They do bandit to get the firms behaving as if they are monopolies. With lots of information, we get them cooperating on higher price otherwise- we get less

ALGORITHMIC COLLUSION: SUPRA-COMPETITIVE PRICES VIA INDEPENDENT ALGORITHMS

KARSTEN T. HANSEN [A], KANISHKA MISRA[B], AND MALLESH M PAI[C]

ABSTRACT: Motivated by their increasing prevalence, we study outcomes when competing sellers use machine learning algorithms to run real-time dynamic price experiments. These algorithms are often misspecified, ignoring the effect of factors outside their control, e.g. competitors' prices. We show that the long-run prices depend on the informational value (or signal to noise ratio) of price experiments: if low, the long-run prices are consistent with the static Nash equilibrium of the corresponding full information setting. However, if high, the long-run prices are supra-competitive—the full information joint-monopoly outcome is possible. We show that this occurs via a novel channel: competitors' algorithms' prices end up running correlated experiments. Therefore, sellers' misspecified models overestimate the own price sensitivity, resulting in higher prices. We discuss the implications on competition policy.

[A] RADY SCHOOL OF MANAGEMENT, UNIVERSITY OF CALIFORNIA, SAN DIEGO. K4HANSEN@UCSD.EDU
[B] RADY SCHOOL OF MANAGEMENT, UNIVERSITY OF CALIFORNIA, SAN DIEGO. KAMISRA@UCSD.EDU
[C] DEPARTMENT OF ECONOMICS, RICE UNIVERSITY. MALLESH.PAI@RICE.EDU
Date: March 10, 2021.
The authors thank Joe Harrington, Jan Svitak and seminar participants at UC Berkeley, UC Los Angeles, UC Riverside, and UC San Diego for helpful comments and discussion. Pai gratefully acknowledges financial support from NSF grant CCF-1763349.

1. Introduction

A defining feature of many of the most profitable companies in the 21st century is *scale*. For example, the song catalog for the music service Spotify contains over 50 million songs.[1] The online retailer Amazon.com sells close to 120 million unique products.[2] The movie streaming service Netflix has over 130 million subscribing customers.[3] This proliferation of products and customers would be impossible to manage if humans did not have help in the form of information technology. Large scale databases and algorithms for decision making are now an essential ingredient of high-tech business management.

The benefits of this revolution are clear. For example, one consequence of this technology has been the creation of the most valuable companies in human history. At the same time consumers benefit from recommendation systems engineered for customized product assortments, ease of purchase, peer reviews, online price comparisons etc. An obvious question to ask is if there are any downsides to algorithmic business management?

In this paper we analyze a novel unforeseen consequence of algorithmic management which has received some attention both in the popular press and research: The possibility that companies may end up colluding *through algorithms*. The context we study is pricing—an area where online retailers increasingly use machine learning algorithms ('White House', 2015). These algorithms set real-time prices for an array of products where the retailer has incomplete demand information. In this setting each algorithm is an automated field experiment that learns about potential profit ("exploration") and set the product's profit-maximizing price ("exploitation"). These algorithms are called multi-armed bandits (where each "arm" is a price). Given the complexity and scale of online retail, field experiments in this context are typically analyzed assuming the focal firm is a monopolist (or - alternatively - best-responding to fixed priced of competitors). Chen et al. (2016) study the best selling product on Amazon.com and estimate only 2.4% of sellers run pricing algorithms that consider competitive prices. Recent academic empirical pricing studies including Cheung et al. (2017); Dubé and Misra (2017); Misra et al. (2019) assume no competitive response to pricing decisions.[4] Several theoretical investigations into the use of dynamic pricing algorithms also assume the firm is a monopoly, e.g. Broder and Rusmevichientong (2012); Handel and Misra (2015); Ban and Keskin (2020).

This assumption greatly reduces complexity: it simplifies the problem to that of a single choice variable (own price), and estimating a profit curve which is only a function of the firm's own price. A consequence of this simplicity is that theoretically optimal algorithms

[1]https://newsroom.spotify.com/company-info/

[2]https://www.scrapehero.com/number-of-products-on-amazon-april-2019/

[3]https://www.cnn.com/2019/01/17/media/netflix-earnings-q4/index.html

[4]See section 4 of den Boer (2015) for an more complete overview of dynamic monopoly pricing with unknown demand, we discuss this literature in more detail in the online appendix.

include index algorithms (Gittins, 1989; Weber, 1992; Auer et al., 2002). These algorithms calculate an index for every arm based on the past performance of the arm (mean, variance and number of times played). In each round, the algorithm simply selects the arm with the highest current index.

In this paper we analyze the outcome if competing firms all use an independent multi-armed bandit indexing algorithm outlined above—and act as if they are in a monopolistic market—but in reality price into an oligopolistic market. In keeping with the bandit analogy, we call pricing in each period an "experiment." In particular, we study what long run prices would result in such a setting. Of course, since a seller's demand (and therefore profit) at a particular price depends on competitors' prices, each seller's implicit statistical model is misspecified.

We show that the long-run prices that result depend on the informational value (or signal to noise ratio) of the underlying pricing experiments. In markets where price experiments have low information value, the resulting long-run prices are statistically indistinguishable from Nash Equilibrium prices, and the misspecified models achieve nearly the first best profits. However, in markets where price experiments have high information value, market prices are supra-competitive. We show that more informative pricing experiments result in correlated price experiments across firms. Competitive prices, therefore, become correlated unobservables in each firms' pricing algorithm. By not accounting for the existing competitive pricing, each firm's price sensitivity will have an upward bias, resulting in supra-competitive prices.

A growing body of literature has raised concerns that algorithms might induce collusive pricing behavior — see for example the white paper (OECD, 2017) listing policy-makers' concerns, and the review article of Harrington (2018) on how competition law should adapt. Recent influential papers such as Calvano et al. (2019, 2020) show that algorithmic collusion can occur in settings where firms observe each others prices. We contribute to this literature by identifying a novel channel/mechanism by which such supra-competitive pricing may occur. By contrast, in our setting, each algorithm/firm do not observe competitors' prices, a requirement in the channels identified in the extant literature. We show that collusion can materialize even in this case.

In the Operations literature, our result is most closely related to Cooper et al. (2015) who also show non-Nash outcomes as the limit prices if duopoly sellers set prices from a mis-specifed monopoly model. The main difference between the papers is that our algorithm considers optimal experimentation balancing the benefits from learning and earning (reinforcement learning), while Cooper et al. (2015) do not explicitly model experimentation (assumed exogenous) and instead consider an estimation-optimization (certainty equivalence) algorithm where prices are set as the static optimal prices from a linear OLS demand model.[5] In terms of implications for policy, Cooper et al. (2015) suggests that the possibility of collusive prices depends on which parameters of a linear demand model are commonly known by both firms (this cannot be empirically tested). By contrast, our results suggest that it hinges on the underlying signal-to-noise ratio in the demand function, i.e. the stochasticity of underlying demand (which *is* empirically testable). We provide a more detailed discussion of the connection to this paper and other prior research (on Algorithmic Collusion, Dynamic Pricing, Learning in Games, Behavioral Game Theory, and Algorithmic Bias) in the Online Appendix.

We believe the identification of this novel channel is useful not only for theory, but also raises fresh practical concerns for managers and policy makers. In our setting firms independently choose algorithms which use misspecified models of the underlying demand system (firms' algorithms assume they are effectively monopolies, although the true market is oligopolistic). Each firm use these algorithms to structure 'relevant learning' (Aghion et al., 1991), or learning about profit at the most profit-relevant prices. Our results show that when multiple firms employ identical algorithms, then learning can be focused away from competitive prices resulting in supra-competitive prices. Jointly, firms may have an incentive to be "willfully misspecified". The question then, is whether all firms choosing such misspecified models constitutes a concern for competition policy.

The remainder of this paper is organized as follows: Section 2 describes the general set-up and assumptions. Section 3 outlines our main results via a set of simulations. Section 4 contains our theoretical results (in a stylized setting) and lays out the mechanism by which supra-competitive prices result. In Section 5 we re-analyze Amazon.com pricing data from Chen et al. (2016) and argue that the observed correlations in prices are consistent with our simulations. Section 6 concludes and discusses avenues for future research.

