of absence*. We do not know when sellers first adopted algorithms and if they did, how long it takes till eventual learning algorithms would converge to long-term strategies. We cannot exclude that the results change in the future. Furthermore, we only measure *posted* prices. We do not observe the prices consumers eventually pay. In particular, prices vary over time, often even throughout the day. If the high points of pricing cycles coincide with times of peak traffic, unsophisticated consumers may end up paying more than if prices were slightly higher on average but constant. Whether and to what extent consumers are aware of algorithmic pricing strategies on online marketplaces and how they adapt and adjust their shopping behavior is an interesting research question that exceeds the scope of this article. In line with the existing literature, we believe that posted prices are at least somewhat informative of market outcomes.

To test the robustness of our result, we use different conventional protocols of identifying algorithmic sellers: either via the frequency of price changes or by correlation with a relevant benchmark. We test 20 absolute price changes during the sample period, which is suggested by the empirical distribution of price changes,[40] as well as a more data-driven approach of frequent price changes that exceed the mean number of price changes in a category by either one or two standard deviations. We also test two protocols where we identify algorithmic sellers as those that show more than 20 price changes during the test period AND those prices correlate highly (Spearman's $\rho > 0.7$) either with i) Bol's price or ii) the lowest price for the product (second-lowest price in case a seller offers the lowest price herself).

Finally, we also compare the results to a scenario where a seller is flagged as algorithmic across all of her products if she has at least one product where she shows algorithmic pricing. Thus, we account for the possibility that sellers with the capability of setting algorithmic prices – for example being active users of a third-party repricer service – may *choose* to apply algorithmic pricing on subsets of their product portfolio.

5.1.2. Results

We present results on how the presence of algorithmic sellers affects the Buy Box price and how this effect changes with different numbers of sellers present at every given period t (Table 4). Coefficients on the number of algorithmic sellers by the total number of sellers are plotted in Figure 12.

The columns in Table 4 refer to the number of firms selling the product at a given time. Column 1 (Comp = 1) is the monopoly case, column 2 (Comp = 2) corresponds to the duopoly scenario and the other columns are analogous. The underlying criterion for algorithmic pricing applied here is prices changing sufficiently frequently, namely more than two standard deviations above the category average number of price change. In Section 5.2 we discuss results using other criteria to flag algorithmic sellers.

[39] Following a suggestion by a referee, we additionally tested several control variables related to seller-size to account for potential endogeneity arising from unobserved capabilities of large sellers in single-seller markets where we find our main result. These variables, portfolio size and presence across multiple categories, only had a negligible impact and are therefore not reported in the paper.

[40] Figure A.3 and Figure A.4 provide summary statistics on the share of algorithmic sellers by product as well as price category according to this definition of algorithmic pricing.

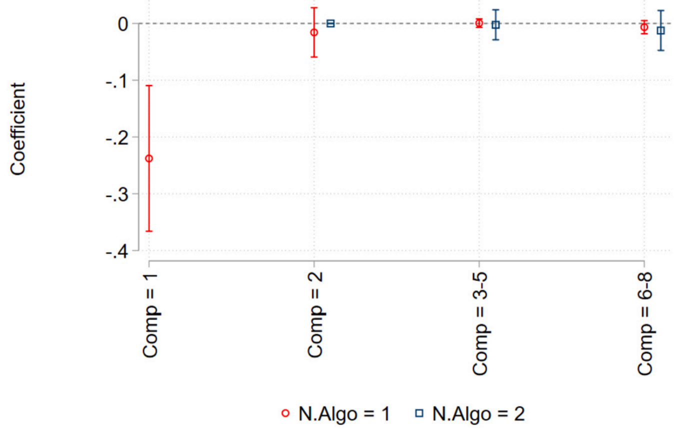


Fig. 12. Buy Box price by subsamples. (Crawl 1).

From the first column of Table 4, the Buy Box price *decreases* by circa 24% if the monopolist seller is algorithmic, compared to the case of a traditional seller acting in the same position. This finding is surprising, novel, and deserves explanation. It stands in stark contrast with the empirical results of Assad et al. (2020), who find that the adoption of algorithmic pricing software did not affect monopoly petrol station prices. One may suspect that the price reduction due to algorithmic sellers in monopoly markets arises because these sellers may focus on lower-price products in the first place. We exclude this explanation for at least two reasons.

First, we see no systematic relationship between the prevalence of algorithmic pricing and product categories or price classes (see Figures A.5 and A.6 in the Appendix). Second, our estimated Equation (1) includes product-date fixed effects. We capture the effect of algorithmic pricing primarily by the variation of the Buy Box price *within* products on the same day. For these reasons, we consider it unlikely that the price reduction due to algorithms in monopoly markets would have to do with the products of algorithmic and non-algorithmic sellers being systematically different.

We provide a simple explanation for the large algorithmic price rebate in single-seller markets. In markets such as petrol stations (Assad et al., 2020), each facility has a manager who makes pricing decisions for a narrow set of petrol products. On e-commerce marketplaces such as Bol, third-party sellers often carry tens of thousands of items.[41] While in a petrol station a manager needs to review merely three or four petrol product prices, pricing in thousands of products on Bol separately would clearly be a challenge even for a large team of product managers.

The most likely and practical way for non-algorithmic sellers to determine the prices of a long list of products on e-commerce platforms – at least for the first upload of prices – seems to be applying a simple formula: by first summing up costs and then adding a margin that – for practical reasons – is likely equal for several products. External pricing tools explicitly recommend this *cost-plus* pricing approach.[42] Clearly, a uniform margin may not be profit-maximizing for all products. However, the effort required to adjust prices for thousands of products separately can prove prohibitive for sellers.

When a traditional seller subscribes to a re-pricer, she likely starts out by uploading a list with regular product prices based on such a *cost-plus* formula. The re-pricer then takes over the pricing of individual products: within the freedom entrusted by the manager, the re-pricer software can apply *if-then* price rules, or even experiment with and adjust those regular prices. In monopoly markets, the re-pricer algorithm may end up pushing prices downwards, at least for some products, where the regular margins may result in less-than-optimal demand. Re-pricer ChannelEngine offers precisely this pricing procedure for monopoly products (*"if no competition"*) on Bol, by automatically setting the *"regular price, with price rules applied"* (ChannelEngine, 2021).

Note that by the same mechanism, whether algorithms reduced or increased prices compared to manual sellers depends entirely on the initial bias made by the human seller. Algorithms entrusted with the objective of maximizing profits would merely adjust the price to the profit-maximizing monopoly level. This is a downward adjustment, as we find if humans initially set these prices excessively high.[43] Clearly, by the same mechanism algorithms could as well result in an *upward* price adjustment, if the human bias meant lower-than-monopoly prices in the first place. Our findings suggest that the net effect is on average a *downward* adjustment of manually set prices.

Consistently with this explanation, we observe that human sellers tend to disproportionately change prices on Thursdays, and around peak office hours, at 8–9 AM and 4–6 PM (Figure 5). In contrast, algorithmic sellers change prices evenly across various weekdays, and are predominantly at work during the night, a few hours after the main human price upload at 4 PM. This lends some support to the view that algorithms continuously adjust human-uploaded prices.

We conclude that algorithmic agents may in some cases reduce prices, and observe this at work in monopoly markets. We explain this phenomenon by automated pricing engines applying different margins product-by-product, in a more granular manner than what would manually be feasible.

We now move on to discussing the results in Table 4 for products with more than one seller present. In column 2, we present the effects for duopoly products. We do not observe products where two algorithmic sellers are active at the same time. Where one algorithmic seller is present, we do not find an effect on prices. The coefficient on the presence of one algorithmic seller is very close to zero. Columns 3 and 4 present results for 3–5 and 6–8 competitors, respectively. Here, too, do we not find evidence of an effect of algorithms on prices. The point estimates are close to zero, sometimes also slightly negative. As Figure 12 shows, estimates are more precise than for single-seller markets in columns 2–4, yet the results are consistently centered around 0.

5.2. Robustness and further results

In this section, we explore the robustness of some of the results established in the previous section.

5.2.1. Seller-level definition of algorithmic pricing

In e-commerce, sellers likely rely on algorithmic pricing software to manage price-setting for a broad range of products. In the analysis so far, we defined pricing as algorithmic on the *product-seller* level. The same seller is considered algorithmic only on some of her products, where prices *behave algorithmically* according to our criteria.

We repeat much of the above analysis by categorizing a seller as algorithmic across *all* products if her prices on *any* product satisfy our

[41] Bol displays the number of products by seller. We browsed several arbitrary sellers to find that products typically range in the thousands, with several sellers having more than 40,000 products.

[42] See for example https://www.woosa.com/software/bol-woocommerce-addon-price-calculator/?v=d3dcf429c679 and https://www.shopify.com/blog/how-to-price-your-product (both retrieved on the 1st of December 2021), explaining the procedure as follows: *"To set your first price, add up all of the costs involved in bringing your product to market, set your profit margin on top of those expenses, and there you have it."*

[43] There is a rich body of business literature documenting managerial overconfidence, among others about demand (Montgomery and Bradlow, 1999; Feiler and Tong, 2022), new product introduction (Simon and Shrader, 2012; Markovitch et al., 2015), and other corporate decision variables (Malmendier and Tate, 2015). Furthermore, managerial overconfidence is documentedly persistent (Huffman et al., 2022). Kahneman (2011) recognizes overconfidence as "*the most significant of the cognitive biases.*"

Table 5
Number of product-seller pairs by algo-definition. **Total:** 6752.

