Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market*

Stephanie Assad[a], Robert Clark[b], Daniel Ershov[c], Lei Xu[d]

June 7, 2023

Abstract

We provide the first empirical analysis of the relationship between algorithmic pricing (AP) and competition by studying the impact of adoption in Germany's retail gasoline market, where software became widely available in 2017. Because adoption dates are unknown, we identify adopting stations by testing for structural breaks in AP markers, finding most breaks to be around the time of widespread AP introduction. Because station adoption is endogenous, we instrument using headquarter adoption. Adoption increases margins, but only for non-monopoly stations. In duopoly and triopoly markets, margins increase only if all stations adopt, suggesting AP has a significant effect on competition.

*The views in this paper do not reflect those of the Competition Bureau of Canada. Helpful comments were provided by John Asker, David Byrne, Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, JP Dubé, Avi Goldfarb, Joe Harrington, JF Houde, Fernando Luco, Alex MacKay, Jeanine Miklós-Thal, Kanishka Misra, Ariel Pakes, Nadia Soboleva, Catherine Tucker, Matt Weinberg and seminar participants at the University of Bologna, UBC, Cambridge, Chicago, Colorado Boulder, Michigan, Montreal, New Mexico, Penn, Paris TelecomTech, Stanford, Tel Aviv Coller, TSE, Yale, CESifo, FTC, University of Amsterdam conference on algorithmic collusion, Netherlands Authority for Consumers and Markets, Bank of Canada - University of Toronto Joint Conference on Collusion, the 2020 NBER Summer Institute IT and Digitization, the 2020 NBER Economics of AI, the 2021 AEA meetings, the 2021 CEA meetings, QME 2021, and the 2022 ASSA meetings. Edited by Chad Syverson. Daniel Ershov would like to acknowledge support received from ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir program) and from ANITI. Correspondence to: [a]Stephanie Assad - Competition Bureau, Ottawa ON; Email: assads@econ.queensu.ca, [b]Robert Clark - Queen's University, Dunning Hall, 94 University Avenue, Kingston, ON; Email: clarkr@econ.queensu.ca, [c]Daniel Ershov - UCL School of Management, 1 Canada Square, London E14 5AA, United Kingdom; Email: d.ershov@ucl.ac.uk, [d]Lei Xu - IESE, Avenida Pearson, 21, 08034 Barcelona, Spain; Email: lei.xu2@gmail.com

1 Introduction

Pricing-algorithm technology has become increasingly sophisticated in recent years. Although firms have made use of pricing software for decades, technological advancements have created a shift from mechanically-set prices to AI-powered algorithms that can handle large quantities of data and interact, learn, and make decisions with unprecedented speed and sophistication. The evolution of algorithmic pricing (AP) software has raised concerns regarding possible impact on firm behaviour and competition. The potential for algorithms to facilitate collusion, either tacit or explicit, has been a popular discussion-point among antitrust authorities, economic organizations, and competition-law experts in recent years (OECD 2017; Competition Bureau 2018; Autorité de la Concurrence and Bundeskartellamt 2019; UK Digital Competition Expert Panel 2019; Ezrachi and Stucke 2015, 2016, 2017; Varian 2018; Goldfarb et al. 2019). Since the goal of algorithms is to converge to an optimal policy, AI agents could learn to play a collusive strategy to achieve a joint-profit maximizing outcome. AP software can also facilitate collusion through increased ease of monitoring and speed of detection, and through punishment of possible deviations.

The literature on algorithmic collusion is expanding, with contributions from economics, law, and computer science. At present, there is no theoretical consensus as to whether algorithms facilitate tacit collusion (Kühn and Tadelis 2018; Calvano et al. 2020; Miklós-Thal and Tucker 2019; Asker et al. 2021; Brown and MacKay 2023). Despite some evidence that collusive algorithmic behaviour can appear in synthetic environments, there are questions about whether it can and will arise in practice. As of yet, there is no empirical evidence linking the adoption and use of pricing algorithms to market outcomes related to competition. The objective of this paper is to supplement existing theoretical literature by conducting the first empirical analysis of the impact of wide-scale adoption of AP software. We focus on the German retail gasoline market, where, according to trade publications and news articles, AP software became widely available beginning in 2017, and for which we have access to a high-frequency database of prices and characteristics for every retail gas station in the country from January 2016 to December 2018.[1]

Investigating the impact of the adoption of algorithmic-pricing software on competition requires overcoming three important challenges. First, even with access to detailed pricing data, adoption decisions are typically not publicly observed. Second, adoption is endogenous, since the decision to

[1] **Legal disclaimer:** This paper analyses the impact of adoption of AP on competition strictly from an economic point of view. We base our understanding of the facts on publicly-available data on prices from the German Market Transparency Unit for Fuels. To our knowledge, there is no direct evidence of anticompetitive behavior on the part of any algorithmic-software firms or gasoline brands mentioned in this paper.

adopt is correlated with factors that are unobserved to the researcher. Finally, even if adoption can be causally linked with higher prices or margins, it is not clear whether these can be attributed to changes in competition intensity rather than to other factors, such as an improved ability to detect underlying fluctuations in wholesale prices or predict demand.

To overcome the first challenge we test for structural breaks in pricing behaviours that are thought to capture the promised impacts of sophisticated pricing software: (i) the number of price changes made in a day, (ii) the response time of a station's price to a rival's price change, (iii) the responsiveness of a station's price to crude oil shocks, and (iv) the responsiveness of a station's price to local weather shocks. We focus on these measures since leading providers characterize their software as performing high-frequency analysis to "rapidly, continuously and intelligently" react to market conditions. For each measure, we test for structural breaks at each station, considering each week in a large window around the time of supposed adoption (Quandt 1960). For each measure, the best-candidate structural break for a given station is the week with the highest F-statistic. Breaking in one of the four measures could occur for any number of reasons, but breaking in multiple markers in close proximity should provide a strong indication of adoption. Therefore, we classify a station as an AP adopter if it experiences a best-candidate break in at least two of four markers within a short time period, which we take to be four weeks, but is robust to alternative specifications. We find that approximately 20% of stations in our dataset experience best-candidate breaks in multiple markers within a four-week window. The majority of these breaks occur in mid-2017, just as AP software supposedly became widely available in Germany. Adopting stations experience noticeably different trends in all four measures, confirming that our data-driven approach for identifying adoption captures meaningful changes.

Having identified adopters, we next examine the impact of their adoption on retail prices and margins.[2] Although we control for time and station-specific effects, as well as time-varying market level demographics, individual station adoption decisions may be correlated with station/time specific unobservables (managerial skills, changing local market conditions, etc). We provide evidence of selection bias and diverging outcomes between non-adopters and adopters before their adoption date that attenuate OLS estimates to zero. We address this challenge by instrumenting for a station's adoption decision. Our main IV is the adoption decision by the station's *brand* (i.e., by brand headquarters). As demonstrated by previous technology-adoption episodes in retail gasoline, brands

[2]Previous studies on coordination in the retail gasoline market use margins (retail prices over wholesale prices) to evaluate competition (Clark and Houde 2013, 2014; Byrne and De Roos 2019), and theory papers on algorithmic competition also make clear predictions related to margins (Calvano et al. 2020; Brown and MacKay 2023).

can facilitate adoption by their stations. "Adopting" brands provide support/subsidies/training to individual stations, reducing adoption costs.[3] Brand-level decisions should not be correlated with individual station-specific unobservables, making this instrument valid. Since brand adoption decisions are also unobserved, we use a proxy as our instrument: the fraction of a brand's stations that adopt AP. If a large fraction of a brand's stations adopts, it is likely that the brand itself adopted and facilitated adoption by the stations.[4]

We find that, following adoption, mean station-level prices and margins increase by approximately 1.3 cents per litre, or roughly 15% for margins.[5] These estimates are similar in magnitude to claimed increases in gross profits achieved by stations employing AP software in Brazil and Denmark. Our findings provide evidence of the causal impact of adoption of AP software; however, it is not clear whether these higher margins can be attributed to changes in the degree of competition intensity rather than to factors such as an improvement in the ability to identify fluctuations in wholesale prices or to better forecast demand.

To isolate the effects of adoption on competition we focus on the role of market structure. We begin by comparing adoption effects in monopoly and non-monopoly markets. If adoption influences competition, its effects may be stronger for non-monopolists than for monopolists. However, there is a lack of clear theoretical predictions on how AP should affect average prices in different market structures if its only function is to improve a seller's ability to tailor prices to time-varying demand or cost conditions. Therefore, we restrict attention to small oligopoly markets (with two or three stations) to hold market structure roughly fixed, and perform a more direct test of the predictions of the literature studying the impact of AP. We compare market-level margins in markets where no stations adopted, where a subset of stations adopted and where all stations adopted. In the first type, competition is between rule-based algorithms. In the second it is between rule-based and AI-powered algorithms, while in the last it is only between AI-powered algorithms. By comparing all three market-types we can learn about the effect of AP on competition.

We observe heterogeneity in outcomes based on market structure suggesting AI-powered AP software may affect margins through competition. Adopting stations with no competitors in their local markets (i.e., monopolists) see no statistically significant change in mean margins or prices. In contrast, adopting stations with local competitors experience a statistically significant margin increase of

[3]Below we provide examples of other episodes of technology adoption in retail gasoline markets.

[4]As an alternative instrument we use an annual measure of broadband-internet availability in the area around each station. See Appendix G.4 for additional discussion.

[5]Estimates using alternative broadband availability IVs are qualitatively similar to the main estimates, although larger. See Appendix G.4 for additional discussion of these results.