2. COMPETITIVE BANDITS SETUP

In this section, we explicitly state the key demand and supply assumptions in our analysis. These assumptions define both the objective functions and the informational assumptions for the agents. We consider two symmetric single-product firms with a constant marginal cost (set to zero). Firms compete by setting prices in each period.

We assume that each firm's demand is static and stable, a common assumption in empirical and theoretical work and field experiments, see e.g. Besbes and Zeevi (2009), Dubé

[5]An implication is that even if the seller is a monopolist, the limit price need not converge to the correct optimal price if the assumed linear demand is incorrect. The UCB algorithm used in our paper considers (undiscounted) dynamic profits without making parametric demand assumptions.

and Misra (2017), Misra et al. (2019), Calvano et al. (2019). Static demand rules out a large number of potential consumer dynamics, such as preference learning, uncertainty, strategic consumers and stockpiling. These assumptions are important in our setting as we focus on the firms' learning about the demand (equivalently, profit) curve by price experimentation. If demand was not static, then the pricing field experiment would be biased by assumption.

We assume that demand is initially unknown to the firms. We assume there is a discrete set of potential prices $P = \{p_1, p_2, \ldots p_K\}$ that each firm can set. The formal results are for $K = 2$, but we show that similar results obtain for arbitrary values of K in simulations. A period of the algorithm can be considered as a fixed interval of time (e.g., 15 minutes), as is common in the operations research literature (Besbes and Zeevi, 2009), or a fixed number of consumers (e.g., 100 consumers), as is common in the Marketing literature (Misra et al., 2019). Upon setting a price (running an experiment), the firm observes profit (with sampling error) in that period, but does not observe the rival firm's price and quantity sold. An equivalent assumption is that the firm observes competitors' price/quantity but ignores these in its own statistical model.

In notation: In period t, each firm j sets a price $p_{j,t} \in P$. The firm observes a resulting profit $\pi_{j,t} \in \mathbb{R}_+$. The additional complication is that in our setting firms do not observe (or, ignore the effect of) competitors' prices. That is to say that each firm's statistical model is $\pi_{j,t} = \pi^*(p_{j,t}) + \varepsilon_{j,t}$, where $\pi^*()$ is the true stable expected profit and $\varepsilon_{j,t}$ is a small sample error or noise from the experiment at time t. However, in truth, $\pi_{j,t} = \pi^*(p_{j,t}, p_{-j,t}) + \varepsilon_{j,t}$. This means that firms' models are misspecified: they attribute the effect of competitors' prices ($p_{-j,t}$) on own profits to be arising from noise from nature.

Viewing each possible price as an arm in a multi-armed bandit (MAB) problem, we suppose that each firm runs an MAB algorithm to balance learning for future profits and current profits. The objective function of the algorithm is to minimize (undiscounted) dynamic *statistical regret*, defined as the difference between average profits achieved with the algorithm and the ex-post optimal profits. The literature provides theoretical analysis and mathematical guarantees of these algorithms.[6]

The bandit algorithm we will use is the UCB algorithm (Upper Confidence Bound), originally due to Auer (2002). This provides an asymptotically optimal non-parametric MAB solution (Agrawal, 1995; Auer et al., 2002): that is to say, (1) it is guaranteed to find the optimal arm without any further assumptions in any MAB problem, and (2) no other algorithm achieves lower regret. In our context, this means that a monopoly seller using a UCB algorithm for pricing is guaranteed to find the monopoly optimal price without any parametric assumptions on the relationship between prices and profits. The idea of

[6]For an overview, see Sutton and Barto (1998). For applications in similar settings, see Hauser et al. (2009) for a marketing application, and Misra et al. (2019) for pricing.

this kind of algorithm is to maintain an index for each arm. The basic algorithm, for each arm p_k, at each period t tracks the empirical average of the profit from that arm far $\overline{\pi}_{k,t}$, and the number of times that arm has been pulled $n_{k,t}$. The index of arm k at period t, $I_{k,t}$ is defined as $\overline{\pi}_{k,t} + \sqrt{\frac{2\ln t}{n_{k,t}}}$ (the index of an arm that has never been pulled is defined as ∞). In every period, the algorithm experiments with the price with the current highest index, randomizing if there is a tie.

Viewing the profit draws from a given arm as i.i.d. draws from a distribution of unknown mean, this index tracks the upper-bound of a confidence interval of this unknown mean. A simple concentration argument tells us that, in period t, the true expected mean reward of any arm k is lower than the current index $I_{k,t}$ with probability $(1 - \frac{1}{t})$.

A price is thus experimented if either the empirical average of past profits from that price is high (exploitation), or $n_{k,t}$ is low relative to t, i.e., that price has not been used sufficiently often to be confident about its reward (exploration).

3. Simulation Results

In this section we describe the details of our simulations to test implications of independent competing firms running pricing algorithm.

3.1. *Simulation settings*

For our main results we consider the algorithm called UCB-tuned.[7] For each price (p_k) in period t we calculate the index is defined below, and select the price with the highest index:

$$\mathrm{V}_{k,t} = \overline{\pi^2_{k,t}} - \bar{\pi}^2_{k,t} + \sqrt{\frac{2\log t}{n_{k,t}}},$$

$$\text{UCB-tuned}_{k,t} = \bar{\pi}_{k,t} + \sqrt{\frac{\log t}{n_{k,t}} \min\left(\frac{1}{4}, \mathrm{V}_{k,t}\right)},$$

where $n_{k,t}$ is the number of periods where the firm has charged price p_k up to period t and $\bar{\pi}_{k,t}$ is the empirical mean of profits in those periods. In this algorithm, $V_{k,t}$ can be interpreted as the empirical variance plus an exploration bonus that depends on (decreases with) the number of times price p_k has been experimented with. Further, we account for the UCB improvement (Auer and Ortner, 2010) and we allow for arm elimination.

[7] In the appendix section A.2, we show our results are robust to considering the untuned version (Auer et al., 2002) called UCB1. Note that the tuned version is shown to have better empirical performance—for example, see Auer et al. (2002) (general) or Misra et al. (2019) (application to demand learning). We also show that our results are robust to other index algorithms, namely the Gittins index (Gittins, 1989; Brezzi and Lai, 2002).

We eliminate an arm k if the upper confidence bound of the arm is lower than the mean minus exploration bonus (i.e. the "lower confidence bound") of another arm.

We consider a parametric data generating process and allow firms to run independent UCB algorithms. To be explicit, the DGP is known to us (the researcher), but not to the firms: the firms run UCB which is purely non-parametric. The UCB algorithm considers (undiscounted) finite time regret – this requires a pre-specification of the number of periods the algorithm will be run.[8] In all our simulations we consider the algorithms running for 2 million periods and analyze the outcomes of the last 1,000 periods as the "long run".[9]

In the main simulation, we assume a linear demand model:

$$d_j^*(p_j, p_{-j}) = \alpha - \beta p_j + \gamma p_{-j} \tag{1}$$

Under this assumption, the competitive and joint-monopoly (i.e., collusive) prices can be analytically computed. They are:

$$\text{Competitive: } p^D = \frac{\alpha}{2\beta - \gamma}$$
$$\text{Collusive: } p^M = \frac{\alpha}{2(\beta - \gamma)}$$

In any experiment (t) a firm observes a noisy estimate of the true profit, i.e.,

$$\pi_t(p_j, p_{-j}) = p_j d_j^*(p_j, p_{-j}) + \varepsilon_{j,t},$$

but their models are misspecified so they think they are observing

$$\pi_t(p_j) = p_j d_j^*(p_j) + \varepsilon_{j,t}.$$

In our simulations, we set the value of $\alpha = 0.48$, $\beta = 0.9$, $\gamma = 0.6$. With these value the duopoly and monopoly prices are $P^D = 0.4$ and $P^M = 0.8$. The firms decide between 91 discrete prices between \$0.10 to \$1.00, i.e. $P = \{0.10, 0.11, \ldots, 0.99, 1.00\}$.