	1 SD	2 SD	20 Changes	Corr. min. price	Corr. Bol price
always	2248	2318	2757	2763	2812
sometimes	3150	1170	480	225	232

criteria.[44] Neither of these approaches is superior. We consider defining a seller as algorithmic on the seller-product level more relevant for policy: if the aim is screening for potential algorithmic collusion, it is useful to draw attention to prices on those products where the bulk of algorithmic pricing is taking place.

The change in definition changes the composition of control and treatment groups. When defining algorithmic pricing on a seller-level, the (algorithmic) treatment group includes not only products of sellers where prices change often but also products of the same seller that do not fulfill this criterion. The (non-algorithmic) control group in turn includes products of sellers who set relatively stable prices on all of their products, so that they do not exhibit patterns of algorithmic pricing on any part of their portfolio.

Table 5 shows the number of product-seller pairs depending on whether a seller changing prices sufficiently often on any product is regarded as algorithmic on all her products (*always algo*) or only those product-seller pairs are deemed algorithmic where such behavior is actually observed (*sometimes algo*). We observe that sellers with algorithmic capabilities *choose* relatively (but not necessarily fully) stable prices for a substantial number of products in their portfolio. By labelling product-seller pairs as algorithmic where we observe such behavior transfers these product-seller pairs from the treatment into the control group.

Table A.1 presents results for the scenarios examined in previous sections (a seller is flagged as algorithmic on all products if they are flagged as algorithmic on at least one product).[45] As we can see, results point in the same direction although at a larger magnitude.

We note that Table A.1 carries an important message for policy. It shows that even with simple methods such as those presented here, practitioners interested in screening the horizon for potential algorithmic collusion are able to restrict attention to 1 out of 4 to 7 product-seller pairs, where algorithmic pricing is actually observed.[46]

5.2.2. Substitution patterns and markets

Our analysis is at the level of individual products. This is a necessary simplification as it is not feasible to conduct a separate market definition (in the competition-policy sense) for every individual product. However, to capture the most salient substitution patterns, we repeat our analysis using two different approaches to categorize products that may be close substitutes.

We have identified products that are potentially close substitutes based on a) the similarity of product description texts and b) the product category variable in the data set. We find that our main results are robust even when considering potential substitution patterns across products.

As one approach, we combined the product title and description into a single string and identified similar products based on the *cosine similarity* between the various product strings. Cosine similarity is a standard text mining metric quantifying similarity between two documents based on the co-appearance of various terms. A cosine similarity between two documents of 1 implies identical content (but ignoring the order of terms) while zero indicates full orthogonality.

For a further robustness check, we also define products with identical third-level product categories (the lowest-level product category in our data) as belonging to a single product market. This includes product categories such as *"puzzles"*, *"PS4 games"*, *"vacuum cleaners"*, *"TV sets"*, *"HDMI cables"*, but also less obvious groups of products, such as *"religion"* and *"biology"*. By this measure, we have 668 markets encompassing 2370 products. Around 45% of these category-3-level markets have one product, the remaining categories combine at least two products.

Regarding the distribution of algorithmic sellers in these two samples, we observe no anomalies.[47]

Under both definitions, we aggregate the number of sellers and the number of algorithmic sellers across *similar products*. Generally, this results in products having an increased number of sellers, because related products become a single product with all their sellers. As we now have similar numbers of observations for *markets* with different numbers of sellers, we run our regression separately by total number of sellers using the same specification and control variables.

The coefficients of the main independent variable (number of algorithmic sellers) are reported in Tables A.8 and A.7. The former confirms our finding of a highly significant, large negative effect in markets with one seller with a coefficient of −0.157. In Table A.8 we also find a negative effect of a single algorithmic seller when there are 3 sellers. Other effects are either insignificant or generally small. In Table A.7, all coefficients are small and insignificant (one effect is marginally significant at the $p<0.1$ level).

5.2.3. Seller prices and margins

In a screen for potentially harmful pricing behavior, margins are particularly relevant, as high margins may serve as markers for collusion or sustained high prices. A drawback of our data-sparse screening approach is that we do not observe costs directly. However, since marginal costs are unlikely to change during the relatively short period of our analysis, we believe seller prices are closely correlated with margins.[48] We explicitly analyze seller prices here, depending on the number of players in the market and in particular, the number of algorithmic sellers. Under the assumption that marginal costs are constant in our sample period, this analysis is analogous to that of margins.

We estimate the following equation:

$$ln(Priceijt\%) = \beta_0 + \beta_1 Bol_{it} + \beta_2 N.Algo_{it} + \mathbf{X} + \mu_{id} + \phi_j + \epsilon_{ijt} \quad (2)$$

with $ln(Price_{ijt}\%)$ corresponding to the logged price of seller j on product i at time t. In addition to product-date fixed effects, we include seller-fixed effects to account for unobserved seller heterogeneity.

Results in Table A.3 are analogous to what we have previously established on Buy Box Prices. In monopolies, prices are around 13% lower if the seller is algorithmic. When there are many (6–8) competitors, a single algorithmic seller tends to charge slightly higher prices. This turns around when two algorithmic sellers are present, where prices are slightly lower on average. With intermediate competition, algorithmic sellers are not associated with different prices than non-algorithmic rivals.

5.2.4. Seller heterogeneity and endogeneity

Sellers that adopt pricing algorithms are potentially different from those that do not. Although we control for seller characteristics, including ratings and delivery time, there may be other unobserved differences

[44] An earlier, working paper version of this article presented the main results focusing on the latter definition of algorithmic sellers.

[45] As we have substantially fewer product-seller pairs that are labeled algorithmic (see Table 5), the coefficient for two algorithmic sellers in a duopoly is now mostly omitted.

[46] This is the approximate share of the "*Sometimes Algo*" row on the total number of product-seller pairs.

[47] We discarded a third approach of grouping potential substitute products: we scraped the *"others also looked at (vaak samen gekocht)"* links for each product and saw whether the recommended products appear among our list of products. We ultimately discarded this approach because a large share of the recommended products appeared complements rather than substitutes.

[48] For example, if marginal costs (mc) are constant over the few weeks of our analysis, then the margins for a seller j on product i at time t is $m_{ijt} = p_{ijt} - mc$, which is perfectly correlated with p_{ijt}.

between sellers. We would not expect these differences to render pricing algorithms meaningless, otherwise, it would be hard to explain why certain sellers would adopt algorithms in the first place. Indeed, including seller-fixed effects supports our finding that the effect of algorithmic pricing on prices is negative when there is a single seller and not distinguishable from zero when there are more sellers. Table A.2 shows the results when adding SellerID-fixed effects to the fixed effects for product-date combinations from the main regression. This refers to the SellerID of the Buy Box seller. This suggests that controlling for unobserved seller differences does not explain away our findings, especially the negative coefficient for products with a single seller.

5.2.5. Flagging algorithmic sellers: price changes and price correlations

We verify whether the results of our analysis hold for different definitions of algorithmic pricing. In the main part of the paper, we have flagged a seller as algorithmic if the prices of the products changed sufficiently often. As a baseline, we assumed that price changes that are more than *two* standard deviations above the normalized distribution within the product category are due to algorithmic pricing. We tested a variety of alternative criteria to flag algorithmic sellers, for example by applying both one standard deviation as alternative cut-off as well as a fixed number of 20 price changes. This cutoff is motivated by our visual inspection of the empirical distribution of price changes.

As a further refinement, we add price correlations with competing sellers to the number of price changes, as discussed in section 4.2. Doing so results in labeling a seller as algorithmic if prices change often *and* closely follow either the lowest price (or the second-lowest price if they offer the lowest price themselves) or Bol's price.

The results of this analysis are saved in a separate Online Appendix.[49] The results from our baseline specification holds for most of these different specifications. Coefficients decrease by 1 to 2 percentage points as we choose a more conservative measure of algorithmic pricing.

5.2.6. Second crawl

Lastly, we establish that the previous largely holds based on the data we have gathered about a year later. The results are presented in Table A.5. Importantly, our main qualitative results do not differ significantly between the two crawls.[50] This lends credence to the robustness of our results over time.[51]

6. Policy discussion and conclusion

To our knowledge, we are the first to document that algorithmic pricing may involve efficiencies in the form of lower prices in monopoly markets. We explain this finding by the inability of traditional sellers to accurately determine the profit-maximizing monopoly price for thousands of products in their portfolio. Algorithmic agents may start with imperfect prices but gradually converge to the monopoly price by experimenting.

We provide a method for identifying algorithmic sellers based on heuristics. This is useful for screening purposes, to focus attention to particular sellers to investigate further. We also highlight several pricing patterns that arise with pricing algorithms and discuss how they may be consistent with collusion.

Overall, we find little robust trace of algorithmic sellers increasing prices and benefiting from each other's presence for multi-seller products. While in some specifications we observe – consistently with collusion – that the presence of algorithmic sellers is associated with higher prices and margins, these results are not particularly robust. Establishing a *correlation* between prices and the presence of algorithmic sellers is the first screen analysts should undertake to uncover algorithmic collusion. Competition agencies *screening* for anti-competitive behavior should be as first step concerned with algorithmic pricing if it is associated with higher prices.