1.3 cents per litre (cpl).[6] Our oligopoly market-level results indicate that, relative to markets where no station adopts, markets where all do see a margin increase of 3.2 cpl, or roughly 38%. Mean prices increase by 6 cpl. Markets where only a subset of stations adopts see no change. These results show that market-wide AP adoption raises margins, suggesting algorithms soften competition. The magnitudes of margin increases are consistent with previous estimates of the effects of coordination in retail gasoline markets (Clark and Houde 2013, 2014; Byrne and De Roos 2019).

To provide further evidence of the impact on competition and to better understand the mechanism we examine whether algorithms actively learn how not to compete (i.e., to tacitly collude) by testing the *timing* of price and margin changes. Updating algorithms operating in fluctuating markets should adjust slowly, as they learn and explore the state space and set of possible outcomes. As a result, convergence to stable strategies can take as long as several years. Asker et al. (2021) show that less-sophisticated *asynchronous* algorithms converge towards the monopoly price, but take many periods to do so. Similarly, Calvano et al. (2020) suggest it can take time for algorithms to train and converge to stable strategies, involving for instance punishment for rival price reductions. We find evidence consistent with these results. Margins start to increase only about a year after market-wide adoption, suggesting algorithms in this market learn tacitly-collusive strategies. We also examine behaviour that emerges in markets where all stations adopt. In these markets a station is more likely to respond to a rival's price decrease with an immediate decrease of its own. There is no comparable change in the propensity to respond to rival price increases. The timing of these effects is consistent with that of the price and margin increases. Altogether, these findings provide further evidence that adoption affects competition and suggest that algorithms learn that undercutting will not be profitable, since lower prices will be followed.

Our results have important policy implications. Globally, antitrust authorities are considering adjustments to their toolkits to address the challenges of the digital economy (Autorité de la Concurrence and Bundeskartellamt 2019; UK Digital Competition Expert Panel 2019). Currently, authorities expend substantial resources pursuing hard-core cartels on individual bases, possibly overlooking a broader set of collusion-facilitating devices that do not even require a conspiracy. AP may be one such mechanism. Communication via earnings calls is another (see Aryal et al. 2022). We provide further policy discussion along with some recommendations in Section 8.

The remainder of this paper is laid out as follows. The next section discusses relevant literature. Section 3 provides a background discussion and an overview of the German market. Section 4

[6] We find that the pricing behaviour of adopting monopolists changes in ways that does not increase average daily prices and is consistent with improved ability to price discriminate.

describes the data and methodology used to identify AP adopters. Section 5 displays results on the impacts of AP adoption on outcomes. In Section 6 we provide evidence that results are driven by algorithms learning to tacitly collude. Section 7 presents a series of robustness results. Finally, in Section 8 we present a brief policy discussion and some conclusions.

2 Related Literature

This paper is most closely related to the recent literature concerning the potential link between AP and collusion. Theoretical and experimental results remain ambiguous. Several papers have shown that when AP competition is modelled in a repeated game framework collusive outcomes are possible under certain conditions (Salcedo 2015; Calvano et al. 2020; Klein 2021); however, others argue that improved price response to demand fluctuations may provide increased incentives for firm deviation from a collusive price (Miklós-Thal and Tucker 2019; O'Connor and Wilson 2020). Klein (2021) and Calvano et al. (2020) use computational experiments to study the effect of Q-learning algorithms on strategic behaviour of competing firms. Both find that these repeated games will converge to collusive outcomes including supra-competitive pricing and profits, as well as punishment of competitor deviation.[7] Asker et al. (2021) find that the sophistication of an algorithm's design affects the extent to which prices increase above the competitive benchmark. While Miklós-Thal and Tucker (2019) find that improved demand prediction may lead to the possibility of collusion in markets where it is previously unsustainable, in other markets it may create incentives for deviation that were absent with less prediction capabilities. O'Connor and Wilson (2020) come to similar conclusions. Brown and MacKay (2023) develop a model where firms compete in pricing algorithms (rather than prices) and show that prices may increase even without collusion.[8] Overall, there is little certainty as to whether algorithmic competition will lead to collusive outcomes in reality. There is, as far as we are aware, no empirical research regarding this question in the economics literature.[9]

[7] Johnson et al. (2023) propose market-design policies to disrupt algorithmic collusive strategies in platform settings.

[8] See also Lamba and Zhuk (2022). Harrington (2022) shows that outsourcing the development of pricing algorithms to profit-maximizing third party developers can also increase prices and reduce consumer welfare by making algorithms more sensitive to changes in demand.

[9] Decarolis and Rovigatti (2021) find that common bidding intermediaries in online advertising markets lead to anti-competitive effects, reducing prices for bidders at the expense of the platform. Bidding is done through algorithms, which leads to regulatory concerns about multiple competitors in a market adopting the same pricing algorithm. Their findings suggest that algorithms could serve as "hubs" in a hub-and-spoke cartel (Garrod et al. 2021, Clark et al. 2023). The primary focus is on increasing intermediary concentration rather than on AP software behaviour. Two recent working papers study the rise of automatic pricing tools at e-commerce sites and investigate whether they can facilitate collusion. See Wieting and Sapi (2021) and Musolff (2022).

The question as to whether the use of algorithms may result in coordinated behaviour has been studied in fields outside economics such as law and computer science. In computer science Kaymak and Waltman (2006, 2008) and Moriyama (2007, 2008) indicate that reinforcement learning algorithms can result in cooperative outcomes; however, these outcomes are not always the most likely and are dependent on various specifications of the algorithm. Legal scholars generally voice more certainty that AP can lead to collusion. Ezrachi and Stucke (2015, 2016, 2017) and Mehra (2015) have expressed concern over this issue and its implications for competition policy.

We also relate to an extensive literature on retail gasoline markets. A number of papers have examined collusion, including Borenstein and Shepard (1996), as well as Slade (1987, 1992). More recently Wang (2009), Erutku and Hildebrand (2010), Clark and Houde (2013, 2014), and Byrne and de Roos (2019) have all studied anti-competitive behaviour in retail gasoline. There is a small set of papers looking at the German retail gasoline market (Dewenter and Schwalbe 2016; Boehnke 2017; Cabral et al. 2021; Montag and Winter 2019).

An associated literature studies the impact of technological advancements on price discrimination. A consequence of the rapid expansion of Big data and AI-driven analysis is that personalized pricing strategies may become increasingly sophisticated. It is possible that more accurate determination of optimal personalized pricing can increase firm revenues (Shiller and Waldfogel 2011; Shiller 2014). Kehoe et al. (2020) find that firm profit, and consumer surplus, may increase or decrease under personalized pricing depending on consumer certainty regarding product tastes, and that total welfare is higher under discriminatory pricing in comparison to uniform pricing. Dubé and Misra (2023) show through experiments that personalized pricing improves firm profits and that a majority of consumers benefit. MacKay et al. (2022) also show efficiency gains under dynamic pricing.

3 Background

3.1 The German Retail Gasoline Market

As in other retail gasoline markets around the world, a large fraction of stations in Germany have a brand affiliation.[10] ARAL and Shell are the largest, combining to make up over 25 percent of stations

[10] Our data set does not specify which stations are vertically integrated and directly owned by the brands and which are owned by independent franchisees who enter into a licensing agreement in exchange for the brand name and some technical support. Both are common in retail gasoline markets (Convenience.org).

in Germany.[11] There are a number of other large brands with over 350 stations each: Esso, Total, Avia, Jet, Star, BFT, Agip, Raiffeisen, and Hem. ARAL, Shell, Jet, BFT, Total and Esso together account for 84 percent of fuel sales in the German retail gas market.[12]

Two features of the German market are important for our analysis. First, price transparency was instituted in August 2013 in response to concerns about high prices and tacit collusion by regulatory authorities.[13] As part of this initiative, stations adjusting their price must report new prices "in real time" to the German Market Transparency Unit for Fuels (www.bundeskartellamt.de), which are then shared with consumer-facing information service providers and integrated into websites and mobile applications as well as into car GPS systems.[14]

Second, Shell introduced a price matching guarantee in 2015. Each Shell station was required to match the lowest price of the 10 nearest stations within a 30-minute period, for consumers with Shell loyalty cards. Dewenter and Schwalbe (2016) and Cabral et al. (2021) find the policy led to retail price increases, attributed to an additional price jump incorporated into the daily price cycles that characterize Germany's retail gasoline market. Before the policy, prices were elevated then gradually decreased starting around 8am before rising sharply again in the evening. After the policy was implemented, a price jump at noon, followed by reversion, emerged. Overall, stations featured considerable price variability throughout the day.[15] We perform several robustness checks to confirm that Shell stations (or their direct competitors) are not driving the main results (see Appendix G.1).

Our paper takes this environment as given and so we study the *additional* effects of algorithmic pricing software in a market with price transparency, daily price variation, and price matching.

[11] Detailed summary statistics of station numbers at the brand level are in Section 4.

[12] 2019 shares of fuel sales are Aral 21%, Shell 20%, BFT 16%, JET 10.5%, Total 9.5%, and Esso 7% (bft.de).

[13] The purpose is to allow "motorists...to gain information on the current fuel prices and find the cheapest petrol station in their vicinity or along a specific route" and to "increase competition" (www.bundeskartellamt.de). Evidence on the effect of this policy on prices and margins in Germany is conflicting (Dewenter, Heimeshoff and Luth 2017, Montag and Winter 2019). See Luco (2019) for analysis of a similar program in Chile.

[14] A full list of consumer facing data providers is here: `https://www.bundeskartellamt.de/EN/Economicsectors/MineralOil/MTU-Fuels/mtufuels_node.html`. We obtained our data from Tanker-Konig, one such provider.