As we previewed earlier, our results suggest that the long-run distribution of prices depend on the informativeness of the price experiments, i.e. the magnitude of the noise. To that end, we vary the distributional assumptions about the error term, $\varepsilon_{j,t}$. In all simulations we will assume that $\varepsilon_{j,t} \sim U[-\frac{1}{\delta}, \frac{1}{\delta}]$, where we vary the value of δ across simulations.

[8] In a setting with infinite time, there is no trade-off between learning and earning as once found the best arm will be played for an infinite time.

[9] To check robustness, in the appendix section A.2 we present results when the number of periods is increased to 10 million. Our results are statistically indistinguishable from those presented in the main results.

If δ is small (large) then the experiment is uninformative (informative). δ signifies the signal to noise ratio (SNR) in an experiment (by construction the highest signal is 1.0). In our simulation experiments, we vary SNR between $\frac{1}{10}$ (large noise), and 10 (large signal).

3.2. *Main Results*

Our main results are shown in figure 1. The dark grey bars represent a setting with two competing single product firms, and the light grey bars represent a setting with a monopolist jointly maximizing profits for both products. The top panel considers the long run prices from our simulations. For all levels of SNR the long-run price when the firm is a monopoly is not statistically distinguishable from the true monopoly price of 0.8. With a small SNR the long-run prices in a duopoly are not statistically different from the Nash equilibrium price, 0.4. However, as the SNR increases, we find that the estimated duopoly price increases to supra-competitive levels. When the SNR levels are large, the estimated duopoly prices are statistically different from the Nash equilibrium, and indeed, statistically indistinguishable from the full-information joint monopoly (collusive) outcome. This suggests that in markets where pricing experiments are very (less) informative, independent algorithms will result in long-run prices that are supra-competitive (competitive). In appendix section A.2 we show that this result is robust to a different demand system (logit demand) and a different index algorithm (Gittins index). Our main findings continue to hold in both these simulations: markets with low (high) SNR, independent bandits result in competitive (supra-competitive) prices.

[Figure 1 about here.]

To understand the mechanics that generate supra-competitive duopoly prices we consider the joint distribution of long-run prices. Figure 2 displays the distribution of long-run duopoly prices by SNR. Our point of emphasis here is the shape of the distribution, and how this changes with SNR. For small SNR values, the shape of this distribution implies independent prices. However, as the SNR increase, the distribution indicates correlated prices across the two firms. [10] In section 4, we explain both the origins of this correlation, and how it impacts long-run prices.

[Figure 2 about here.]

This suggests that alternative pricing algorithms that force no (force large) correlation in prices should result in competitive (supra-competitive) prices across all SNR settings. The long run prices for such alternative algorithms are shown in figure 3. The top panel

[10]Numerically, we can see this as the decrease in the median difference between the two firms prices (top left of the chart) – this decreases from 0.15 to 0.00. Alternatively, we consider the percentage of simulations in which resultant prices are within 1 cent of each other (bottom left of chart). Again we see this number increase from 3% in the low SNR setting to 69% in the high SNR setting. This suggests that the signal strength of experiments is critical to coordinating prices across different algorithms.

of this figure corresponds to the market outcomes when both firms induce random price experimentation with the heuristic-based ε-greedy algorithm. In this algorithm at period t, with probability ε (we set ε to be 0.01) the firm sets a randomly selected price, otherwise, the firm sets the price with the highest mean profits from past experiments. Consistent with this intuition, we find the long-run prices are indistinguishable from competitive Nash Equilibrium prices for all levels of SNR. The bottom panel of this figure corresponds to the market outcomes where we force a correlation of 1 between the firms' prices. Here, we consider a setting where one firm uses the UCB algorithm for pricing (as in our main results) and the other firm price matches in real time. The competitive outcomes are indistinguishable from the monopoly outcomes for all levels of SNR.

[Figure 3 about here.]

4. Theoretical Foundations

In this section, we explain the theoretical basis for the results above. In the online appendix section B we explain the mechanism by a simple intuition for the implication of correlated prices. This is incomplete as the correlation of prices is endogenously generated by the algorithm. We provide correct theoretical foundations in a setting where each firm chooses between two prices, i.e. $K = 2$.

We prove our result in a model where each firm only chooses between two prices, $p_H > p_L$. A firm's observed profit when it charges price p_x and its competitor charges p_y, for $x, y \in \{H, L\}$, is $\pi_{xy} + \varepsilon$. Here, ε is a mean-0 shock, independently drawn across periods. The pricing structure therefore implies $\pi_{LH} > \pi_{HH}$ and $\pi_{LL} > \pi_{HL}$, i.e., charging the low price is a dominant strategy for each firm. However, both firms charging the high price results in higher joint profits, i.e. $\pi_{HH} > \pi_{LL}$. Putting this together, we have that the true game for both firms is essentially a prisoner's dilemma.

Both firms do not know the true π, indeed, their statistical models are misspecified in that they assume the profit when they charge price p_x is $\pi_x + \varepsilon$. They both run UCB algorithms, so from each agent j's perspective, in any period t, the relevant part of past play can be summarized by four numbers: for each price p_x, the empirical average of profits in past periods where p_x was charged, $\overline{\pi}_{x,t}$, and the number of times it was charged, $n_{x,t}$. Note that $n_{H,t} + n_{L,t} = (t-1)$ by definition. The UCB algorithm calculates an index for each price $x \in \{H, L\}$ as

$$I_{x,t} = \overline{\pi}_{x,t} + \sqrt{\frac{2\ln t}{n_{x,t}}},$$

and always pulls the arm with the highest index. By definition, the index of an arm that has never been pulled is ∞, i.e. every arm is pulled at least once before any arm is pulled twice. It will be clear from the proofs that follow that similar results obtain for other "index" algorithms, where indices are deterministic functions of $\overline{\pi}$ and n.

We study this system analytically for the extreme case where the noise in the demand system vanishes, i.e, demand in each period is a deterministic function of prices. Our results for each of these cases mirror correlated-past-price intuition in Section 3. Informally, when the noise in the demand vanishes, we show that price paths end up correlated (even though the algorithms are independent) and prices converge to the (p_H, p_H) collusive outcome. Formally:

THEOREM 1. *Suppose the true demand function is deterministic and both firms use independent UCB alogrithms. Then,*

(1) *the prices are always exactly correlated from the third period onward (i.e. in any period $t \geq 3$, either both firms charge p_H or both firms charge p_L), and*
(2) *the fraction of times in the first t periods that either seller charges p_L converges to 0 as t grows large.*

Proof provided in appendix section A.1.

Our result is shown for the specific case two possible prices, deterministic demand, and when both sellers use the UCB algorithm. These features make the system feasible to study analytically. However, the mechanics of this proof, and the intuition provided illustrate how similar phenomena should obtain more generally (and indeed, why such results obtain in the simulations we conducted previously). Independent algorithms can end up having correlated price paths. Since the algorithms are misspecified, this results in an omitted variable bias (the omitted variable being the competitor's price), and an overestimate of own price sensitivity. The absence of sufficiently large demand shocks to force independent experimentation then results in this being self-reinforcing. Both sellers's algorithms then settle on high prices, even though this is neither the equilibrium of the underlying game, or indeed, even the best response given the competitor's strategy. When demand shocks are small and the sellers' algorithms are close to deterministic, such mis-learning may occur and the competing firms may settle down on charging collusive prices. This could occur forever (as in the Theorem above), or with small amounts of noise, may occur for arbitrarily long periods or with very high probability[11].

So when does such coordination fail? Demand has to be stochastic enough to prevent both firms from settling in to the correlated price paths displayed above. It should

[11]Consistent with this intuition in the online appendix we show a simulation where we allow demand shocks to increase or decrease by round and show our results are robust further suggesting that importance of initial noise to make the algorithms uncorrelated.

be clear that with highly stochastic demand, the early stages of experimentation under UCB will be close to independent/ stochastic for many periods. Since charging p_L is a dominant strategy in the underlying pricing game, both agents will learn this and their associated index of p_L will be higher, and their historical average payoff when playing p_L will converge to π_{LL} by standard concentration inequalities. Subsequently, the stochastic differences in their estimates of the payoff under p_H will ensure that both players are unlikely to switch to playing p_H at the same time. As a result, the historical average when playing p_H will fall to $\pi_{HL} < \pi_{LL}$, and therefore the path of play will feature both players playing the low price almost always in the limit. How much noise/ stochasticity is needed in demand to ensure this happens with high probability is a subject left for future research.