Screening for algorithmic collusion should remain on the agenda of academics and increasingly, of competition practitioners. Tacit collusion is a grey zone in competition policy. Competition law in the US and the EU requires a *"meeting of minds"* between rival firms. Price parallelism in itself does not constitute illegal behavior. However, abnormally high prices, moving in parallel over time, without obvious common costs should be suspicious. Antitrust agencies can deploy further investigative measures – such as unannounced inspections – to verify whether the sellers are aware of each others' actions. These investigative measures can uncover whether a *"meeting of minds"* took place. Data analysis like presented in this article can help practitioners to zoom into potential antitrust law violators.

Previous research on algorithmic collusion highlighted as the main policy question whether collusive prices arise due to algorithms *"failing to learn to compete"* (Hansen et al., 2020) or actually *"learning to coordinate"* (Calvano et al., 2020). Some commentators (e.g. Assad et al., 2020) argue that competition policy *"should mostly be concerned with algorithms actively learning not to compete."* We believe while this distinction is intellectually interesting, it carries only limited policy relevance. Antitrust agencies rightly argue that *"companies cannot hide behind algorithms"* (Laitenberger, 2017; Busse, 2017).[52] Competition policy looks at algorithmic agents applying the same criteria as to human decision-makers. Whether a manager charged with collusion *"learned to coordinate"* or *"failed to learn to compete"* would likely make little difference for most judges deciding in antitrust matters. The first-order policy question is whether algorithmic pricing is associated with higher prices. We find little trace of this on our chosen e-commerce platform.

A limitation of our analysis is that we do not *identify* collusion. We do not intend to *prove* or even allege collusion. We aim to investigate the likelihood of pricing algorithms to increase prices in a real market environment. We furthermore aim at creating simple screens that can be used to narrow the search for algorithmic collusion. Our descriptive results carry relevance for competition authorities, researchers, and managers scanning the horizon for potentially collusive practices: we propose a list and frequency of potentially problematic price patterns that can serve as a simple first screen to shortlist firms and products for further analysis. We do so in an extraordinarily data-sparse manner, relying solely on publicly available price information.[53]

Overall, our recommendation to authorities is to look specifically for pricing patterns where several sellers in a market use "rocks and balloons" as this is the pricing pattern consistent with collusive pricing, undercutting, and gradual "forgiveness" of a deviation from a high price. Furthermore, our impression is that sellers on *Bol.com* using re-pricer software are relatively unsophisticated. We see little trace of complex

[49] See https://github.com/philiphanspach/Algorithms-in-the-marketplace.

[50] The first crawl includes data from the holiday and New Year period. While our primary aim is to devise collusion screens and not to perform analysis that is representative of any one platform, this period may be special in terms of pricing strategies and product availability. The fact that we see similar results in Crawl 2 strengthens the external validity of our findings. As opposed to COVID in the second crawl, the end of the holiday period in Crawl 1 is not special in terms of demand to Bol.com (see Appendix, Figures 13 and 14).

[51] We provide further robustness tests in the appendix, which are not discussed in detail in this article: Table A.4 presents results using product-week fixed effects. Table A.6 presents results using fixed effects by date and quartile of the average Buy Box price. Table A.10 presents results using the minimum price rather than the Buy Box price as dependent variable. Table A.11 presents results excluding products on which Bol competes as a seller. None of the overall conclusions of this article are called into question by these ancillary results.

[52] They undoubtedly try, see Feier et al. (2021).

[53] We obtain our data by scraping the platform. Other methods of obtaining similar data may include API access, which Bol used to offer to individual sellers and recently to professional sellers only. Depending on the precise question studied, third-party providers (such as Keepa.com) also offer similar data from e-commerce platforms for sale, more extensive in some directions but less detailed in others.

learning behind the documented pricing patterns. The strongest evidence consistent with algorithmic learning appears to be the *random jumps* pattern illustrated in Figure 8, which we observe in just 11% of affected (algorithmic) products. This pattern may indicate algorithms being *trained* as sellers experiment with different prices.

As also illustrated in Figure 1, the bulk of algorithmic pricing on Bol.com appears to consist of a finite set of *if-then* statements. This apparent lack of sophistication may not make pricing algorithms less harmful. On the contrary: a secret to successful collusion may lie in managers' ability to commit to simple strategies, such as leader-follower prices shown in Figure 7a via the simple *if-then* formulae in off-the-shelf repricer software.[54]

While we find exceptional examples where algorithmic sellers appear associated with higher average prices, we observe instances where algorithmic pricing appears non-profitable for *some* algorithmic sellers.[55] The Dark Green seller in Figure 7a provides an example. This seller automatically follows Dark Purple to raise prices. But Dark Green never wins the Buy Box and therefore appears to sell very little, making it unlikely to benefit much directly from the increased prices. This has practical implications for competition policy, where firms accused of algorithmic collusion may (truthfully) argue that algorithmic pricing never benefited them. It also shows the limits of purely data-driven analysis to *prove* collusion from pricing data: These techniques can serve as a useful first screen, but traditional cartel investigative tools such as unannounced inspections and formal information requests appear irreplaceable to understand whether and how algorithmic sellers work together, communicate and allocate profits.

Our econometric analysis of algorithmic prices yields nuanced results. We document that algorithmic pricing may involve pro-competitive effects, namely in monopoly markets. It likely simplifies price-setting product-by-product and may counter-steer human errors in over-estimating demand and setting excessive prices. We see little evidence of algorithmic collusion as the number of sellers increases.

We also find that algorithmic sellers tend to price very differently within their own product palette. On some products, they change prices frequently, and these prices are on average higher than those of comparable non-algorithmic firms. However, on a broad range of their products, algorithmic agents cannot be distinguished from non-algorithmic sellers. This also implies that (relatively) stable prices are not particularly useful markers in screening for potential algorithmic collusion. Stable prices of sellers with algorithmic capabilities are not higher than of sellers without the ability to engage in algorithmic pricing.[56]

While in this paper we focus on *what comes out* of algorithms, designing appropriate policies to improve market outcomes would also require a detailed understanding of *what goes in*. For example, algorithms that take as input the prices of rivals may cause more harm in competitive markets than algorithms that merely experiment and observe the resulting demand (Morton, 2012). In a similar vein, understanding what happens *inside* the algorithm – the formula the algorithm uses to recommend a price based on the input signals – is very likely also necessary for an apt policy response. Better understanding these questions constitutes possible avenues for further research.

[54] Salcedo (2015) and Brown and MacKay (2020) emphasize the role of algorithmic agents as commitment devices to elevated prices.

[55] Normann et al. (2021) also report a similar finding.

[56] While in theory, stable prices may be markers of successful collusion, they are not particularly useful as *screens* for collusion. Neither managers deciding about dedicating resources to investigating a group of sellers for collusion, nor courts would likely be particularly persuaded by the argument that these sellers display "*suspiciously stable prices.*" While no *proof*, parallel prices appear to be a more workable collusion *screen*.

CRediT authorship contribution statement

Philip Hanspach: Writing – review & editing, Visualization, Validation, Software, Project administration, Investigation. **Geza Sapi:** Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Marcel Wieting:** Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A

A.1. Cosine similarity and product categories to identify potential substitute products

We have identified products that are potentially close substitutes based on a) the similarity of product descriptions and b) the hierarchy of product descriptions. For approach a), we have combined the product title and description into a string and identified similar products based on the cosine similarity between the various product strings. However, we cannot reject the hypothesis that there is no effect of algorithmic pricing on the products that we can classify into larger markets using this approach.

A *document* in our analysis consists of a product's *Title*, the *Brand* (including the *Series* if available) as well as the detailed *Product Description*. For example, for a product the Title may be *"RITUALS The Ritual of Sakura – Trial Set"*, the Brand "*RITUALS*", the Series *"The Ritual of Sakura"*, and the Description:

> *"Trakteer jezelf of een van je dierbaren op deze heerlijk geurende set van The Ritual of Sakura. Deze trial set is de perfecte kennismaking met de producten van Rituals, een ideaal cadeau voor vrienden en familie. Bevat een shower foam, een bodycrème en een 91% natuurlijke suikerscrub. Alle producten zijn verrijkt met de vernieuwende geur van kersenbloesem en rijstmelk."*

A document for the purpose of our analysis is a merge of all these fields into a single string. We pre-processed the text and converted all strings into lower case, removed special characters followed by a total of 280 Dutch and English stopwords (using those included in the Python NLTK package). We then perform stemming using the Snowball stemmer, a standard step in text analysis.

We used Python's "Sklearn Countvectorizer" to create a document-term matrix, where each row corresponds to a document and each column corresponds to a term, with as many columns in the matrix as the total number of terms across all documents. The cells in the document-term matrix count the occurrence of a particular term in a particular document.

Based on this matrix, we calculate the cosine similarity across documents. The cosine similarity is a standard text mining metric quantifying similarity between two documents based on the co-appearance of various terms. A cosine similarity between two documents of 1 implies identical content (but ignoring the order of terms) while zero indicates full orthogonality. We experiment with thresholds and choose 0.75 above which we consider two documents as similar. For example, this threshold categorizes several watches of different brands as similar.

Furthermore, toys that appear as potential substitutes such as various chessboards, puzzles, Playmobil, and LEGO models show up as sufficiently similar. The null result remains for a variety of other thresholds in the neighborhood of 0.75. The results are reported in Table A.7. Much higher or lower thresholds lead to either fewer products that are identified as similar to other products or to sets that clearly do not include substitutes.