[15] Similar patterns have been documented in other markets. For instance, Wang (2009) provides evidence that in Australia prices fluctuated multiple times throughout the day prior to the implementation of a pricing reform in 2001. The same is true of some markets in the US and Canada (see CTV.ca).

3.2 Use of Algorithmic Pricing Software in Retail Gasoline Markets

3.2.1 History of Algorithmic Pricing in Retail Gasoline

Fuel retailers are typically not forthcoming about the pricing technologies they employ. AI-driven algorithmic pricing software providers are also mostly secretive about their customer base, and little is known about the structure of the market, or the market shares of particular software providers. A *Wall Street Journal* article on the subject mentions certain firms, including the Danish company *a2i Systems* and Belgian company *Kantify*, as notable providers (WSJ.com). Other firms, not listed in the article but prominently featured on the internet as algorithmic software providers, include *Kalibrate* (Kalibrate.com), *Revionics* (Revionics.com) and *PDI* (PDIsoftware.com).

The use of algorithmic pricing software in European fuel retail markets began in the early 2010s. *a2i* sold their software to Danish fuel retail company OK Benzin in 2011 (a2i Systems). However, the main penetration of machine learning and AI based software appears to have happened in the mid 2010s, roughly coinciding with the publication of several newspaper articles about the subject in 2017 (WSJ.com,CSPDailyNews.com).[16] Kalibrate began explicitly distinguishing between rule-based pricing and algorithmic pricing on its website in mid-2017 (2016 Kalibrate.com, 2017 Kalibrate.com). *a2i*'s software was tested in workshops with stations in the Netherlands and Belgium in 2015 (servicestationmagazine.be) and adopted by some Shell stations in the Netherlands by 2017 (WSJ.com).

In Germany, a number of trade publications and news articles mention that algorithmic pricing software became available in 2017, noting in particular the introduction of *a2i*'s software.[17] The websites of some software providers also suggest that they became active around this time. Evidence of this introduction and of adoption activity in Germany is presented in Appendix B.1.

Promotional materials by retail gasoline AP software providers around the world make claims that stations using their pricing software outperform stations with human pricing agents. The Brazilian pricing start-up Aprix estimates that gas stations using its AI-based pricing software increased station gross profits by approximately 10% (towardsdatascience.com). a2i similarly estimated that its software could increase station profits by at least 5% (a2i.com).

[16] It is possible that providers sold algorithmic pricing software in Germany before 2016 (the start of our sample). We should not observe any structural breaks for stations that adopted before the start of our sample. This means that we would be labelling some adopters as non-adopters. If adopters have higher average margins than non-adopters, this would bias our station-level estimates towards zero.

[17] In conversations with us, *a2i* claims that, contrary to statements in these advertising materials, they were never active in the German market.

3.2.2 How does algorithmic pricing software work?

Software providers reveal little about their algorithms, but promotional materials describe the software as based on "machine learning" or "artificial intelligence," with references to "neural networks" and "deep learning" (Kalibrate.com, PDISoftware, a2i.com). They are characterized as able to help station owners "master market volatility with AI-powered precision pricing and respond rapidly to market events and competitor changes" (Kalibrate.com) and take advantage of "superhuman expertise" (a2i.com). Additional promoted benefits include optimizing for long-term revenues and avoiding price wars (Kantify).

Providers stress the ability of algorithms to incorporate market conditions and variables such as own and competitor prices, sales volumes, costs, and weather and traffic events into their decision-making. *a2i Systems* provides more detail, outlining its algorithm in Derakhshan et al. (2016).[18] It is described as a "multi-agent-system" based on the interaction of two agents: a consumer and a gas station. Agent behaviour is described by a "belief-desire-intentions" (BDI) model, a popular approach in computer science and information systems research. An agent's "beliefs," "desires" and "intentions" roughly correspond to information, payoffs and actions/strategy in decision-theory.[19]

a2i's algorithm works in three repeating steps. The first is "observation," where the agent collects data from the environment and forms "beliefs." As mentioned previously, these data include own prices, sales, traffic and environmental factors. Competitor behaviour is not explicitly modelled but rival prices are included as inputs at this step. In the second step, "learning," the gas station agent uses an Artificial Neural Network (ANN) to map inputs into outcomes.[20] The outcomes are not explicitly outlined in Derakhshan et al. (2016), but likely correspond to sales, revenues and/or profits.[21] These are the "desires"/payoffs in the BDI model. The last step is "adaptation," where the agent sets prices to achieve their "desires"/maximize the objective function.[22]

[18]This algorithm is based on the earlier papers Derakhshan et al. (2006) and Hammer et al. (2006). These papers look at interactions of children at a playground with the goal of encouraging more physical activity.

[19]Individual station owners can set different goals such as market share maintenance or constraints such as minimum price. They can also change the goals over time or adjust them. However, substantial changes by station owners does not happen much in practice. One algorithmic software provider states that approximately 80-90% of station owners *do not customize or interfere* with the default operations of the algorithm (Kalibrate.com).

[20]This step also implicitly models consumer behaviour, but this is not described.

[21]In the earlier papers on children's playgrounds that form the basis of this algorithm, outcomes are categories that capture whether children are playing fast or slow, continuously or discontinuously, etc (Derakhshan et al. 2006).

[22]Similarly, The Brazilian provider Aprix claims to "simulate the demand reaction for different price and market scenarios" by cycling through three stages: modelling station and consumer behaviour, simulating (or mapping) the relationship between inputs (the state) and desired outputs (margins, profits, market shares), and optimization by setting station prices to reach maximum outputs conditional on the state. As with a2i, the algorithm continuously performs these stages and re-optimizes (towardsdatascience.com).

An interpretation of AI-driven AP software is that it makes stations more sensitive to the state of the market. *The Wall Street Journal*, presents a summary of its functioning, describing constant learning about the state of the market (wsj.com). For retail gas this means collecting demand-related information such as weather and traffic that can change driving behaviour and the probability of stopping for gas, cost-related information such as crude oil price fluctuations, or other relevant information such as competitor prices. This can now be done at high frequency by scraping websites (e.g., weather websites or Google maps). The algorithm then relates this information to outcomes and decides on the best price to set conditional on the state. Human- or rule-based pricing agents operators can also collect information about what competitors and consumers do and set prices in response (Time.com), but algorithms collect and process more information than any human could. Algorithms can also respond faster to changes in the state, and to very subtle changes that humans might miss, consistent with evidence from hotels showing that human pricing agents exhibit more inertia and higher price-adjustment costs than algorithms (Garcia et al. 2021).

For further discussion about how the algorithms operate, see Appendix B.2.

3.3 Algorithmic Pricing Software Adoption

As in other cases of corporate technology adoption (e.g., Tucker 2008, Ryan and Tucker 2012), technology adoption in gasoline retail happens at two levels: at the brand HQ (headquarters) level and at the individual station level. Brands make big-picture decisions about technologies they would like their stations to use, and provide stations with employee training, technical support, maintenance and subsidies. Individual station owners make adoption decisions specific to their stations, potentially incurring substantial investment costs that are not necessarily fully subsidized by the brand.

An example electronic payment system adoption in the 1990s. As with algorithmic pricing software, brands wanted their stations to adopt this technology, but some stations may have been reluctant because of the costs involved. As part of a brand-wide roll-out of a contactless electronic payment system in 1997 Exxon Mobil offered a $1,000 rebate towards the $17,000 installation fee (per pump).[23] Partial subsidies help explain staggered / delayed technology adoption in this market.[24]. We look at the adoption of electronic payments from 1991 to 2001 by Canadian gasoline retail stations and document that it takes years after the first appearance of this technology for a substantial fraction of stations belonging to the five biggest brands in the market to adopt. Even after 10 years

[23] See BusinessWeek.

[24] We provide additional evidence for staggered technology adoption in the gasoline retail market in Appendix F.

of availability, fewer than 50% of stations owned by leading brands adopted the technology (Figure F1).

There is no reason to suspect that algorithmic pricing software adoption is different. Anecdotal evidence suggests that gasoline brands have entered into long-term strategic partnerships with AI pricing and analytics providers, either directly or indirectly.[25] However, should a brand decide to "adopt" or enter into a partnership with an AI pricing software provider, its stations do not necessarily automatically and instantaneously adopt for a variety of reasons. Cloud-based AI-pricing software may require substantial infrastructure investments and not all station owners are in a position to incur these costs when the technology becomes available.[26] In Germany, many areas do not have access to stable high speed internet connections.[27] Station operators also require training with the software to set its parameters and deal with potential errors.

4 Data

This section provides a general description of the datasets used in our analysis. The Online Data Appendix contains more details about data construction. The main dataset comes from the German Market Transparency Unit for Fuels and includes all price changes for the most commonly used fuel types (E5, E10, diesel) for approximately 14,500 German gas stations. For each station, the raw data also include location information (5-digit ZIP code, latitude and longitude coordinates), as well as an associated brand.[28] Our sample covers January 2016 to December 2018.[29] We focus on E5 fuel, which has over 80% market share in Germany (bdbe.de).[30]

[25]For example, in Denmark *a2i* directly entered into a partnership with Danish retail fuel company OK Benzin (a2isystems.com). More indirectly, AI-pricing software providers enter into partnerships with IT companies that provide integrated services to brands. *Tankstop*'s June 2017 issue mentions that *a2i*'s services are supported by WEAT Electronic Data Service GmbH, a provider of cash-free payment systems and technical and logistical support for a number of petrol brands within Germany (WEAT.de). *a2i* also has a strategic partnership with Wincor Nixdorf, a retail-technology company providing services such as PoS terminals and self-checkout solutions (DieboldNixdorf.com).