5. Empirical Relevance

To provide empirical relevance of our results, we re-analyze pricing data from Chen et al. (2016). These data include a price tracker for best-selling products on Amazon.com in two web crawls from September 15, 2014 to December 8, 2014 and from August 11, 2015 and September 21, 2015. In each web-crawl individual sellers' prices and ratings were collected every 25 minutes. For each product, we define the top two sellers as sellers the highest listings (defined by page and rank within the page) with Prime shipping (to remove variation in shipping prices and times). We consider 830 products with variations in prices for both sellers.

We consider two measures of interest: (1) pairwise correlations in prices for the top two sellers over time and (2) the relative prices of the top two sellers versus all other sellers. For each product we define the relative price as the mean price for the top two sellers minus the mean price for all other sellers. The results are shown in top row of figure 4. The top left chart plots the distribution of pairwise correlations and shows that the distribution is bi-modal with modes near 0 and 1. This is consistent with our simulation settings, where correlations were either near 0 (small SNR) or near 1 (large SNR)[12]. The top right chart plots the distribution of relative prices. The median relative price in the data is $-\$1.04$ (the top two sellers price \$1.04 below the average price of all other sellers), however there is large cross-sectional variation across products.

To investigate the cross-sectional variation across products, we consider a proxy for the variance in demand shocks (i.e., the magnitude of δ in our model). For this we consider the variance in number of new product reviews per day for each product. Intuitively,

[12]We investigate if high correlations due to sellers identified as 'algorithmic competitive sellers' in Chen et al. (2016). We find that seller pairs with no identified sellers have a higher median correlation (0.54), than seller pairs with one (0.22) or two (0.14) identified sellers. This suggests that the higher correlations are not due to algorithmic competitive sellers.

if agents leave reviews at roughly the same rate, then variance in rate of new reviews equates to variance in demand. The charts in the bottom row of figure 4 plot the CDF (cumulative distribution) for the products with the lowest demand variation (lowest 25%) versus all other products. The bottom left chart shows the CDF of pairwise correlations and the bottom right chart shows the CDF for relative prices. We observe that the CDF for products with the lowest demand variation is shifted to the right for both charts (statistically significant for relative prices). This suggests that the top two sellers for products with the lowest demand variation have (a) a higher correlation in prices over time and (b) higher relative prices relative to other sellers.

Both these trends in the Amazon.com data are consistent with our results where we find that markets with the higher SNR (lower noise) have the higher correlations in prices and higher levels of prices. Of course, since we do not observe (a) the algorithms used by sellers on Amazon, (b) demand for each product, (c) supply conditions (marginal costs), and (d) the Chen et al. (2016) dataset has a non-random selection of products, we do not assert that these correlations and prices are uniquely due to the mechanism described in our paper.

[Figure 4 about here.]

6. Conclusion

With the growth of e-commerce, we have seen increased usage of algorithms automating pricing decisions. The pricing algorithm runs automated field experiments to learn about the demand curve and each product's profit-maximizing price. Given the complexity and scale of online markets, field experiments typically assume firms are monopolists or oligopolists best-responding to fixed competitors' prices. While this assumption greatly reduces complexity, it also results in a misspecifed pricing algorithm. In this paper, we study outcomes in an oligopoly setting where all competing sellers independently use such misspecified algorithms for pricing.

We show that the long-run prices that result depend critically on the informational value (signal to noise ratio) of pricing experiments. If low, the long-run price are competitive and misspecified algorithms achieve nearly first best profits. However, if high, the long-run prices are supra-competitive. We show this occurs via a novel channel: competitors' algorithms' prices end up being correlated through the experiments. Therefore, sellers' misspecified models overestimate own price sensitivity, resulting in higher prices. We believe the identification of this novel channel raises important new concerns for competition policy.

In terms of future research, while we revealed a new channel by which collusive-seeming pricing is possible, there is much work needed to understand how robust this finding is. For instance, our theoretical results investigates a stylized model where symmetric firms simultaneously price using the UCB algorithm, and there is no noise. What if firms set prices asynchronously (as in Brown and MacKay (2020)) or demand is not stationary (as in Keskin and Zeevi (2017) or with price discrimination (e.g. Dubé and Misra (2017))? How much noise is necessary to "break" the collusion? What are properties of algorithms other than UCB that can sustain such outcomes? Our simulations provide insight and intuition, but a full theoretical analysis would be desirable. Relatedly, while we provide indicative evidence, it would be interesting to see if empirical work could robustly show the existence of such pricing patterns in the real world.

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation," *The Review of Economic Studies*, 58, 621–654.

Agrawal, R. (1995): "Sample Mean Based Index Policies with O(log n) Regret for the Multi-Armed Bandit Problem," *Advances in Applied Probability*, 27, 1054–1078.

Auer, P. (2002): "Using Confidence Bounds for Exploitation-Exploration Trade-offs," *Journal of Machine Learning Research*, 397–422.

Auer, P., N. Cesa-Bianchi, and P. Fischer (2002): "Finite-time Analysis of the Multiarmed Bandit Problem," *Machine Learning*, 47, 235–256.

Auer, P. and R. Ortner (2010): "UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem," *Periodica Mathematica Hungarica*, 61, 55–65.

Ban, G.-Y. and N. B. Keskin (2020): "Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity," *Available at SSRN 2972985*.

Besbes, O. and A. Zeevi (2009): "Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms," *Operations Research*, 57, 1407–1420.

Brezzi, M. and T. L. Lai (2002): "Optimal Learning and Experimentation in Bandit Problems," *Journal of Economic Dynamics and Control*, 27, 87–108.

Broder, J. and P. Rusmevichientong (2012): "Dynamic pricing under a general parametric choice model," *Operations Research*, 60, 965–980.

Brown, Z. and A. MacKay (2020): "Competition in Pricing Algorithms," *Available at SSRN 3485024*.

Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello (2019): "Algorithmic Pricing What Implications for Competition Policy?" *Review of Industrial Organization*, 55, 155–171.

——— (2020): "Artificial intelligence, algorithmic pricing and collusion," *American Economic Review*, Forthcoming.

Chen, L., A. Mislove, and C. Wilson (2016): "An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace," in *Proceedings of the 25th International World Wide Web Conference (WWW 2016)*, Montreal, Canada.

Cheung, W. C., D. Simchi-Levi, and H. Wang (2017): "Dynamic pricing and demand learning with limited price experimentation," *Operations Research*, 65, 1722–1731.

Cooper, W. L., T. Homem-de Mello, and A. J. Kleywegt (2015): "Learning and pricing with models that do not explicitly incorporate competition," *Operations research*, 63, 86–103.

den Boer, A. V. (2015): "Dynamic pricing and learning: historical origins, current research, and new directions," *Surveys in operations research and management science*, 20, 1–18.

DUBÉ, J.-P. AND S. MISRA (2017): "Scalable Price Targeting," Working Paper 23775, National Bureau of Economic Research.

GITTINS, J. C. (1979): "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society, Series B*, 41, 148–177.

——— (1989): *Multi-Armed Bandit Allocation Indices*, Chichester, UK: John Wiley and Sons, 1 ed.

HANDEL, B. AND K. MISRA (2015): "Robust new product pricing," *Marketing Science*, 34, 864–881.

HARRINGTON, J. E. (2018): "Developing Competition Law for Collusion by Autonomous Artificial Agents," *Journal of Competition Law & Economics*, 14, 331–363.

HAUSER, J. R., G. L. URBAN, G. LIBERALI, AND M. BRAUN (2009): "Website Morphing," *Marketing Science*, 28, 202–223.

KESKIN, N. B. AND A. ZEEVI (2017): "Chasing demand: Learning and earning in a changing environment," *Mathematics of Operations Research*, 42, 277–307.