Two examples showing the difficulty in categorizing products as substitutes based on text mining are that the same computer game on different console models also shows up as similar, while these are unlikely to be considered substitutes by consumers, who typically either have one console or the other. Also, clothes of the same brand of different sizes (e.g. S, M, or L) arise as similar. Despite these exceptions, most products classified as "similar" appear reasonable. Overall, this exercise yields 706 unique pairs of products that are sufficiently similar. For the purpose of this exercise, we assume that they form a single relevant market.

In a further robustness check, we define products with an identical value in the lowest-level product category in our data (such as "wall decoration", "dustbins", and "USB sticks") as belonging to one product market. By this measure, we have 668 markets encompassing 2370 products. Around 45% of these category-3-level markets have one product, or conversely, around 12% of the 2370 products form a single category on their own, the remaining 88% of products share a category with at least one other product.[57] The results are reported in Table A.8. The negative coefficient for the presence of algorithms on prices when there is a single seller is still obtained when we form sets of similar products by product category.

Table A.9 allows a quantitative comparison of algorithmic and non-algorithmic markets, adding to the qualitative examination of pricing patterns in Section 4.3. This table shows the descriptive statistics of the total Buy Box price minus the average Buy Box price per product, aggregated across product categories. We compare markets where no algorithmic sellers are present (left-hand-side) with those where at least one is present (right-hand-side), dropping categories where we observe none of one kind (Auto & Motor, Living, DIY).

A result consistent with higher collusive prices associated with algorithmic pricing patterns such as "rockets and feathers" would be an increase in skewness in algorithmic versus non-algorithmic markets. A consistently positive and high skew in algorithmic markets would indicate that the deviations of Buy Box prices from the product mean are more concentrated towards the right side, and the left tail is spread out. Hence, the statistical results are bent towards the left-hand side.

In our case, the comparison of algorithmic and non-algorithmic markets with respect to skewness and kurtosis does not show a clear picture. The skewness and kurtosis are sometimes higher and sometimes lower in algorithmic markets. The limited availability of data, especially the lack of demand data that prevents a classic market definition exercise, is a hindrance. Ideally, one would analyze how price residuals changed in an individual product market. Still, we believe that the statistical properties of algorithmic pricing patterns deserve further examination in a more focused study.

A.2. Crawl 2

Table 6 presents summary statistics for the underlying dataset with respect to both waves of the data.[58] We present the main results using data from Crawl 1 and provide the most corresponding results for Crawl 2 as a robustness check in the Appendix. Crawl 1 is particularly interesting for several reasons. First, it has more products than Crawl 2, as shown in Table 6. The number of products is reduced by around 900 in the second crawl. This is because we kept the same list of URLs for both crawls, which consisted of top-selling products in early 2019. By the time of Crawl 2, in the spring of 2020, a large number of these products, such as music, computer equipment, and fashion items, became outdated and disappeared from the marketplace platform.

Second, for the remaining products, the number of sellers increased rapidly, with the consequence that the number of monopoly products reduced significantly in Crawl 2 (Figure A.7). Monopoly products constitute an interesting benchmark in our analysis and are better captured in Crawl 1. Finally, part of Crawl 2 coincides with the first lockdowns during the COVID-19 pandemic, and a steady increase in visits to the Bol platform (see Appendix, Figures 13 and 14). During these days many sellers had to swiftly expand their online presence due to store closures, some products experienced large demand shocks, and there were delivery bottlenecks that may have affected seller ratings. This may imply potentially confounding factors in Crawl 2 that may deserve a dedicated analysis.

Overall, we are confident that the first crawl constitutes business as usual for Bol, with average website visits. The second crawl is relatively

Table 6
Summary statistics.

	Crawl 1 Mean (sd)	Crawl 2 Mean (sd)
BuyBox Price in EUR	45.04 (87.29)	39.34 (88.55)
Price in EUR	50.03 (87.40)	43.04 (89.59)
Seller Rating (1–10)	8.78 (.44)	8.75 (.58)
Delivery Time in Days	2.99 (2.92)	3.77 (2.47)
Nr. of Sellers per Product	6.05 (2.74)	5.51 (2.65)
Shipping Fees in EUR	.03 (.27)	.03 (.31)
Crawl Frequency in Min.	122.85 (453.82)	32.89 (439.67)
N	2437557	17066561
Products	2846	1949
Sellers	1871	2190
Period	Dec 18–Jan 19	Feb–Mar 20

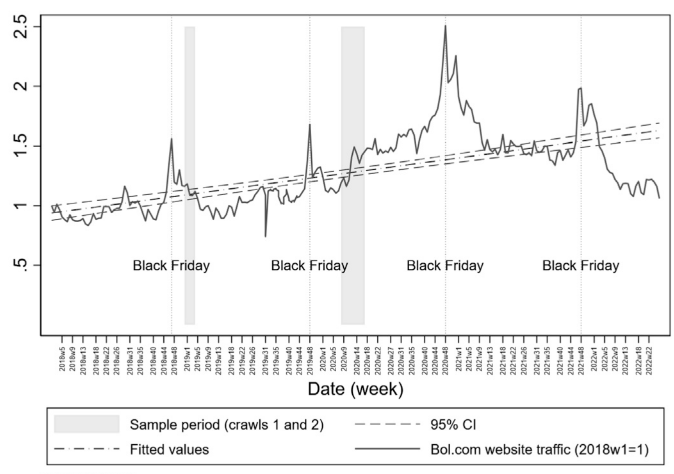


Fig. 13. Website traffic to Bol.com. Source: similarweb.com, for the period prior to 2020, Congiu et al. (2022).

[57] More detail on product categories is available upon request from the authors.

[58] The second crawl comprises substantially more observations than the first crawl due to the higher crawl frequency of scraping approximately circa every 30 minutes, rather than every 120 minutes.

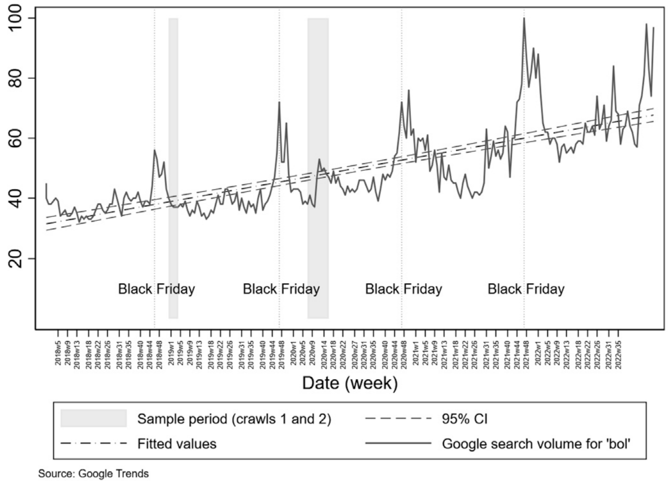


Fig. 14. Google search volume for the term 'bol', Netherlands. Source: Google Trends.

normal in terms of demand to the Bol platform, but it partly coincides with the first Covid lockdowns, and with a steady increase in website visits over time. Website traffic to Bol.com over time is shown in Figures 13 and 14.[59]

Figure A.7 shows the distribution of products by the number of sellers.[60] Products offered by ten or more sellers are grouped in the last bin of *ten sellers*. Two things are noteworthy in Figure A.7. First, the typical product tends to see three to four sellers. Second, as explained above, the number of monopoly products reduced significantly in Crawl 2. This is because the products that were best-selling in Crawl 1 likely became less novel in more than a year of time by Crawl 2, as more sellers had time to stock up.

Figure A.1 shows the average Buy Box price over product categories. The Buy Box price tends to be highest in electronic products, men's fashion, and bike accessories. Health care, books, and music are the lowest-price categories. Average Buy Box prices are very close in both crawls, but there was a large reduction in Crawl 2 in the top categories. This is consistent with the view that many products in computer, men's fashion and bike accessories became outdated till Crawl 2: these are fast-moving product categories. In fashion and computers product life cycles are particularly short.

Table 6 provides the main summary statistics for both crawls. It shows that the average Buy Box price is lower than the average price per product. This is not surprising, since Bol takes into account prices when awarding the Buy Box to a particular seller (see Section 2.2). Both, the seller prices as well as the Buy Box price are highly left-skewed. Prices and the number of products reduced on average between the two crawls. This is consistent with the view of a subset of products churning over time as they become outdated.