[26]For example, high-speed internet *and* high-speed internet enabled PoS terminals and pumps are likely required for the software to work. Equipment upgrades of this sort are expensive, costing thousands to tens of thousands of Euros (mobiletransaction.org). Again, this is analogous to previous cases of technology adoption and upgrading decisions by gas station owners, including allowing for chip cards or automated payment at the pump (Chicago Tribune).

[27]Reports suggest many German regions receive sub-par services and speeds comparable to the "old dial-up days" (NPR.org). We use broadband internet availability as an alternative instrument. See discussion in Appendix G.4.

[28]We do not observe the ownership structure of the stations.

[29]Additional data exist for 2014-2015, but we choose to start our sample two years after the beginning of the transparency initiative and one year after Shell's price matching policy described in Section 3.1. Results are robust to alternative samples (Appendix G.1).

[30]Super E5 is an ethanol based fuel with 5% ethanol and 95% unleaded petrol. We find similar results using E10 fuel and diesel as reported in our robustness section.

We also make use of regional wholesale fuel prices from Oil Market Report (OMR), a private independent German gasoline information provider, and we merge in annual regional demographics from Eurostat. We include data on total population, population density, median age, employment (as a share of total population) and regional GDP. These data are at the "Nomenclature of Territorial Units for Statistics 3 (NUTS3) level, which is frequently used by EU surveys. A NUTS3 region is roughly equivalent to a US county and larger than a 5-digit ZIP code. We also incorporate weather information from the German Meteorological Service (DWD) and oil price data from FirstRate Data. Finally, we collect data on local fixed-line broadband internet from the EU Commission's netBravo initiative (netBravo): whether the local area around the gas-station has widespread availability of 10 Mb/s internet in a given year.[31] Unfortunately, we do not have access to sales data, so, to confirm there were no significant changes during our sample period we study aggregate volumes obtained using snapshots from the Wayback Machine of website https://www.bdbe.de/daten/marktdaten-deutschland.[32] Throughout the time period, sales of E5 hover around 15 million tons (14.6 million in 2014, 15.0 in 2015, 15.1 in 2016, 15.0 in 2017, 14.7 in 2018).

4.1 Station-Level Descriptive Statistics

Table 1 presents summary statistics, including the number of stations per brand, the number of stations per market, and the distance between stations. Out of the 14,565 stations in our data set, single-operating stations account for approximately 7 percent. With our IV strategy, these stations are not part of our final estimating sample. The remaining stations are affiliated with brands.[33] There are 232 distinct brands in the data, of which 215 have between 2 and 100 stations and 17 have more than 100. The top 5 brands account for 46 percent of stations and the 19 largest brands (those with more than 100 stations) account for 74 percent of total stations (10,720 stations total).

The market definition we use is based on the clustering algorithms developed in Carranza, Clark and Houde (2015) and Lemus and Luco (2019). The algorithm is implemented using distances between station pairs. Details are provided in Appendix C. Using this approach there are 3,957 markets of which 526 have a single station (monopoly markets), 789 have two stations (duopolies)

[31] We define 10 Mb/s to be widely available in an area if average speed-tests in that area in that year exceed that speed. More details on the construction of this variable are in the Data Appendix.

[32] See https://web.archive.org/web/20181116121312/https://www.bdbe.de/daten/marktdaten-deutschland and https://web.archive.org/web/20190929194231/https://www.bdbe.de/daten/marktdaten-deutschland.

[33] The data set does not specify whether the stations are vertically integrated and directly owned by the brands, or whether they are owned by independent franchisees who have entered into a licensing agreement in exchange for the brand name and some technical support. Both are common in retail gasoline markets (Convenience.org).

Table 1: Brand and Market Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	P25	P75
Stations per Brand	232	59	224	2	2256	3	18
Stations per Market	3957	4	2	1	19	2	5
Stations per 5 Digit ZIP Code	5488	3	2	1	16	1	3
Distance to nearest station (KM)	14565	1.42	1.77	0	17.19	.32	1.69

and 879 have three stations (triopolies). The full distribution is presented in the appendix. The mean number of stations per market is around four. Only 60 markets have more than 10 stations.[34]

Table 2: Station/Month Summary statistics

Variable	Observations	Mean	Std. Dev.
Prices and Margins			
Mean Monthly E5 Price (EUR/litre)	448,221	1.362	.083
Mean Monthly E5 Margin (EUR/litre)	448,221	.083	.032
Regional Demographics and Weather Controls			
ln(Total Regional Population)	448,221	12.419	.816
Regional Population Density (pop/km^2)	448,221	758.238	1022.41
Regional Median Population Age	448,221	46.018	3.125
ln(Regional GDP)	448,221	9.083	.976
Regional Employment Share (employed/pop)	448,221	.527	.134
Mean Temperature (degrees Celsius)	448,221	10.417	6.87
Std. Dev. Temperature (degrees Celsius)	448,221	3.079	.806
Mean Precipitation (mm)	448,221	1.94	1.399
Std. Dev. Precipitation (mm)	448,221	3.603	2.605
Broadband Availability			
At Least 10 Mb/s Internet Available Dummy	443,752	.834	.371

4.2 Station/Month-Level Descriptive Statistics

Using prices in the raw data, we calculate a mean weekday (non-weekend or holiday) price from 7am to 9pm for each station. To construct margins, we merge these with regional wholesale prices (average daily ex-terminal prices in eight major German refinery and storage areas). We calculate the distance between each station and all refinery and storage areas and use wholesale prices from the

[34]In the appendix, we also consider a market definition based on 5-digit ZIP codes. In Europe, this is the most detailed ZIP code available. There are 5,488 5-digit ZIP codes in our data of which 2,039 have a single station (are monopoly markets), and 1,286 have two stations (are duopolies).

nearest refinery.[35] Prior to subtracting wholesale price, we also subtract German VAT (19%) from retail prices. We compute station-level daily margins and take the monthly mean for our station-month level analysis. In addition to daily average prices, we also calculate prices at different points during the day for each station. For each station and weekday, we calculate the price at 9am, noon, 5pm and 7pm. Once again, we take a monthly mean for our station-month level analysis.

Table 2 displays summary statistics at the station/month level, for prices, margins and regional demographics and weather. The mean price charged is 1.36 Euros per litre, and the monthly mean margin earned by the average station over wholesale regional prices (after subtracting VAT) is 8.3 cents per litre. The average station is located in a fairly dense NUTS3 region, with population density of 760 persons per square-km. The median age of the population around a station is 46 years and 53 percent of the population is employed. Over 83 percent of gas station/month observations are for areas with widely available 10Mb/s internet access. The weather data are collected daily from thousands of weather stations. We compute the average distance between each station and all local weather stations and use data from the nearest weather station. We include monthly means and standard deviations of temperature (in degrees Celsius) and precipitation (in mm).

4.3 Identifying Algorithmic Pricing Adoption

4.3.1 Station-Level Adoption

We do not have information on the algorithmic-pricing software adoption of individual stations or brands. Our approach is to take advantage of the detailed price data to identify changes in station pricing technology, since we see price changes at 1-minute intervals. As discussed in Section 3.2, algorithms use machine learning to optimize prices conditional on a "state" that includes variables such as competitors' prices, weather conditions, and traffic. Changes in a station's prices responsiveness to the "state" should indicate the adoption of AP software. We consider the following four variables that capture a gas station's responsiveness to the "state":

1. Number of price changes made in a day: we calculate the number of times each gas station changes their price in each non-holiday weekday. We then average across each week.
2. Average response time to a rival's price change: we define a rival to station i as the nearest station j that belongs to a different brand and is within 1km of station i. After each price

[35]This is a standard approach in the retail gasoline literature. We may be understating retail margins if stations belong to vertically integrated retailers.

change by station j in each non-holiday weekday, we calculate the average time in minutes it takes station i to respond. We then average across each week.

3. Responsiveness of a station's prices to crude oil price shocks: using data from FirstRate Data, we observe an intra-day time series for crude oil prices. In each non-holiday weekday, we separate fluctuations in crude oil prices from the moving average. We define a crude oil price shock as large deviations from the moving average.[36] We define a response to a crude oil price shock as a price change within 5 minutes of the shock. We calculate the number of shocks for each week, and the number of responses and response probability for each station and week.
4. Responsiveness of a station's prices to local weather shocks: using data from the German Meteorological Service (DWD), we observe a high-frequency time series of local air temperature around each station.[37] We separate fluctuations in temperature from the moving average. We define a local weather shock as a large deviation from the moving average,[38] and a response to a local weather shock as a price change within 5 minutes of the shock. We calculate the number of local shocks, number of responses and response probability for each station and week.

Similar measures have also been used previously in the literature to identify heterogeneity in pricing technology (Brown and MacKay 2023; Aparicio et al. 2021; Chen et al. 2016).[39]

Formally, we look for structural breaks in these measures using Quandt-Likelihood Ratio (QLR) tests (Quandt 1960, Andrews 1993). This method tests for the best-candidate structural break in a time-series measure for each period in some interval of time and takes the largest resulting test statistic. It is useful when an exact break date is unknown, and has been used in previous work involving collusive behaviour (Harrington 2008; Clark and Houde 2014; Boswijk et al. 2018; Byrne and de Roos 2019). We conduct a QLR test for each station and for each measure. Further details can be found in Appendix D.

Appendix A shows the distribution of structural breaks for each measure.[40] We find a large

[36] Defined as deviations from the moving average above the 90th percentile of all deviations in a given year-month, helping to account for changing volatility of oil prices over time.

[37] We also observe local precipitation. Fluctuations and shocks in precipitation are generally highly correlated with shocks in temperature, but there are some areas and time periods that are drier and so have no precipitation and no fluctuations. Variation in air temperature exists for all stations throughout our sample period.