MISRA, K., E. M. SCHWARTZ, AND J. ABERNETHY (2019): "Dynamic Online Pricing with Incomplete Information Using Multiarmed Bandit Experiments," *Marketing Science*.

OECD (2017): "Algorithms and Collusion: Competition Policy in the Digital Age," Tech. rep.

SUTTON, R. S. AND A. G. BARTO (1998): *Reinforcement Learning: An Introduction*, Cambridge, MA: MIT Press.

WEBER, R. (1992): "On the Gittins Index for Multiarmed Bandits," *The Annals of Applied Probability*, 2, 1024–1033.

'WHITE HOUSE', C. O. E. A. (2015): "Big Data and Differential Pricing," February.

APPENDIX A. APPENDIX

A.1. *Proof of Theorem 1*

PROOF. We simulate the evolution of the UCB algorithm. By the definition of UCB, each firm will, in the first two periods, experiment exactly once with each of the two prices p_H and p_L. There are now exactly two possible classes histories at the end of the first two rounds:

(1) Matched prices, i.e. (p_H, p_H) in one round, (p_L, p_L) in the other.
(2) Mismatched price, i.e., (p_L, p_H) and (p_H, p_L) in the other.

We consider each of these in turn.

Case 1: Matched Prices Note that at the end of the first 2 periods, the "state"[13] of each firm's algorithm can be summarized as: $(\pi_{HH}, 1, \pi_{LL}, 1)$. To see part (1) of Theorem 1, note that the algorithms and demand system are deterministic, and both firms' algorithms share a common state at the end of period 2. Therefore the firms will take the same action in period 3, and then inductively, in every subsequent period.

Now $\pi_{HH} > \pi_{LL}$, both firms will charge the high price as long as.

$$\pi_{HH} + \sqrt{\frac{2\ln t}{t-2}} > \pi_{LL} + \sqrt{2\ln t}$$
$$Or, \pi_{HH} - \pi_{LL} > \sqrt{2\ln t}\left(1 - \frac{1}{\sqrt{t-2}}\right)$$

Note that for some t large enough this inequality will be violated (as the right hand side is strictly increasing in t and the left hand side is constant), however, since demand is deterministic, it will be violated at the same time for both firms. At such t, both firms will switch to charging p_L. In period $t+1$, the state of both firms' algorithms will therefore be $(\pi_{HH}, (t-2), \pi_{LL}, 2)$. However, since (1) p_H was chosen in t-1 and (2) the exploration bonus decreases with the number of attempts,[14] we must have

$$\pi_{HH} - \pi_{LL} > \sqrt{2\ln(t+1)}\left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{t-2}}\right).$$

In other words, both firms will immediately switch back to charging p_H. Further, both firms will charge p_L for the n^{th} time in period t_n s.t.

$$\pi_{HH} - \pi_{LL} < \sqrt{\frac{2\ln(t_n)}{n-1}} - \sqrt{\frac{2\ln(t_n)}{t_n - n}}.$$

---

[13]The state variables are (1) average profits for price H, (2) the number of times price H has been charged, (3) average profits for price L, and (4) the number of times price L has been charged

[14]Formally, (1) $\pi_{HH} - \pi_{LL} > \sqrt{2\ln(t-1)}\left(1 - \frac{1}{\sqrt{t-3}}\right)$, and (2) $\frac{\partial\left(\frac{1}{\sqrt{x}} - \frac{K}{\sqrt{x+1}}\right)}{\partial x} < 0, \forall K < 1$.

Therefore n cannot be larger than $O(\ln t_n)$ as it grows large. As a result $\frac{n}{t_n}$ goes to 0 as n grows large.

Case 2: Mismatched Prices As in Case 1, at the end of the first two periods, the two firms share a common state, $(\pi_{HL}, 1, \pi_{LH}, 1)$. As a result both firms take the same action in period 3, and then inductively in every subsequent period. This shows (1) of Theorem 1.

Consider any subsequent period where both firms have charged the high price n_H times and the low price n_L times. The corresponding indices in period $t = n_H + n_L + 1$ are:

$$I_{H,t} = \left(\frac{1}{n_H}\pi_{HL} + \frac{n_H - 1}{n_H}\pi_{HH}\right) + \sqrt{\frac{2\ln t}{n_H}}$$
$$I_{L,t} = \left(\frac{1}{n_L}\pi_{LH} + \frac{n_L - 1}{n_L}\pi_{LL}\right) + \sqrt{\frac{2\ln t}{n_L}}$$

Observe that the first term of the first equality is increasing to π_{HH} as n_H grows large, while the first term of the second is decreasing to π_{LL}. Part (2) of Theorem 1 now follows from an analogous argument to that above. ■

A.2. *Robustness*

Figure 5 shows robustness to the specific algorithm in the main paper. (1) We show that our results are robust to using the UCB-untuned (Auer et al., 2002) and the number of rounds to run the algorithm. (2) The UCB is based on a finite time analysis of the MAB problem and requires an ex ante specification of number of rounds. In this figure, we run the algorithm for 10 million rounds to show the robustness of our results. The results are statistically indistinguishable from the main results in figure 1.

The UCB index for the online appendix is given by

$$\text{UCB-untuned}_{kt} = \bar{\pi}_{kt} + \sqrt{\frac{2\log t}{n_{kt}}}$$

In the top panel of 5 we show that our results are robust to considering the UCB-untuned algorithm. Note consistent with the prior literature (including in Auer et al. (2002)) the untuned version of the algorithm takes longer to converge and requires a larger number of rounds (10 million). The bottom panel of 5 we show that our results are robust to considering 10 million round with the the UCB-untuned algorithm (as opposed on 2 million in the main paper)

[Figure 5 about here.]

Figure 6 shows the robustness of our results in two ways. First, we consider a different demand systems (logit demand system). Second, we use a different MAB index algorithm (Gittins index). Our main findings (see figure 1) are replicated in both these simulations: markets with low (high) SNR, independent bandits result in competitive (supra-competitive) resultant prices.

[Figure 6 about here.]