A.3. Further figures and tables

Table A.1
Seller Level Algo Definition, main specification.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
Bol	-0.090*	-0.228***	-0.145***	-0.090***
	(-1.65)	(-5.21)	(-6.33)	(-5.20)
N.Algo = 1	-0.043	0.011	-0.001	-0.008
	(-0.44)	(0.55)	(-0.14)	(-0.99)
N.Algo = 2	0.069	0.039*	-0.005	-0.003
	(0.52)	(1.69)	(-0.54)	(-0.35)
N.Algo = 4		0.029	-0.018*	-0.001
		(0.87)	(-1.74)	(-0.17)
N.Algo = 3			-0.011	-0.003
			(-1.16)	(-0.29)
N.Algo = 5			-0.020*	-0.012
			(-1.85)	(-1.11)
N.Algo = 6			-0.014	-0.024*
			(-1.27)	(-1.80)
Rating	0.069	0.037**	0.016	0.034***
	(0.61)	(2.13)	(1.34)	(3.00)
Deliverytime	0.003*	-0.001	-0.013***	-0.015***
	(1.86)	(-1.50)	(-4.55)	(-2.99)
Constant	2.650***	3.178***	3.303***	3.161***
	(2.64)	(20.42)	(30.16)	(30.70)
N	166352	156319	316877	98387
R2	1.000	1.000	1.000	1.000
ProductxDate FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2
Seller fixed effects added.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
Bol = 1	-0.126***	-0.225***	-0.149***	-0.0881***
	(-3.02)	(-5.31)	(-6.49)	(-5.40)
N.Algo = 1	-0.238***	-0.0157	0.000567	-0.00649
	(-3.63)	(-0.71)	(0.15)	(-1.10)
N.Algo = 2			-0.00239	-0.0124
			(-0.18)	(-0.69)
Rating	0.0733	0.0377**	0.0145	0.0334***
	(0.55)	(2.14)	(1.25)	(2.97)
Deliverytime	0.00254	-0.000564	-0.0129***	-0.0156***
	(1.85)	(-1.51)	(-4.54)	(-2.86)
Constant	2.626**	3.194***	3.313***	3.166***
	(2.24)	(20.58)	(31.85)	(31.82)
N	166352	156319	316877	98387
Seller,ProductxDate-FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Seller prices. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[59] To verify that website traffic is not abnormal during our crawl periods, we obtained website traffic data for bol.com from Similarweb (similarweb.com) reaching back till 2018. We thank Similarweb for providing us this data. In terms of website visit volume to bol.com, Crawl 1 is fully within the 95% confidence interval. Crawl 2 overall coincides with a moderate run-up in website visits. Traffic on Bol.com tends to peak before and after Black Fridays in late November. Neither of our crawl periods falls within such a peak demand time range. The same is confirmed in Google Trends search volumes for the term "bol" in the Netherlands.

[60] While we talk about *monopoly* and *duopoly* products, it is important to clarify that this is not meant in the *antitrust sense* and does not intend to serve as a *market definition* in any way other sense than the natural means to refer to products in the data being sold by one or two sellers. It appears likely that even *monopoly* sellers of a particular product face competition from other products in our data. Since our main aim is to develop workable screens for collusion and to investigate whether algorithmic sellers may be associated with higher prices, looking at the number of sellers per product as classified by Bol suffices for our purposes.

Table A.3
Seller prices.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
Bol comp. = 1		-0.226***	-0.144***	-0.157***
		(-16.86)	(-68.42)	(-73.02)
N.Algo = 1	-0.128***	-0.00696	0.000750	0.00444***
	(-7.69)	(-1.06)	(1.18)	(7.18)
N.Algo = 2			0.00163	-0.00338***
			(0.39)	(-3.40)
N.Algo = 3			0.00154	
			(0.37)	
N.Algo = 4			-0.00603	
			(-1.38)	
N.Algo = 5			-0.0113*	
			(-2.44)	
Rating	0.0920***	-0.000312***	-0.0000314	0.000229**
	(3.84)	(-4.92)	(-0.73)	(3.23)
Delivery Time	-0.000239	-0.0000300*	0.000247***	-0.0000677*
	(-1.27)	(-2.21)	(15.06)	(-2.51)
Constant	3.020***	3.529***	3.416***	3.208***
	(18.34)	(422.47)	(2302.61)	(1918.16)
N	129377	314221	1208299	540168
R2	1.000	1.000	1.000	0.998
ProductxDate FE	Y	Y	Y	Y
Seller FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4
Product-week fixed effects.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
N.Algo = 1	-0.123	-0.019	0.012	-0.007
	(-1.10)	(-1.61)	(1.46)	(-0.84)
N.Algo = 2			0.001	-0.032*
			(0.04)	(-1.94)
Bol	-0.170***	-0.159***	-0.131***	-0.082***
	(-5.85)	(-7.04)	(-6.76)	(-4.51)
BBoxRating	0.009	0.047**	0.016**	0.024**
	(0.59)	(2.35)	(2.02)	(2.22)
BBoxDeliveryTime	0.001	-0.000	-0.008***	-0.015***
	(1.06)	(-0.77)	(-4.57)	(-3.49)
Constant	3.232***	3.073***	3.278***	3.247***
	(24.31)	(17.63)	(45.92)	(35.12)
N	169394	157687	318814	99005
R2	0.999	0.999	0.999	0.999
ProductxWeek FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5
Buy Box price by subsamples. (Crawl 2).

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
Bol comp. = 1	-0.073***	-0.019	-0.041***	-0.124***
	(-8.27)	(-1.62)	(-19.94)	(-30.96)
N.Algo = 1	-0.062***	0.036*	-0.002***	0.009***
	(-4.64)	(1.76)	(-3.74)	(9.84)
N.Algo = 2		0.035*	-0.022***	0.008***
		(1.71)	(-4.07)	(7.36)
N.Algo = 3			-0.023***	
			(-4.20)	
N.Algo = 4			-0.037***	
			(-6.76)	
Rating	0.010*	0.008***	-0.006***	-0.003***
	(1.95)	(4.66)	(-7.85)	(-3.00)
Deliverytime	0.000***	-0.001***	0.001***	0.001***
	(5.46)	(-12.28)	(13.35)	(4.44)
Constant	3.494***	3.267***	3.203***	3.230***
	(78.77)	(209.27)	(472.25)	(386.74)
N	155632	557198	1066364	318770
R2	1.000	1.000	1.000	0.999
ProductxDate FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses. Dependent Variable: Log Bbox Price. Product-clustered SE.
Note: Coefficients on more than four algorithmic sellers are not shown.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6
Fixed effects by date and quartile of average Buy Box price within category.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
N.Algo = 1	-0.113	-0.096	-0.070	0.032
	(-1.42)	(-0.87)	(-1.32)	(0.42)
N.Algo = 2		-1.201***	0.321	0.176
		(-12.41)	(0.75)	(1.12)
Bol	-0.219***	-0.207***	-0.229***	-0.124
	(-3.99)	(-4.64)	(-4.50)	(-1.35)
BBoxRating	-0.348***	-0.128	-0.008	0.187**
	(-3.26)	(-1.45)	(-0.21)	(2.35)
BBoxDeliveryTime	-0.012***	-0.007	-0.027***	-0.038**
	(-2.92)	(-1.52)	(-3.28)	(-2.41)
Constant	6.458***	4.666***	3.613***	1.857***
	(6.94)	(5.94)	(10.10)	(2.62)
N	168116	152134	312259	96637
R2	0.629	0.583	0.551	0.599
Quartile x Date FE	Y	Y	Y	Y
Algo	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7
Product definition by high cosine similarity.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3	(4) Comp = 4	(5) Comp = 5	(6) Comp = 6	(7) Comp = 7	(8) Comp = 8	(9) Comp = 9	(10) Comp = 10
N.Algo = 1	-0.000 (-0.02)	-0.002 (-1.11)	-0.000 (-0.60)	0.002 (1.43)	0.000 (0.48)	-0.001 (-1.11)	0.000 (0.01)	0.000 (0.34)	-0.000 (-0.78)	-0.002* (-1.93)
N.Algo = 2				-0.002 (-1.00)				-0.001 (-0.90)	-0.005 (-0.70)	-0.002 (-0.88)
N	12051	17967	17203	15272	14859	12126	12141	10669	8689	10455
R2	1.000	0.999	1.000	1.000	1.000	0.999	1.000	0.999	0.999	0.999
ProductxDate FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8
Product definition by the most granular product category.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3	(4) Comp = 4	(5) Comp = 5	(6) Comp = 6	(7) Comp = 7	(8) Comp = 8	(9) Comp = 8	(10) Comp = 9	(11) Comp = 10
N.Algo = 1	-0.157*** (-3.14)	0.015 (0.73)	-0.019*** (-3.40)	0.006 (0.72)	-0.024 (-1.12)	-0.006 (-0.92)	-0.006 (-1.38)	0.003 (0.97)	0.005 (0.72)	0.007 (1.05)	-0.000 (-0.47)
N.Algo = 2					-0.024 (-1.12)	-0.021 (-1.27)	-0.063 (-1.19)	0.007 (1.32)	0.004 (0.43)	-0.001 (-0.03)	-0.001 (-0.94)
N.Algo = 3											-0.000 (-0.70)
N.Algo = 4											-0.000 (-0.41)
N.Algo = 5											-0.001 (-1.51)
N.Algo = 6											-0.001 (-0.96)
N.Algo = 7											-0.002 (-1.47)
N.Algo = 8											-0.001 (-0.93)
N.Algo = 9											-0.000 (-0.35)
N.Algo = 10											-0.001 (-0.83)
N	29293	43206	37343	35430	30273	24010	22503	21190	20220	22542	400427
R2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
ProductxDate FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9
Comparison of higher statistical moments of price variation around product mean.