[38] Defined as deviations from the moving average above the 90th percentile of all deviations in a given year-month.

[39] Brown and MacKay (2023) use the number of price changes by retailers in a given period and the speed of reaction to identify new pricing technology. Similarly, Aparicio et al. (2021) document a higher frequency of price changes by online retailers who use algorithmic pricing. Chen et al. (2016) also identify algorithmic pricing users in Amazon Marketplace by measuring the correlation of user pricing with certain target prices, such as the lowest price of that given product in the Marketplace.

[40] A concern is that other breaks may occur at different dates if we considered F-statistics that are not the maximum, but close to it. We find that F-statistic distributions are generally unimodal and that stations do not have significantly

number of statistically significant breaks in the data. Nearly 50% of best-candidate breaks in the number of price changes are in the spring of 2017 when we believe AP technology became available. Similarly, 40% of best-candidate breaks in the responsiveness to local weather shocks, 20% in rival response time, and nearly 20% in responsiveness to oil price shocks happen around that time.[41]

We also find that the breaks capture quantitatively important changes in pricing behaviour. On average, stations without structural breaks in the number of price changes adjust their prices approximately 5 times per day, or roughly once every 3 hours, assuming average opening hours from 7am to 9pm.[42] Stations with structural breaks change their prices approximately 8 times per day during the sample period, or approximately once every hour and a half (9.3 times per day in 2018).[43]

For response time to rivals, the average for stations without breaks is over 84 minutes, compared to 50 minutes for stations with a break.[44] This is at least as fast as the responsiveness identified in Brown and MacKay (2023) and Aparicio et al. (2021), who find that even firms with the most sophisticated pricing technology (e.g., Amazon) do not respond to competitors' price changes for several hours (on average).[45] We identify approximately 3.7 weather shocks per week and 0.9 responses for stations with structural breaks after their break date, compared to 0.5 for stations without breaks. Similarly, we identify an average of 12 oil price shocks per week and find that stations with breaks respond 1.2 times per week after their break date, compared to once for a station without a structural break.[46]

Figure D2 in Appendix D.3 confirms these effects, presenting estimates from a station-month level regression of: $y_{it} = \sum_t \gamma_t^1 D_t + \sum_t \gamma_t^2 D_t \times \text{Break}_i + \delta_i + \epsilon_{it}$, where D_t is a dummy for month t, δ_i are a set of station FE, and Break_t is a dummy equal to 1 for stations that experienced a structural break. γ_t^2 represents the average difference in outcomes between stations with and without best-candidate

different dates that may be identified as breaks. Examples of F-statistics distributions are in Figure D3.

[41]Since variation in responsiveness to oil price shocks appears to be less clear-cut in the data as compared to the other three measures, we also test a definition of adoption that excludes it. Our main results hold.

[42]The number of price changes increases slightly throughout the sample period, from 4.8 to 5.3.

[43]This frequency of price changes is similar to the most rapid (hourly) average price change frequencies identified in online markets by Brown and MacKay (2023) and Aparicio et al. (2021). See also Musolff (2022) who finds that sellers on Amazon using third party repricers adjust prices every 0.29 hours.

[44]The difference in raw averages is larger than difference in Fig. D2 since Fig. D2 accounts for station fixed effects that absorb some of the heterogeneity. We should also note that average response speed hides substantial heterogeneity. We show a substantial increase in very rapid (five minute) responsiveness to competitors' price changes in duopoly and triopoly markets in Section 6, consistent with the description of retail gasoline algorithms in Section 3.2.

[45]Online pricing in multiproduct markets may be more complex than offline pricing in the relatively homogenous retail gasoline market, but online retailers should have data that are at least as good as German gas station data. Online retailers can easily and continuously scrape competitor prices and demand proxies such as sales ranks. Online technology should be at least as advanced as offline technology, so this appears to be the frontier of algorithmic pricing in retail markets in terms of the number of price changes and the average speed of competitive response.

[46]Since variation in responsiveness to oil price shocks appears to be less clear-cut in the data as compared to the other three measures, we also test a definition of adoption that excludes it. Our main results hold.

breaks in month t. The figure shows that for each measure stations that break behave differently than those that do not: 20% more changes, 10% faster response to rivals' price changes, 20% more frequent response to weather shocks and 5% more frequent response to oil shocks. Importantly, the changes between stations with without breaks all appear very rapidly around the middle of 2017.

Classification: Many factors may influence a single measure of pricing behaviour on its own, but breaking in multiple markers in close proximity should provide a strong indication of an actual change in pricing technology, which in our case is the adoption of AP pricing. We label a station as an adopter of AP software if it experiences best-candidate structural breaks in at least two measures of pricing behaviour within 4 weeks.[47] Our results are robust to stricter definitions of adoption.[48,49] We classify 2,728 stations as adopters. Figure 1 shows the distribution of the average break date for all adopters, defined as the average year-week between best-candidate break dates of the two or three measures in which a station experiences a significant break.[50] Over 50% of these average break dates occur in the middle of 2017, consistent with the supposed increased availability of algorithmic pricing software at this time in Germany (see Section 3.3).[51]

The stations we classify as adopters show meaningful differences in their pricing behaviour compared to stations without best-candidate structural breaks and stations with best-candidate structural breaks that are not classified as adopters. Figure 2 compares outcomes between adopter and non-adopter stations throughout our sample period.

We also find ex-ante average differences in local demographic characteristics and local markets between stations labelled as adopters and those not.[52] In Table 3 we find statistically significant differences in market characteristics between adopter and non-adopter stations before any adoption

[47] Any combination of two measures will result in a station being classified as an adopter.

[48] In Appendix G.3 we require stations to experience best-candidate breaks in at least 2 of the 4 measures within 2 weeks. We also include an additional definition that only labels stations as adopters if they experience multiple best-candidate breaks in 2 out of 3 measures (excluding rival response time and oil shock responsiveness), if they experience best-candidate breaks in Diesel, or best-candidate breaks in *both* E5 and Diesel within 4 weeks.

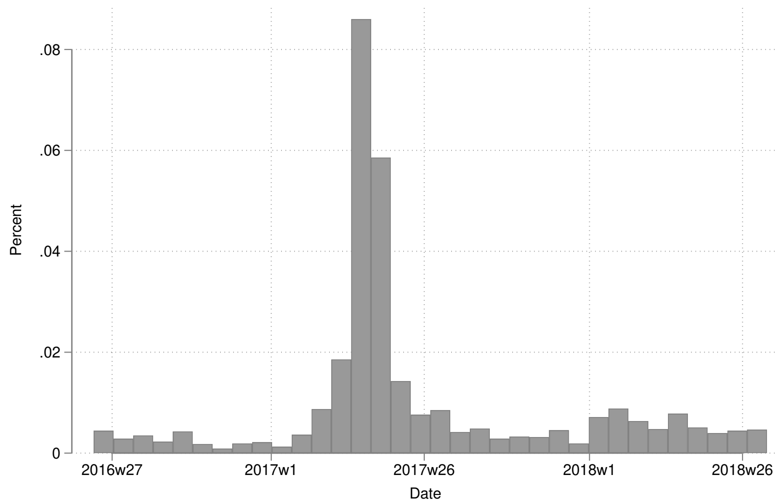
[49] In Appendix D.5 we present information on the break combinations across the four measures.

[50] This is a conservative approach. We may be "missing" some adopters, either due to measurement errors in our markers or to other signs of adoption that we did not consider. As a result some adopters are classified as non-adopters, biasing our station-level estimates towards zero and under-state the true effects of adoption.

[51] A possible concern with our classification is that non-adopting stations may be mistakenly labelled adopters if response to an adopting rival's pricing makes them behave as though they also adopted. This is not a regular occurrence. We observe a large number of duopoly markets where one station is classified as an adopter and not its competitor. Among 717 duopoly markets with full data in December 2018, 544 had no adopters, 142 had at most one, and 31 had two. More generally, Figure D5 in the appendix shows the distribution of adoption shares in markets with more than one adopting station in December 2018 (the last month in our data). There are 1,700 such markets out of a total of more than 4,000, and relatively few where adoption shares are higher than 50%.

[52] We do not observe individual station characteristics.

Figure 1: Frequency of Average Break Date for Measures Breaking Within 4 Weeks (2,728 stations)



Notes: This histogram shows the distribution of dates at which stations are labelled as adopters. We define an adoption date as the average best-candidate break date among the at least 2 best-candidate break dates for one of four measures described in Section 4.3.