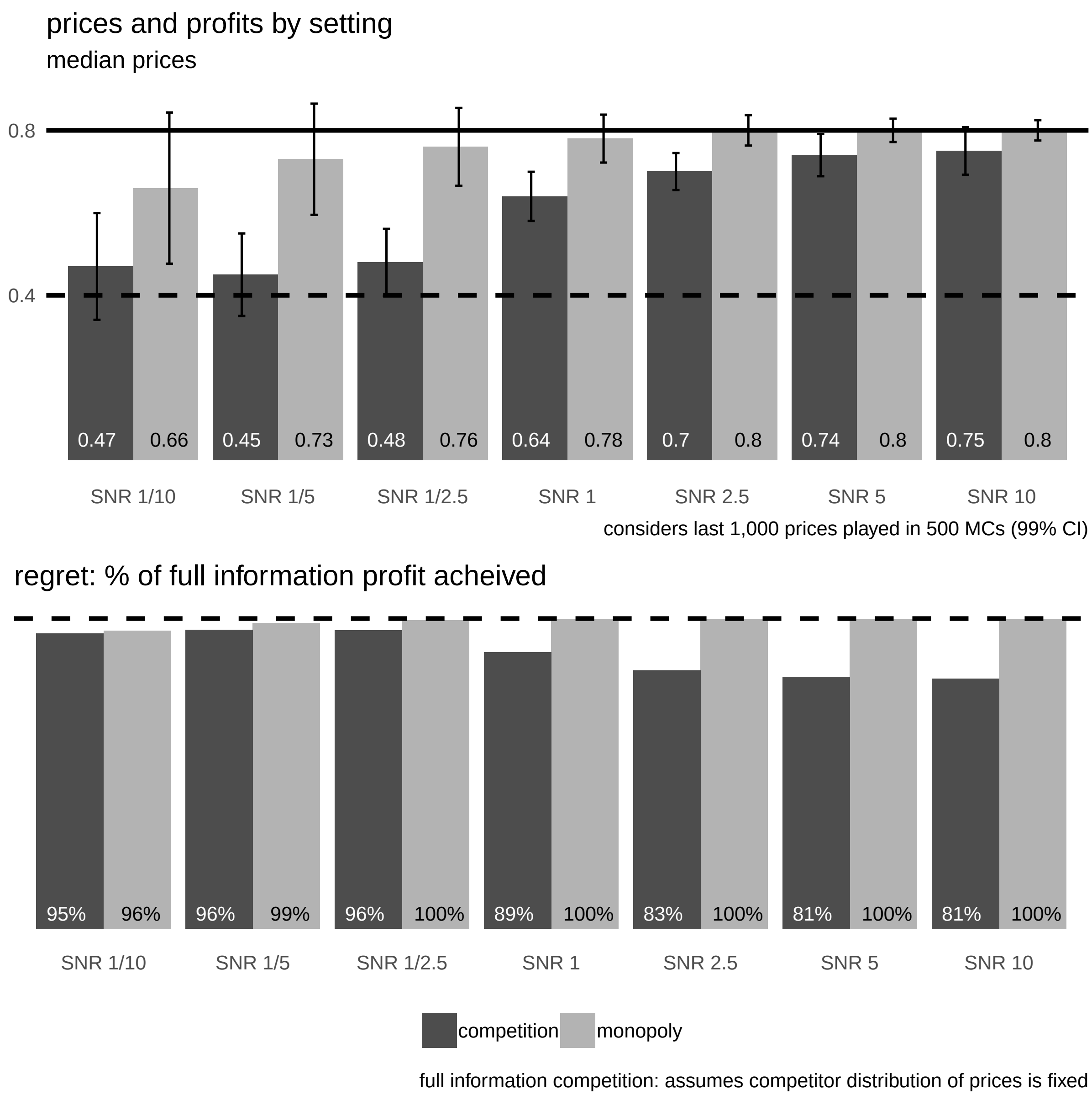

FIGURE 1. Estimated median prices and percent of optimal profits by simulation setting. The dark gray bars represent a setting where two firms are running simultaneous algorithms, while the light gray bars represent a setting when a monopolist is jointly pricing the products. For each simulation, we consider the last 1,000 periods out of 2 million periods. In the top chart, the dashed lines reflect the competitive equilibrium prices; the solid lines reflect monopoly prices. In the bottom chart, the dashed line reflects 100% profit achieved. [Algorithm used: UCB tuned]

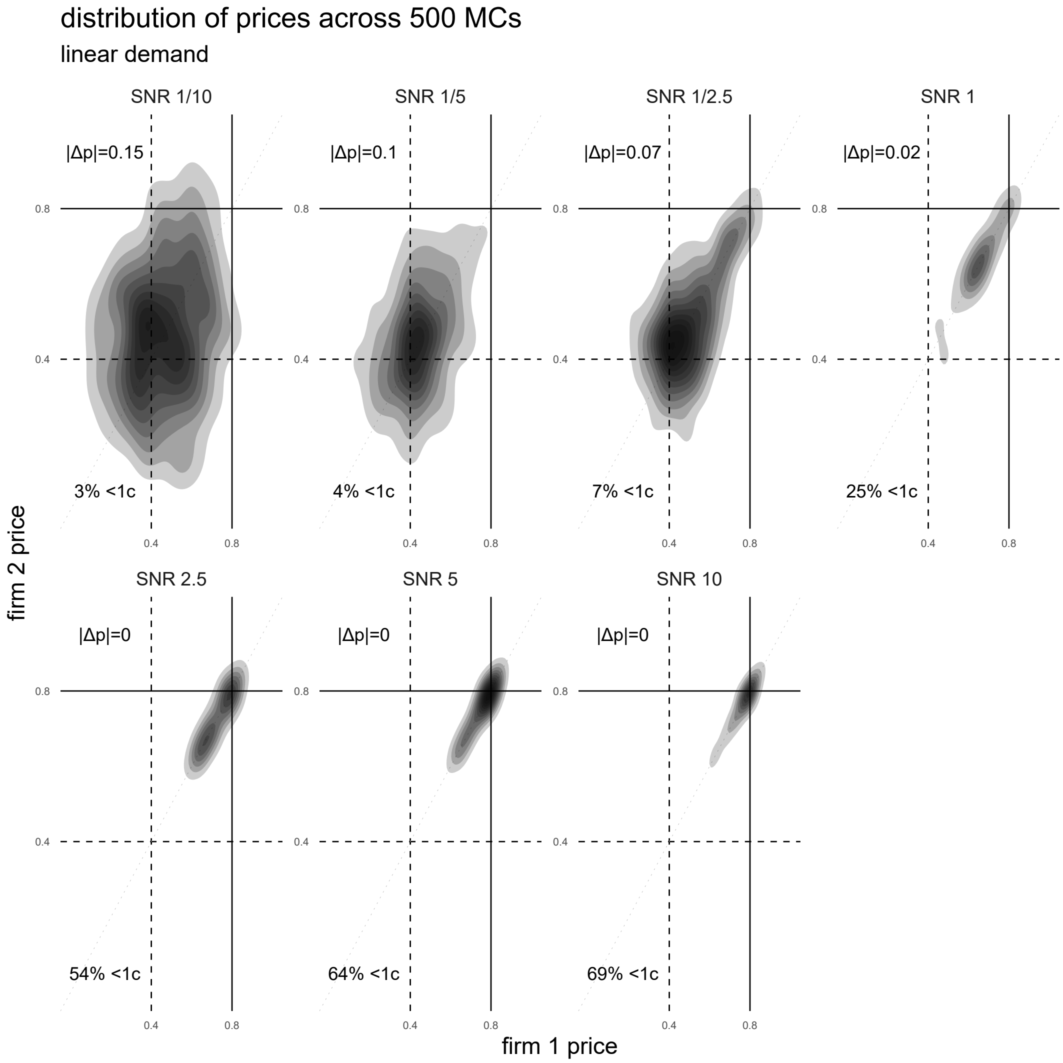

FIGURE 2. 2D Density plot of the distribution of prices for the two firms across 500 MC simulations per scenario. Each chart represents a market setting described by signal to noise ratio (SNR), a small SNR means large noise and a large SNR means small noise. For each simulation, we consider the median price charged in the last 1,000 rounds out of 2 million rounds. The dashed lines reflect the competitive equilibrium prices; the solid lines reflect monopoly prices. The light gray dotted line presents the 45-degree line. The number on the top left of each chart shows the median difference between the two firms' prices, and a number of the bottom left represents the percentage of simulations with the difference in price less than 1c. [Algorithm used: UCB tuned]