Category	Mean	Var	Skew	Kurt	N	Mean	Var	Skew	Kurt	N
Baby	0.07	8.96	-0.42	11.10	12371	-0.70	13.81	-1.75	10.84	1305
Beverages & Liquor	0.01	5.66	0.21	12.12	22852	-0.02	6.06	-0.24	7.07	12021
Bikes & Accessories	0.00	1.42	-1.01	188.45	173713	0.06	2.28	-14.06	346.27	15275
Books	-0.71	590.31	-0.93	52.02	18356	4.38	311.51	3.01	11.39	2987
Camping & Outdoor	0.01	15.11	-11.22	194.09	18304	-0.09	1.57	-2.76	25.52	1937
Computer	0.00	1.35	-4.62	62.49	3573	0.06	0.11	-5.43	39.00	180
Cooking & Dining	0.27	5.56	2.72	31.82	12636	-0.19	4.07	-9.92	125.83	1834
Electronics	0.22	529.23	10.13	455.51	88480	-1.06	1134.80	-16.98	325.36	18102
Erotica	-0.03	4.94	0.00	8.46	8168	0.04	0.47	-1.23	11.64	6048
Fashion (Kids)	-0.10	0.79	1.36	10.05	3759	0.12	0.44	-1.53	3.51	308
Fashion (Man)	-0.97	7.74	-4.08	51.08	17922	1.48	7.05	0.76	2.80	1168
Fashion (Women)	0.14	47.65	-2.49	40.98	21233	-0.19	27.78	0.89	7.15	1537
Games	0.94	24.60	-1.57	11.70	9604	-0.23	101.38	-15.60	248.88	3898
Garden	0.33	5.88	-0.33	15.11	15974	-0.50	2.49	-0.21	2.41	1045
Healthcare	0.00	1.32	1.97	23.78	2675	0.00	0.20	-0.48	1.30	434
Household	-0.01	12.05	0.24	38.62	14355	0.09	3.36	1.50	6.75	1520
Music	0.35	5.62	5.84	36.33	26049	-0.21	1.68	-2.52	15.61	4342
Office & School	0.02	0.95	-1.56	20.78	11136	-0.61	1.32	1.35	28.71	2904
Personal Care	0.06	18.40	11.91	223.35	17786	-0.32	9.71	-8.53	87.35	3233
Pets	0.08	6.66	0.28	19.29	89946	-0.16	3.32	-0.59	14.34	46507
Sports	0.00	4.61	-0.13	19.38	13584	0.01	1.52	-1.33	3.68	371
Toys	0.00	1.61	0.10	17.47	9558	0.01	1.39	-2.69	8.63	370
Travel & Accessories	-0.06	6.24	-2.17	47.03	41838	0.59	32.24	0.93	10.38	3963
	non-algorithmic markets					*algorithmic markets*				

Table A.10
Minimum price instead of Buy Box price.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
N.Algo = 1	-0.237***	0.004	-0.007	-0.015**
	(-3.62)	(0.17)	(-1.57)	(-2.53)
N.Algo = 2			-0.014	-0.012
			(-1.34)	(-0.80)
Bol	-0.122***	-0.156***	-0.150***	-0.064***
	(-2.96)	(-3.39)	(-7.25)	(-4.12)
BBoxRating	0.071	-0.002	-0.000	0.014***
	(0.53)	(-0.27)	(-0.06)	(2.67)
BBoxDeliveryTime	0.003*	-0.000	-0.005***	-0.001
	(1.85)	(-1.26)	(-2.95)	(-1.23)
Constant	2.648**	3.471***	3.371***	3.227***
	(2.26)	(44.20)	(49.31)	(65.72)
N	166423	155700	314794	97767
R2	1.000	1.000	1.000	1.000
ProductxDate FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11
Excluding products on which Bol is seller.

	(1) Comp = 1	(2) Comp = 2	(3) Comp = 3–5	(4) Comp = 6–8
N.Algo = 1	-0.220**	0.084	-0.007	-0.001
	(-2.57)	(1.14)	(-1.58)	(-0.26)
N.Algo = 2			-0.038***	-0.001
			(-6.70)	(-0.07)
BBoxRating	0.147	0.040**	0.034*	0.048**
	(0.90)	(2.12)	(1.92)	(2.18)
BBoxDeliveryTime	0.010**	-0.006***	-0.020***	-0.017***
	(2.19)	(-3.71)	(-5.40)	(-5.20)
Constant	2.253	3.209***	3.219***	3.196***
	(1.57)	(19.62)	(21.19)	(16.44)
N	59392	59779	101925	32555
R2	1.000	1.000	1.000	1.000
ProductxDate FE	Y	Y	Y	Y
Algo criterion	2 SD	2 SD	2 SD	2 SD

t statistics in parentheses.
Dependent Variable: Log Bbox Price. Product-clustered SE.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

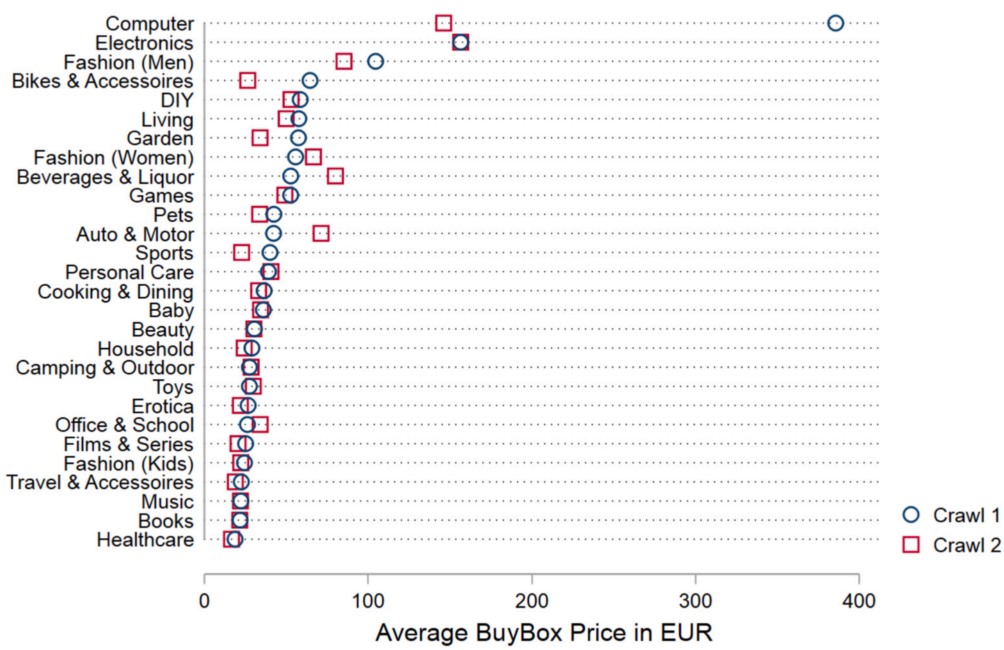


Fig. A.1. Average Buy Box price by category.

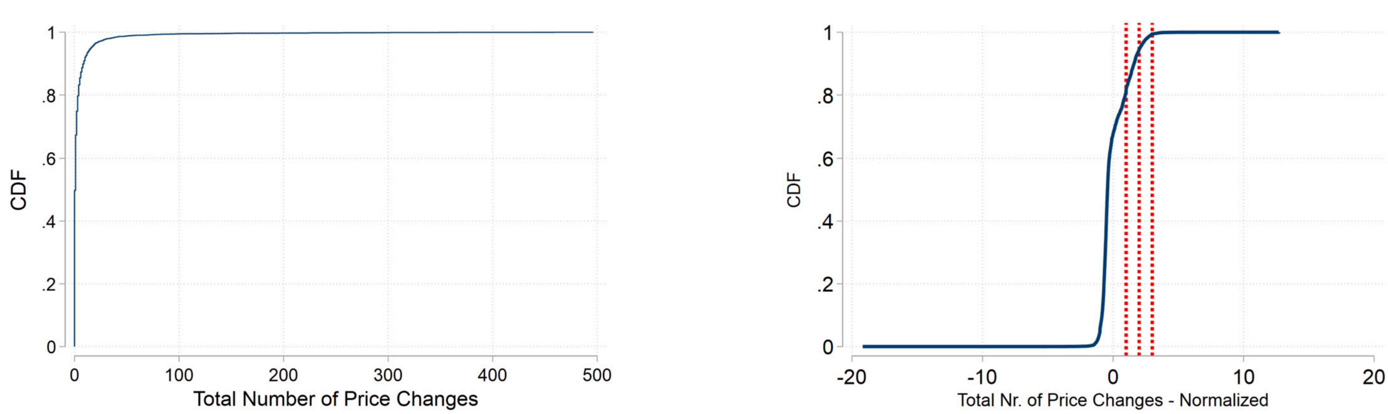


Fig. A.2. CDF of the number of total price changes (left) and normalized on product-level (right, 1–3 standard deviations indicated in red).

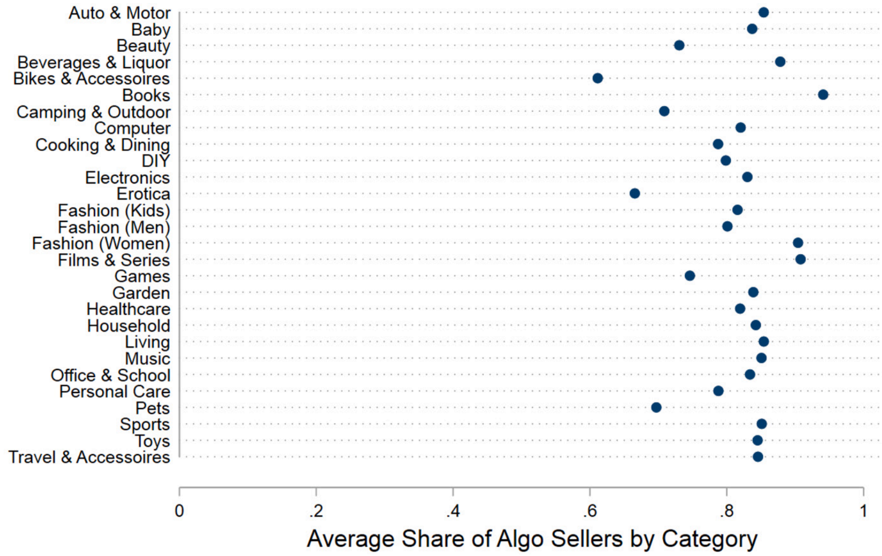


Fig. A.3. Number of algorithmic sellers by Product and Product Category. Product-Seller-pairs are flagged algorithmic if we document more than 20 price changes. Data from Crawl 1.