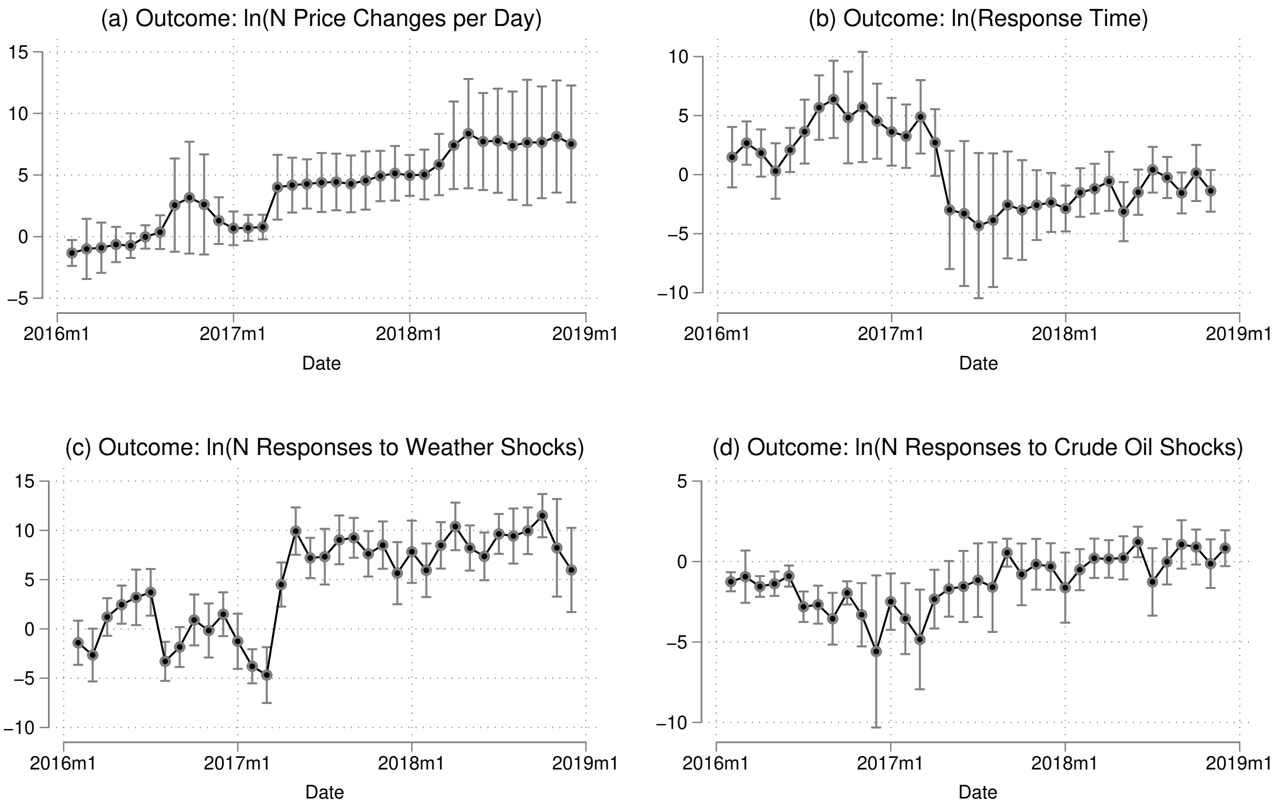
takes place (in 2016). Adopter stations are located in denser areas with different demographic profiles. They also face more competition, suggesting that adoption decisions may be endogeneous, with stations choosing to adopt in response to observable and unobservable market conditions.

4.3.2 Brand-Level Adoption

We do not observe if a brand entered into strategic partnership with an AP software provider. However, we can use findings from the station-level classification to infer brand-level adoption. We use a probabilistic definition, computing the probability that a brand adopted by time t as the percentage of a brand's stations classified as adopters by t. This captures underlying brand-level decisions. As described in Section 3.3, brand-level decisions should facilitate the adoption by individual stations. A brand for which a small percentage of stations adopted by t is unlikely to be an adopter at t, while a brand for which a large percentage of stations adopted is more likely to be an adopter.[53]

[53] Alternative definitions could classify a brand as an adopter if *any* one of its stations is classified as adopting, or only after *all* of its stations are. These alternatives do not reflect technology adoption in this market: brand adoption is *not a necessary condition* for station adoption. Many providers of algorithmic pricing software cater to small or medium enterprises (e.g., Prisync.com or Comptera.net). *a2i*'s 2017 advertisements target individual station owners and emphasize that all stations, regardless of brand, can adopt their technology. Defining a brand as an adopter if *any* one of its stations is classified as an adopter would be sensitive to outliers and amplify noise from our station-level adoption measure. Defining a brand as an adopter only if *all* of its stations adopt is inconsistent with the history of

Figure 2: % Difference Between Adopters and Non-Adopters



Notes: Each panel shows γ_t^2 estimates and their 95% confidence intervals from a station-month level regression of: $y_{it} = \sum_t \gamma_t^1 D_t + \sum_t \gamma_t^2 D_t \times \text{Adopter}_i + \delta_i + \epsilon_{it}$, where D_t is a dummy for month t, δ_i are a set of station FE, and Adopter_t is a dummy equal to 1 for stations that is labelled as an adopter. γ_t^2 represents the average difference in outcome y in month t between a station that is labelled as an adopter of AP and a station that is not labelled as an adopter. The outcome in panel (a) is the natural log of the average number of daily price changes a station has. The outcome in panel (b) is the natural log of the average response time to a rival's price change. The outcome in panel (c) is the natural log of the average number of station responses to a weather shock. The outcome in panel (d) is the natural log of the average number of stations responses to a crude oil price shock. Additional controls in that regression include the number of competitors in the market, average number of competitors' price changes.

Figure 3 shows the evolution of the share of a brand's adopting stations throughout our sample period. Panel (a) displays results for the top 5 brands (by station count). Adoption happens at a staggered rate that varies across brands. All brands experience an increase in adoption around early/mid 2017, likely reflecting the increased availability of the technology. Aral is an early adopter, with nearly 30% of its stations adopting by mid 2017. Total's and Avia's adoption rates increase at

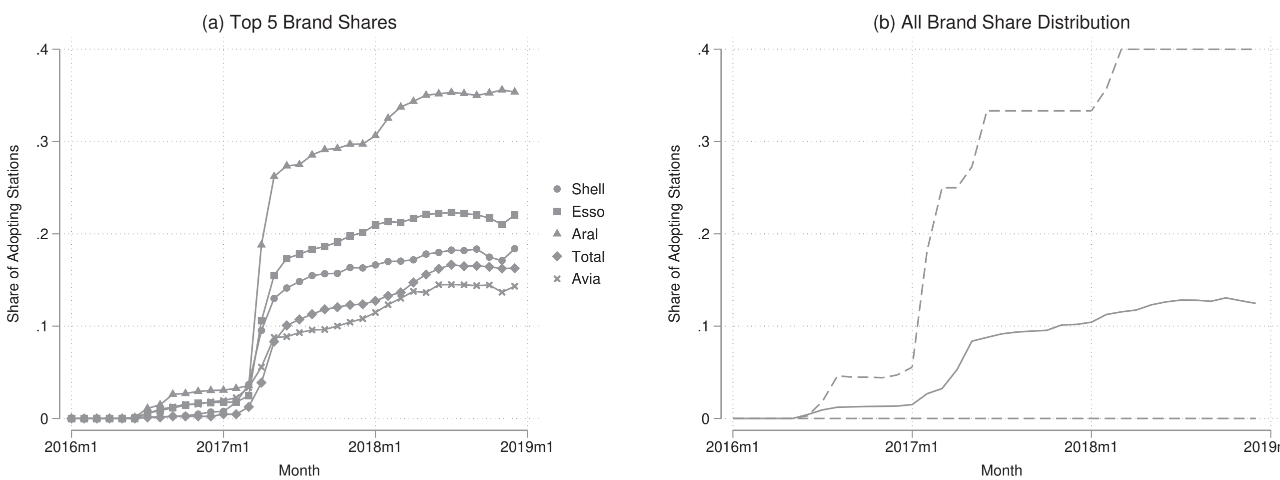
technology adoption in the market. As explained in Section 3.3, brand subsidies to stations for technology adoption are often incomplete and technology adoption is highly staggered. In Appendix F we show that it took years for a substantial share of gasoline stations belonging to top brands to adopt electronic payments in the 1990s.

Table 3: Adopter and Non-Adopter Station Characteristics in 2016

	(1)
Outcome:	Will Station i Adopt AP?
Population Density	0.00003***
	(0.00001)
ln(Population)	0.00443
	(0.03513)
Median Population Age	0.00707***
	(0.00211)
Employment Share	0.09257
	(0.07782)
ln(region GDP)	0.00056
	(0.03241)
N Competitors in Market	0.00297*
	(0.00165)
Observations	165,810

Notes: The sample for this regression includes gas station/month observations from January 2016 until December 2016. The outcome is a dummy variable equal to 1 if the station will eventually be labelled as an adopter in 2017 or 2018, and zero otherwise. Population Density, ln(Population), Median Population Age, Employment Share and ln(regional GDP) are all computed at the NUTS3-year level. "N Competitors in Market" is equal to the number of other stations present in the market of station i in month t. We include month fixed effects. Standard errors clustered at the market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3: Share of Brand Stations that are AP Adopters



Notes: Panel (a) shows the share of brand stations that are AP adopters in each month for the top five brands in our data (by count of stations). Panel (b) shows the distribution of brand adoption shares for all 258 brands. The solid line in Panel (b) shows the mean brand adoption rate, while the area between the dashed lines shows the distribution of brand adoption shares between the 5th and 95th percentiles (for each month, we calculate the 5th and 95th percentile brand adoption share).

a steadier (albeit slower) pace compared to other brands. The heterogeneity in adoption rates across brands suggests there is a brand-specific component to AP adoption, possibly reflecting that some brands were more likely to support the new technology. None of top 5 have adoption rates over 40% by the end of the period. Panels (b) and (c) display results for the top 20 brands and for all brands, respectively. The share of adopters for smaller brands is typically lower and occurs later than for the large brands: the mean adopter share for top 5 brands is 21% by December 2018, while for brands outside the top 5 it is 12%.[54] This likely reflects the better support that larger brands can provide to their stations, lowering adoption costs. Panel (c) presents the mean for all brands (the red line) and the area captured by the 5th and 95th percentiles. The difference between the 5th and 95th percentile grows over time, suggesting heterogeneity in within-brand adoption rates is growing.