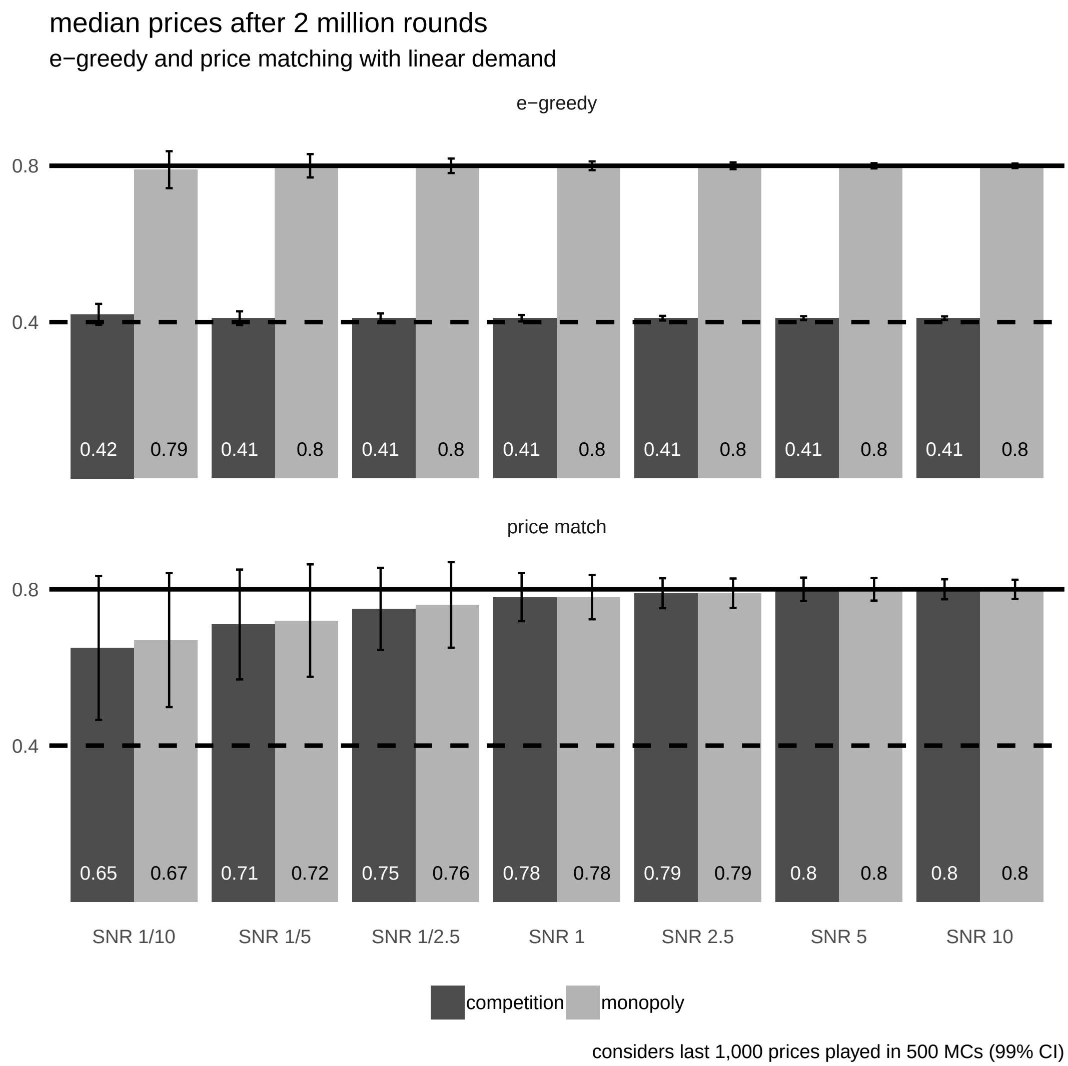

FIGURE 3. Estimated median prices by algorithm and level of competition. The error bars represent 99% confidence interval for the estimated median across MC simulations. The dark gray bars represent a setting where two firms are running simultaneous algorithms, while the light gray bars represent a setting when a monopolist is jointly pricing the products. The top panel corresponds to firms using the ε-greedy algorithm ($\varepsilon = 0.01$), and the bottom panel corresponds to one firm using UCB and the other price matching. For each simulation, we consider the median price charged in the last 1,000 period out of 2 million period. The dashed lines reflect the competitive equilibrium prices; the solid lines reflect monopoly prices. [Algorithms used: ε greedy or price matching with UCB tuned]

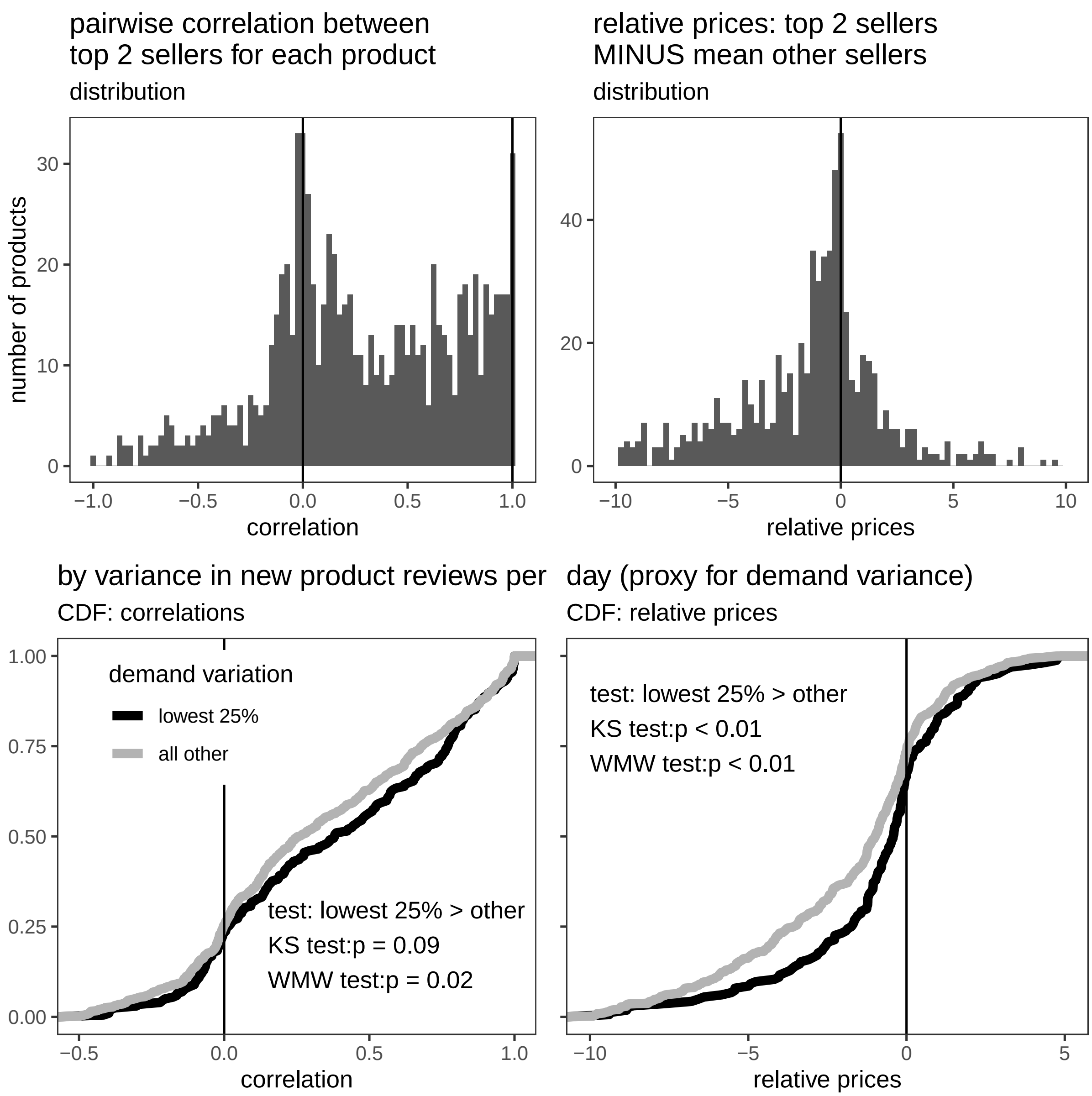

FIGURE 4. Analysis of Amazon pricing data in Chen et al. (2016). The left column consider pairwise correlations between the top 2 sellers (based on highest listings) for each product. The right column considers prices defined as the prices for the top two seller minus the mean price of all other sellers. The top row plots histograms to show the distribution of the raw data. The bottom chart considers the CDF of each measure by the number of new product reviews per hour (as a proxy for demand variation). We consider the products with the lowest 25% of demand variation and all other products. The two test preformed are the Kolmogorov-Smirnov Test and the Wilcoxon-Mann-Whitney Test to see if the distributions are shifted to the right for lower demand variation (the CDF of all other lies above that of lowest 25% demand variation).