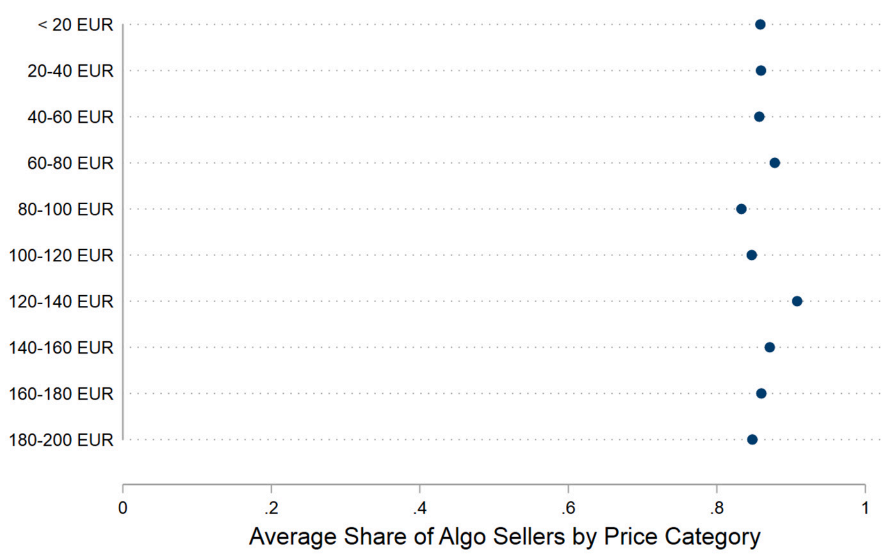


Fig. A.4. Number of algorithmic sellers by Product and Price Category. Product-Seller-pairs are flagged algorithmic if we document more than 20 price changes. Data from Crawl 1.

Fig. A.5. Share of Algorithmic Monopolies by Product Category. Product-Seller-pairs are flagged algorithmic as in the main specification. Data from Crawl 1.

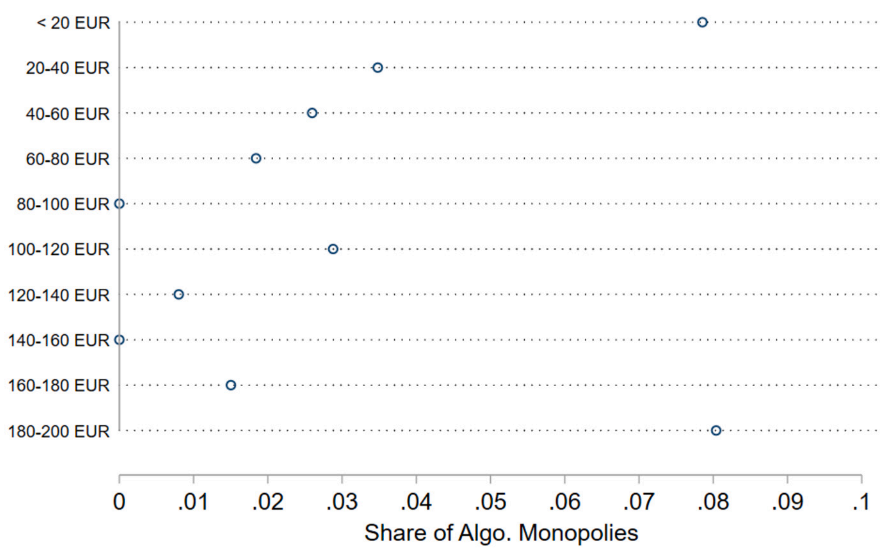


Fig. A.6. Share of Algorithmic Monopolies by Buy Box Price. Product-Seller-pairs are flagged algorithmic as in the main specification. Data from Crawl 1.

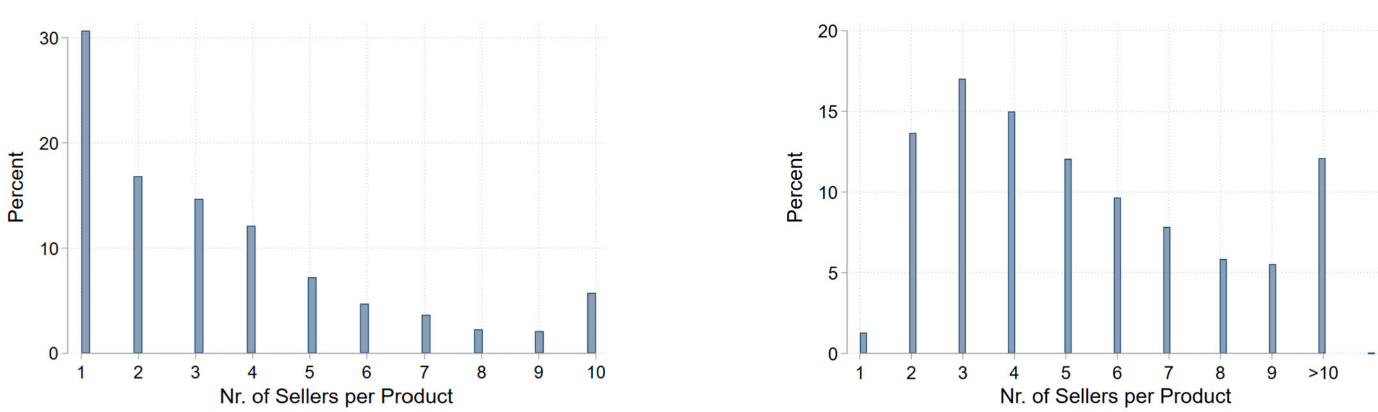


Fig. A.7. Number of sellers per product. Left: Crawl 1, Right: Crawl 2.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.infoecopol.2024.101111.

References

Albæk, S., Møllgaard, P., Overgaard, P.B., 1997. Government-assisted oligopoly coordination? A concrete case. J. Ind. Econ. 45 (4), 429–443.

Albano, G., Buccirossi, P., Spagnolo, G., Zanza, M., 2006. Preventing collusion in procurement: a primer. In: Handbook of Procurement.

Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., Wrobel, S., 2020. Visual Analytics for Data Scientists. Springer.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Aparicio, D., Metzman, Z., Rigobon, R., 2023. The pricing strategies of online grocery retailers. Quant. Mark. Econ., 1–21.

Asker, J., Fershtman, C., Pakes, A., 2021. Artificial intelligence and pricing: the impact of algorithm design. w28535. National Bureau of Economic Research.

Assad, S., Clark, R., Ershov, D., Xu, L., 2020. Algorithmic pricing and competition: empirical evidence from the German retail gasoline market. CESifo Working Paper.

Autoridade da Concorrência, 2019. Digital ecosystems, big data and algorithms. Fair Play Issue Pap.

Berentsen, A., Rojas Breu, M., Waller, C.J., 2019. What is the value of being a superhost? Federal Reserve Bank of St. Louis Working Paper, 2019(19A).

Bernheim, B.D., Whinston, M.D., 1985. Common marketing agency as a device for facilitating collusion. Rand J. Econ., 269–281.

Bigoni, M., Potters, J., Spagnolo, G., 2019. Frequency of interaction, communication and collusion: an experiment. Econ. Theory 68 (4), 827–844.

Bol.com, 2021a. Bol.com website, "the commission". https://partnerplatform.bol.com/nl/hulp-nodig/financien/de-commissie/. (Accessed 16 February 2012).

Bol.com, 2021b. Het koopblok. https://partnerplatform.bol.com/nl/hulp-nodig/aanbod-online-brengen/het-koopblok/. (Accessed 17 February 2021).

Bol.com, 2021c. Allowable retail price. https://developers.bol.com/allowable-retail-price/. (Accessed 17 February 2021).

Borenstein, S., Cameron, A.C., Gilbert, R., 1997. Do gasoline prices respond asymmetrically to crude oil price changes? Q. J. Econ. 112 (1), 305–339.

Bremmer, D.S., Kesselring, R.G., 2016. The relationship between US retail gasoline and crude oil prices during the great recession: "rockets and feathers" or "balloons and rocks" behavior? Energy Econ. 55, 200–210.

Brown, Z., MacKay, A., 2020. Competition in pricing algorithms. Available at SSRN 3485024.

Busse, C., 2017. Bundeskartellamt rügt lufthansa. SZ.

Byrne, D.P., 2012. Petrol price cycles. Aust. Econ. Rev. 45 (4), 497–506.

Byrne, D.P., De Roos, N., 2019. Learning to coordinate: a study in retail gasoline. Am. Econ. Rev. 109 (2), 591–619.

Cabral, L., Fishman, A., 2012. Business as usual: a consumer search theory of sticky prices and asymmetric price adjustment. Int. J. Ind. Organ. 30 (4), 371–376.

Calvano, E., Calzolari, G., Denicolo, V., Pastorello, S., 2020. Artificial intelligence, algorithmic pricing, and collusion. Am. Econ. Rev. 110 (10), 3267–3297.