The pattern in Figure 3 can be compared to the observed timing of electronic-payment adoption at Canadian gasoline stations in the 1990s (see Figure F1 in Appendix F). Despite differences in time, geography and technology, we also see a staggered pattern of technology adoption that appears to be highly brand specific, suggesting that our AP classification captures technology adoption.[55]

5 Results – Effects of AI Adoption

This section presents our estimates of the effects of algorithmic pricing adoption on prices and margins in the German retail gasoline market.

5.1 Impact of Adoption on Station Outcomes

5.1.1 OLS Estimation and Results

Our objective is to capture the effects of station i's adoption of algorithmic pricing on average daily margins (above regional wholesale prices) and prices in period t. We use a station-month specification where we calculate average *monthly* daily outcomes and characteristics for each station in month t ($t \in \{1, 2, ..., T\}$). Our OLS specification is as follows:

$$y_{it} = \alpha_i + \alpha_t + \beta(\text{Adopter} \times \text{Post Adoption})_{it} + \gamma X_{it} + \epsilon_{it}, \quad (1)$$

[54] 7% of single-station independents are labelled as adopters by the end of 2018.

[55] In Appendix B.1 we also provide an example of staggered role out of algorithmic pricing at a brand. According to news articles, Lekkerland, a company that operates gas station convenience stores in Germany and teams up with brands/stations, adopted dynamic pricing in its stores around 2017. The articles mention explicitly two stores with attached gas stations that were the test cases for the introduction of dynamic pricing technology in 2017.

where y_{it} is the outcome variable for station i in time t, α_i and α_t are, respectively, station and time fixed-effects, and $(\text{Adopter} \times \text{Post Adoption})_{it}$ is a dummy variable equal to 1 if station i has adopted algorithmic pricing before period t and 0 otherwise. X_{it} are time-varying station-specific controls (local demographics and weather). X_{it} also includes the number of other gas stations that are in the same market as station i. The key coefficient in this regression is β, which captures the effect of AI adoption on y_{it}. Columns (1) and (2) in Table 4 present the main average OLS station-level estimates. These show that the adoption of AP increases average margins and prices by approximately 0.1 cpl.

We are concerned that OLS estimates are biased because of endogeneity. The OLS specification assumes that adoption is exogenous and as-good-as-random (conditional on observables). Despite the inclusion of fixed effects and a rich set of station-level observables, this is likely not the case. AP adoption could be correlated with unobservable time-varying station characteristics (ϵ_{it}). The adoption of any new technology but especially of new pricing technology is an important and potentially costly decision with long-term consequences. Adopters are going to be stations that expect limited profits if they do not adopt the new software and that can afford to make the investment. This would mean that stations that have had better unobservable shocks in the past and that expect worse future unobservable shocks will be more likely to adopt - such patterns in the unobservables would generate negative correlation between the adoption decision and the ϵ_{it} shocks. Such stations would also have different market outcomes. This would invalidate a difference-in-differences (or event study) research design and attenuate estimated adoption effects towards zero.

Table 3 shows that adopter and non-adopter stations are very different in their local market demographics and in their competitive environment. They are also likely to be different in their unobservable characteristics. We provide further evidence of endogeneity in the OLS regressions using a formal test of parallel trends between adopters and non-adopters before adoption.[56] We estimate the following specification:

$$\begin{aligned} y_{it} = \alpha_i + \alpha_t &+ \beta_1(\text{Adopter} \times \text{Post Adoption})_{it} + \beta_2(\text{Adopter} \times \text{1-6 Months Pre})_{it} \\ &+ \beta_3(\text{Adopter} \times \text{7-12 Months Pre})_{it} + \beta_4(\text{Adopter} \times \text{13+ Months Pre})_{it} + \gamma X_{it} + \epsilon_{it}, \end{aligned} \tag{2}$$

where the key coefficients to estimate are the time varying βs, which represent the differences between adopter and non-adopter stations at various times. For example, $(\text{Adopter} \times \text{7-12 Months Pre})_{it}$ is

[56] Such endogeneity would be mitigated by a "flat" specification where we do not consider time-varying adoption but simply calculate average outcomes and characteristics for adopters and non-adopters before and after the middle of 2017 ($t \in \{\text{pre mid-2017, post mid-2017}\}$). However, even this specification would be subject to a downward bias if time-varying outcomes are correlated with time-varying shocks. See additional discussion in Online Appendix E.1.

a dummy equal to 1 for adopter stations 7-12 months before their actual adoption. The baseline period for each adopter station i is the month immediately before adoption.

Columns (3) and (4) in Table 4 present the time-varying β coefficients, and reveal statistically significant differences in mean prices and margins between adopter and non-adopter stations prior to adoption. Adopters had margins that were 0.2 cpl higher than for comparable non-adopters 7-12 months before they adopted. Similarly, adopters had prices that were 0.2 cpl higher than for non-adopters 7-12 months before adoption and 0.3 cpl a year before adoption. This suggests the parallel trends assumption does not hold in our setting, invalidating a difference-in-differences / event-study based research design. These results also confirm the intuition described above: adoption of AP technology is a strategic decision made by stations that can afford to adopt and that expect limited profits if they do not. These stations likely had better unobservable shocks (and higher margins/prices) some time before adoption and may expect worse shocks in the future.

Table 4: OLS Station-Level Estimates

	(1)	(2)	(3)	(4)
Outcome:	Mean Margin	Mean Price	Mean Margin	Mean Price
Adopter × post-Adoption	0.001**	0.001**	0.001***	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Adopter × 1-6 Months pre-Adoption			0.000	0.000
			(0.000)	(0.000)
Adopter × 7-12 Months pre-Adoption			0.002***	0.002***
			(0.001)	(0.001)
Adopter × 13+ Months pre-Adoption			0.000	0.003***
			(0.001)	(0.001)
Non-Adopter Mean Outcome	0.0821	1.361	0.0821	1.361
Station FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES
Observations	478,172	478,172	309,280	309,280

Notes: Sample is gas station/month observations from January 2016 until December 2018 in Columns (1) and (2). Sample for columns (3) and (4) only includes stations with a history of more than 12 months in the data. Margins are computed above wholesale gasoline prices at a regional terminal nearest to station i. Mean Margin/Price is the monthly average pump price for station i in month t. "Adopter × post-Adoption" is a dummy equal to 1 in month t if the gas station experienced a structural break in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. "Adopter × X Months post-Adoption" is a dummy equal to 1 in month t if the gas station experienced a structural break in at least 2 of 4 relevant measures X months prior to month t. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the number of stations belonging to station i's brand in month t, for the number of competitors in the market and for the number of adopting competitors in the market. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. Market clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5.1.2 IV Estimation and Results

Since we are not able to use an event study research design, we turn to an instrumental variables approach to identify the causal effect of AP adoption on mean margins and prices. We need to instrument for $(\text{Adopter} \times \text{Post Adoption})_{it}$. Our instrument should be correlated with an individual station's adoption decision but should not be affected by station-specific unobservable shocks. We propose *brand-HQ level adoption* as an instrument.[57] As explained in the previous section, we measure brand-level adoption by computing the share of stations belonging to each brand that have been identified as AP adopters by month t. For station i at time t our IV is the share of stations in station i's brand that adopted algorithmic pricing by time t. We exclude station i from this share.

The intuition behind this instrument is similar to the commonly used Hausman-Nevo instruments (e.g., Dubois and Lasio 2018). These instruments are valid if they appropriately recover common cost shocks across groups of observations, for example by using prices from "nearby" observations as an instrument for own prices.[58] In our case, adoption costs should be correlated for stations within a brand because of the aforementioned brand subsidies for technology adoption. Brand-level decisions likely influence the adoption decisions of individual stations (see Section 3.3 for additional discussion). Brands provide individual stations with employee training, technical support and maintenance (Convenience.org). This happens for both chain-operated stations as well as for more independent lessees. For previous waves of technology adoption (such as electronic payments) brands also directly subsidized some costs associated with required station upgrades. This support is important for drastic technical changes such as AP adoption. At the same time, brand level decisions should not be influenced by station-level specific demand or supply conditions.[59]

Station-level IV estimates are presented in Table 5. Column (1) shows the first stage of the IV

[57] As a robustness check, we propose an alternative set of instruments: the availability of broadband internet in the local area around a gas station. As with *brand-HQ level adoption*, the availability of broadband internet should have an effect on a station's decision to adopt AP software. Most AP software are "cloud" based and require constant downloading and uploading of information. Without high speed internet, adoption of such software is not particularly useful for a station. However, the availability of broadband internet in the region should be uncorrelated with station unobservables after conditioning on observable local characteristics. Our estimates with these IVs are qualitatively similar to our main estimates. See Table G6 for results and Appendix G.4 for additional discussion. We also test a "placebo IV" which uses the brand HQ-level adoption decision by a random brand (not the brand of station i) as an instrument and find null effects. Additional discussion is also in Appendix G.4.

[58] Dubois and Lasio (2018) effectively use the prices of pharmaceutical molecule combinations in Germany, Italy, Spain and the UK as an instrument for the prices of the same molecule combination in France.

[59] Table D1 shows that, conditional on brand size, brand adoption shares are uncorrelated with market characteristics. We also test a "placebo IV" which uses the brand HQ-level adoption decision by a random brand (not the brand of station i) as an instrument and find null effects. These findings make sense if the brand-level IV actually recovers each brand's costs, rather than some other time-varying common cost shocks. Additional discussion is in Appendix G.4.