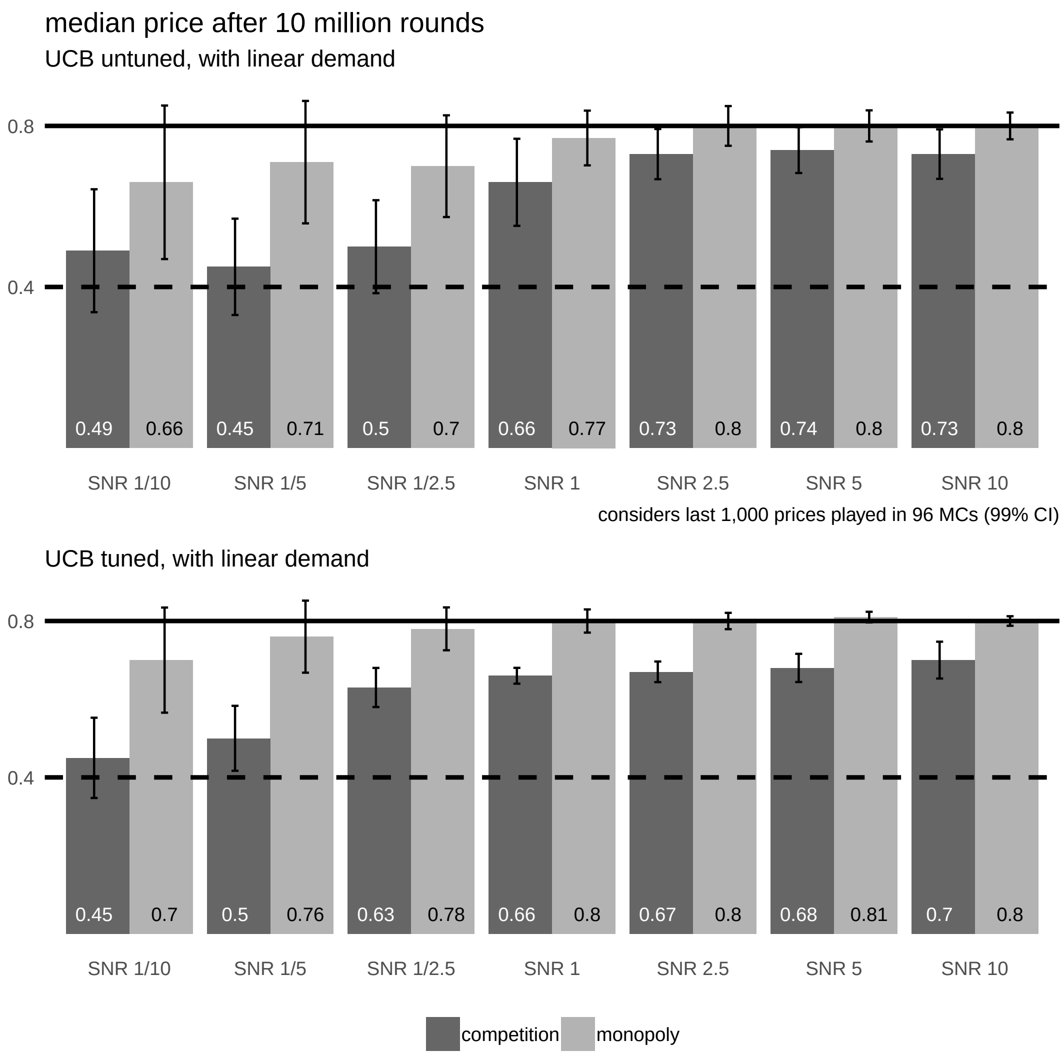

FIGURE 5. Estimated median prices by algorithm and level of competition. The error bars represent 99% confidence interval for the estimated median across MC simulations. The dark gray bars represent a setting where two firms are running simultaneous algorithms, while the light gray bars represent a setting when a monopolist is jointly pricing the products. For each simulation, we consider the median price charged in the last 1,000 rounds out of 10 million rounds. The dashed lines reflect the competitive equilibrium prices; the solid lines reflect monopoly prices. [Algorithm used: UCB tuned]

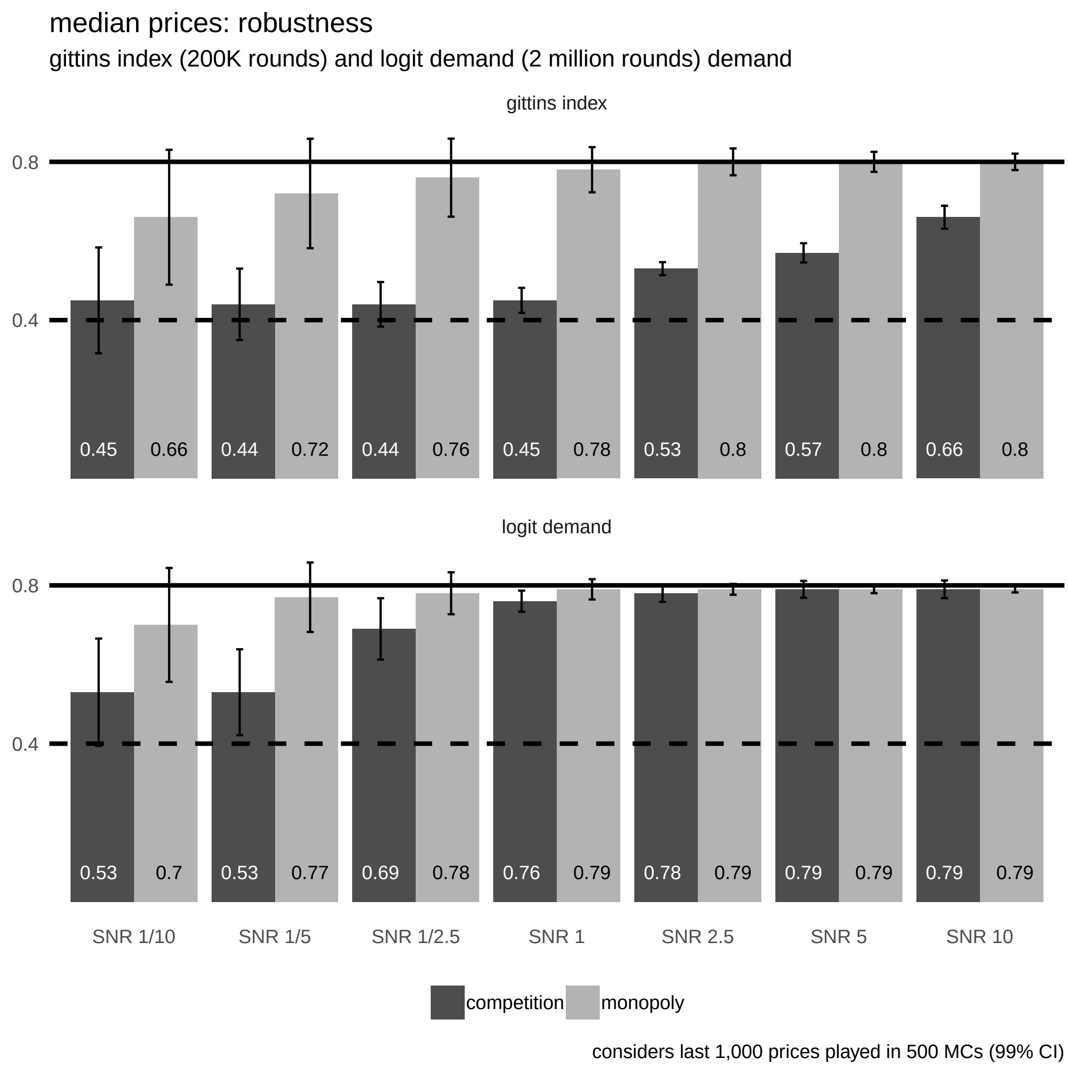

FIGURE 6. Estimated median prices by algorithm and level of competition. Gittins Index Gittins (1979): linear demand model with using the closed form Gittins index as suggested in Brezzi and Lai (2002), we assume a Beta prior and approximate the profit as an aggregate Binomal distribution. Note, this algorithm converges faster than the UCB and hence we consider 200,000 rounds. Logit demand model: $u_j = 4.1 - 4.74p_j + \varepsilon_j$ for $j \in \{1, 2\}$, with an outside option $u_0 = \varepsilon_0$. with ε i.i.d. type 1 extreme value. Under this demand system the NE price $P^D = 0.4$ and the monopoly price is $P^M = 0.8$. The error bars represent 99% confidence interval for the estimated median across MC simulations. The dark gray bars represent a setting where two firms are running simultaneous algorithms, while the light gray bars represent a setting when a monopolist is jointly pricing the products. For each simulation, we consider the median price charged in the last 1,000 rounds in each MC. The dashed lines reflect the competitive equilibrium prices; the solid lines reflect monopoly prices. [Algorithm used: Gittins index and UCB tuned]