Capobianco, A., Gonzaga, P., 2020. Competition challenges of big data: algorithmic collusion, personalised pricing and privacy. In: Legal Challenges of Big Data. Edward Elgar Publishing.

Channable, 2019. Configuring the bol.com repricer. https://helpcenter.channable.com/hc/en-us/articles/360016027219-Configuring-the-Bol-com-Repricer. (Accessed 16 February 2021).

ChannelEngine, 2018. How is your performance score? https://www.channelengine.com/marketplace-blog/win-the-buy-box-by-optimizing-performance-on-bol-com/. (Accessed 17 February 2021).

ChannelEngine, 2021. Channelengine: repricing. https://help.channelengine.com/article/47-repricer. (Accessed 17 February 2021).

Chen, L., Mislove, A., Wilson, C., 2016. An empirical analysis of algorithmic pricing on Amazon marketplace. In: Proceedings of the 25th International Conference on World Wide Web, pp. 1339–1349.

Clark, R., Houde, J.-F., 2013. Collusion with asymmetric retailers: evidence from a gasoline price-fixing case. Am. Econ. J. Microecon. 5 (3), 97–123.

Clark, R., Houde, J.-F., 2014. The effect of explicit communication on pricing: evidence from the collapse of a gasoline cartel. J. Ind. Econ. 62 (2), 191–228.

Colombo, N., 2018. Virtual competition: human liability vis-a-vis artificial intelligence's anticompetitive behaviours. Eur. Compet. Regul. Law Rev. 2 (11).

Congiu, R., Sabatino, L., Sapi, G., 2022. The impact of privacy regulation on web traffic: evidence from the gdpr. Inf. Econ. Policy 61, 101003.

Connor, J.M., 2007. Forensic economics: an introduction with special emphasis on price fixing. J. Compet. Law Econ. 4 (1), 31–59.

Decarolis, F., Rovigatti, G., 2019. From mad men to maths men: concentration and buyer power in online advertising. CEPR Discussion Paper No. DP13897.

Douglas, C.C., 2010. Do gasoline prices exhibit asymmetry? Not usually! Energy Econ. 32 (4), 918–925.

Eckert, A., 2013. Empirical studies of gasoline retailing: a guide to the literature. J. Econ. Surv. 27 (1), 140–166.

EcommerceNews, 2018. Third party sellers make up 40% of bol.com's sales. https://ecommercenews.eu/third-party-sellers-make-up-40-of-bol-coms-sales/. (Accessed 17 February 2021).

EffectConnect, 2021. How does the repricer work? https://www.effectconnect.com/en/repricer-tool. (Accessed 16 February 2021).

Eschenbaum, N., Melgren, F., Zahn, P., 2022. Robust algorithmic collusion. arXiv preprint. arXiv:2201.00345.

Ezrachi, A., Stucke, M., 2016a. Virtual Competition. Titolo Collana. Harvard University Press.

Ezrachi, A., Stucke, M.E., 2016b. How pricing bots could form cartels and make things more expensive. Harv. Bus. Rev. 27.

Ezrachi, A., Stucke, M.E., 2017. Artificial intelligence & collusion: when computers inhibit competition. Univ. Ill. Law Rev., 1775.

Feier, T., Gogoll, J., Uhl, M., 2021. Hiding behind machines: when blame is shifted to artificial agents. ArXiv preprint. arXiv:2101.11465.

Feiler, D., Tong, J., 2022. From noise to bias: overconfidence in new product forecasting. Manag. Sci. 68 (6), 4685–4702.

GOV.UK, 2021. Algorithms: how they can reduce competition and harm consumers. https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/. (Accessed 17 December 2021).

Hansen, K., Misra, K., Pai, M., 2020. Algorithmic collusion: supra-competitive prices via independent algorithms. CEPR Discussion Papers. No. 14372.

Harrington, J.E., 2018. Developing competition law for collusion by autonomous artificial agents. J. Compet. Law Econ. 14 (3), 331–363.

Huffman, D., Raymond, C., Shvets, J., 2022. Persistent overconfidence and biased memory: evidence from managers. Am. Econ. Rev. 112 (10), 3141–3175.

IndustryNews, 2018. Don't be a prisoner in Amazon price wars. https://prospershow.com/media/prosper-blog/dont-be-a-prisoner-in-amazon-price-wars/. (Accessed 16 February 2021).

Jiang, B., Jerath, K., Srinivasan, K., 2011. Firm strategies in the "mid tail" of platform-based retailing. Mark. Sci. 30 (5), 757–775.

Johnson, J., Rhodes, A., Wildenbeest, M.R., 2020. Platform design when sellers use pricing algorithms. CEPR Discussion Paper No. DP15504.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York.

Klein, T., 2021a. Autonomous algorithmic collusion: Q-learning under sequential pricing. Rand J. Econ. 52 (3), 538–558.

Klein, T., 2021b. (Mis)understanding algorithmic collusion. https://www.competitionpolicyinternational.com/misunderstanding-algorithmic-collusion. (Accessed 28 November 2021).

Knittel, C.R., Stango, V., 2003. Price ceilings as focal points for tacit collusion: evidence from credit cards. Am. Econ. Rev. 93 (5), 1703–1729.

Konkurransetilsynet.no, 2021. What effects can algorithms have on consumers? https://konkurransetilsynet.no/wp-content/uploads/2021/03/Report-Algorithms-english-version-2021.pdf. (Accessed 17 December 2021).

Kühn, K.-U., Tadelis, S., 2017. Algorithmic collusion. Presentation prepared for CRESSE. http://www.cresse.info/uploadfiles/2017_sps5_pr2.pdf.

Laitenberger, J., 2017. Enforcing EU competition law – recent developments and a glance to the future. In: CMS EU Competition Conference.

Leisten, M., 2021. Algorithmic competition, with humans. Working Paper.

Luco, F., 2019. Who benefits from information disclosure? The case of retail gasoline. Am. Econ. J. Microecon. 11 (2), 277–305.

Malmendier, U., Tate, G., 2015. Behavioral ceos: the role of managerial overconfidence. J. Econ. Perspect. 29 (4), 37–60.

Markovitch, D.G., Steckel, J.H., Michaut, A., Philip, D., Tracy, W.M., 2015. Behavioral reasons for new product failure: does overconfidence induce overforecasts? J. Prod. Innov. Manag. 32 (5), 825–841.

Maskin, E., Tirole, J., 1988. A theory of dynamic oligopoly, II: price competition, kinked demand curves, and Edgeworth cycles. Econometrica J. Econom. Soc., 571–599.

Mehra, S.K., 2015. Antitrust and the robo-seller: competition in the time of algorithms. Minn. Law Rev. 100, 1323.

Miklós-Thal, J., Tucker, C., 2019. Collusion by algorithm: does better demand prediction facilitate coordination between sellers? Manag. Sci. 65 (4), 1552–1561.

Montgomery, A.L., Bradlow, E.T., 1999. Why analyst overconfidence about the functional form of demand models can lead to overpricing. Mark. Sci. 18 (4), 569–583.

Morton, F.M.S., 2012. Contracts that reference rivals. Antitrust 27, 72.

Musolff, L., 2021. Algorithmic pricing facilitates tacit collusion: evidence from e-commerce. Technical report, Working Paper.

Normann, H.-T., Sternberg, M., et al., 2021. Hybrid collusion: algorithmic pricing in human-computer laboratory markets. Technical report. Max Planck Institute for Research on Collective Goods.

O'Connor, J., Wilson, N.E., 2020. Reduced demand uncertainty and the sustainability of collusion: how AI could affect competition. Inf. Econ. Policy, 100882.

RepricerExpress, 2021. How to win the Amazon Buy Box in 2021. https://www.repricerexpress.com/win-amazon-buy-box/. (Accessed 16 February 2021).

Ritter, C., 2017. Bibliography on antitrust and algorithms. Available at SSRN 2982397.

Salcedo, B., 2015. Pricing algorithms and tacit collusion. Manuscript. Pennsylvania State University.

Schwalbe, U., 2018. Algorithms, machine learning, and collusion. J. Compet. Law Econ. 14, 568–607.

Shirato, G., Andrienko, N., Andrienko, G., 2023. Identifying, exploring, and interpreting time series shapes in multivariate time intervals. Vis. Inform. 7 (1), 77–91.

Simon, M., Shrader, R.C., 2012. Entrepreneurial actions and optimistic overconfidence: the role of motivated reasoning in new product introductions. J. Bus. Venturing 27 (3), 291–309.

Statista, 2019. Leading online retailers based on revenue in the Netherlands from 2015 to 2018. https://www.statista.com/statistics/624008/ranking-of-online-retailers-based-on-revenue-in-the-netherlands/. (Accessed 17 February 2021).

Tappata, M., 2009. Rockets and feathers: understanding asymmetric pricing. Rand J. Econ. 40 (4), 673–687.

Werner, T., 2021. Algorithmic and human collusion. Technical report, DICE Discussion Paper.

Yang, H., Ye, L., 2008. Search with learning: understanding asymmetric price adjustments. Rand J. Econ. 39 (2), 547–564.

Zhu, F., Liu, Q., 2018. Competing with complementors: an empirical look at amazon.com. Strateg. Manag. J. 39 (10), 2618–2642.

Zitzewitz, E., 2012. Forensic economics. J. Econ. Lit. 50 (3), 731–769.