Table 5: IV Station-Level Estimates

Outcome:	(1) 1st Stage Adopter	(2) 2SLS Mean Margin	(3) 2SLS Mean Price	(4) Reduced Form Mean Margin	(5) Reduced Form Mean Price
Adopter × post-Adoption		0.012***	0.012***		
		(0.002)	(0.002)		
Share Brand Adopters	0.660***			0.008***	0.008***
	(0.041)			(0.001)	(0.001)
Non-Adopter Mean Outcome		0.0821	1.361	0.0821	1.361
Station FE	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES	YES
Observations	448,221	448,221	448,221	448,221	448,221

Notes: Sample is gas station/month observations from January 2016 until December 2018. Margins are computed above wholesale gasoline prices at a regional terminal nearest to station i. Mean Margin/Price is the monthly average pump price for station i in month t. Margins are computed above wholesale gasoline prices at a regional terminal nearest to station i. Mean Margin/Price is the monthly average pump price for station i in month t. "Adopter × post-Adoption" is a dummy equal to 1 in month t if the gas station experienced a structural break in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. "Share Brand Adopters" is the excluded instrument used in the 2SLS regression. It is equal to the share of stations that belong to the brand of station i that adopted by period t (excluding i). Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the number of stations belonging to station i's brand in month t, for the number of competitors in the market and for the number of adopting competitors in the market. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. Market level clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

regression. The first stage is strong, with an F-statistic of 35. A 10% increase in the number of other stations affiliated with station i's brand (excluding i) that adopt by period t increases the probability that i adopts by t by 66%, consistent with intuition that adoption of AP is at least in part a brand-level decision. Columns (2) and (3) of Table 5 show 2SLS estimates with margin and price outcomes, and Columns (4) and (5) show the reduced form estimates. Column (2) shows that mean margins increase by 1.2 cpl on average after AP adoption, or about 15% relative to the average non-adopter margin of 8.2 cpl.[60] Prices also increase by 1.2 cpl after adoption. Reduced form estimates confirm that there is a direct positive correlation between the instrument and the main outcomes.

The 2SLS estimates are approximately ten times larger than OLS estimates. One potential reason for this difference is measurement error in our adoption classification. Indeed, if we weight observations by the inverse of the noise (uncertainty) created by our structural break tests, we obtain estimates that are approximately three times our baseline unweighted OLS estimates, bridging part,

[60] 2SLS regressions using alternative instruments based on broadband availability and quality also show that mean margins and mean prices increase after adoption (see Table G6). See Appendix G.4 for additional discussion.

but not all, of the gap between the OLS and IV results and indicating that measurement error is responsible for some of the downward attenuation of the OLS results.[61] The bulk of the difference can be attributed to the fact that adopters will be stations that expect limited profits if they do not adopt the new software and that can afford to make the investment (stations that have had better unobservable shocks in the past and expect worse future unobservable shocks) since such patterns in the unobservables would generate negative correlation between the adoption decision and the ϵ_{it} shocks. Anecdotal evidence supports this: stations struggling to maintain high margins benefit most from adoption.[62] It is also the case that OLS is estimating the average treatment effect over the entire population, while the IV is estimating the local average treatment effect. That is, the instrument affects the behavior of a subgroup of stations for whom the returns to adoption are greater than average, or put differently, for the stations whose choice of treatment was affected by the instrument. The full population used in the OLS includes (i) stations that could never adopt because the fixed costs of doing so are prohibitive, (ii) stations for which adoption was almost automatic because they were heavily subsidized by the brand, and (iii) stations in between. The IV zooms in the stations in between, estimating a local treatment effect that is naturally much larger than the OLS effect.

Reassuringly, the magnitude of the 2SLS effects is in line with estimates of the effects of AP software on gas station profitability released by software providers. The Brazilian pricing start-up Aprix estimates that gas stations using its AI-based pricing software increased their gross profits by approximately 10% (towardsdatascience.com). a2i similarly estimated that its software could increase station profits by at least 5% (a2i.com).

Moreover, we confirm that our IV approach resolves concerns about diverging parallel pre-trends between adopters and non-adopters. First, we verify parallel trends in the outcome for the instrument - testing for parallel trends in the reduced form. We do this by estimating the same regression as

[61] See Table E2 in Appendix E.

[62] See towardsdatascience.com, which describes adoption motives in Brazil, pointing to retail margins restrained by both sides of the supply chain. The upstream segment reduces station margins through increased fuel purchase costs, while at the same time consumers in Brazil were facing tighter budget and becoming more price sensitive, making it difficult for stations to pass on cost increases. AI-driven algorithmic pricing can help stations "survive in the new highly competitive environment, turning the threat into an opportunity." Another article quotes the president of retailer operating in California: "Maybe the manager hasn't done a price survey yet, or we need them to update it because the markets are moving intraday rapidly during the hurricanes, and we need to do price surveys more than once a day because we are struggling to keep up with what is happening...That's the reason why AI is interesting: Is there a way for us to get information faster so we can react faster to changes in the market?" See cspdailynews.com.

Table 6: IV Station-Level Estimates: Pre-Trends

Outcome:	(1) 2SLS Mean Margin	(2) 2SLS Mean Price	(3) Reduced Form Mean Margin	(4) Reduced Form Mean Price
Adopter × post-Adoption	0.013***	0.009***		
	(0.003)	(0.003)		
Adopter × 1-6 Months pre-Adoption	-0.000	0.000		
	(0.001)	(0.001)		
Adopter × 7-12 Months pre-Adoption	0.001	0.001		
	(0.001)	(0.001)		
Adopter × 13+ Months pre-Adoption	0.000	0.002		
	(0.001)	(0.002)		
Share Brand Adopters × post-Adoption			0.007***	0.005***
			(0.002)	(0.002)
Share Brand Adopters × 1-6 Months pre-Adoption			-0.002	-0.000
			(0.003)	(0.004)
Share Brand Adopters × 7-12 Months pre-Adoption			0.005	0.006
			(0.004)	(0.005)
Share Brand Adopters × 13+ Months pre-Adoption			0.003	0.017
			(0.010)	(0.012)
Station FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES
Observations	290,585	290,585	290,585	290,585

Notes: Sample is gas station/month observations with a history of more than 12 months in the data. Margins are computed above wholesale gasoline prices at a regional terminal nearest to station i. Mean Margin/Price is the monthly average pump price for station i in month t. "Adopter × post-Adoption" is a dummy equal to 1 in month t for stations labelled as adopters after they adopted. "Adopter × X Months pre-Adoption" is a dummy equal to 1 for stations that we labelled as adopters in the X months prior to their adoption. "Share Brand Adopters" interactions are the excluded instrument used in the 2SLS regression in Columns (1) and (2). They are equal to the share of stations that belong to the brand of station i that adopted by period t, interacted with dummies reflecting whether the station has adopted AP or if it is going to adopt AP in the future. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the number of stations belonging to station i's brand in month t, for the number of competitors in the market and for the number of adopting competitors in the market. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. Market level clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Equation (2) but with the instrument in place of the endogenous "treatment" variable:

$$\begin{aligned} y_{it} = \alpha_i + \alpha_t &+ \beta_1(\text{Share Brand Adopters} \times \text{Post Adoption})_{it} \\ &+ \beta_2(\text{Share Brand Adopters} \times \text{1-6 Months Pre})_{it} \\ &+ \beta_3(\text{Share Brand Adopters} \times \text{7-12 Months Pre})_{it} \\ &+ \beta_4(\text{Share Brand Adopters} \times \text{13+ Months Pre})_{it} + \gamma X_{it} + \epsilon_{it}, \end{aligned} \tag{3}$$

As above, the baseline period is a month before adoption for adopting stations.

Estimates from this regression are presented in Table 6 and they show that there is no correlation between the outcome variables and the instrument interacted with pre-adoption time dummies, suggesting that parallel trends in the instrument hold. We also estimate an IV version of Equation (2), where we instrument for each lead variable with an appropriately constructed instrumental variable. For example, the variable (Adopter × 13+ Months Pre)$_{it}$ is instrumented with (Share Brand Adopters × 13+ Months Pre)$_{it}$. 2SLS estimates from this regression are also in Table 6 and they similarly show that there is no correlation between instrumented leads of the treatment variable and the outcomes. These results suggest that our instrument helps to effectively correct for the endogeneity between station-level adoption and other station-specific unobservable factors that can affect station-level margins and prices.

5.2 Impact of Adoption on Competition

The previous section presented causal estimates of the effects of AP on station-level prices and margins. AP can increase margins and prices through a reduction in competition and increased market power but there can be other reasons for such changes. An algorithm could better understand underlying fluctuations in wholesale prices, or identify how demand elasticity changes over time and adjust prices accordingly. In this section we describe how we isolate the effects of adoption on competition. We start by splitting the sample according to whether or not the station is a monopolist (e.g., without any nearby competitors). If adoption affects competition, the impact may be stronger for non-monopolists than for monopolists. Since there is a lack of clear theoretical predictions on how adoption should affect average prices in monopoly versus non-monopoly markets if its only function is to improve a seller's ability to tailor prices to time-varying demand or cost conditions,[63] in a second step we evaluate how adoption affects strategic interaction between stations holding market structure roughly constant by focusing on small oligopoly markets (duopoly and triopoly) and testing whether the adoption of only a subset or of all competitors triggers changes in outcomes.

[63]Monopolists can typically achieve higher profits by price discriminating. Extending these findings to the case of oligopoly is not straight forward since, in addition to market-level elasticities, firm-level elasticities must also be taken into account (see Holmes 1985). As a result, firms might even be worse off under price discrimination than uniform pricing (see Thisse and Vives 1988 and Corts 1998). More generally, there can be considerable variation in potential effects of adoption in both monopoly and oligopoly depending on (i) what the algorithm does, (ii) the informational environment, and (iii) supply and demand fundamentals (market- vs firm-level elasticities).

5.2.1 Impact of Adoption on Monopolist and Non-Monopolist Stations

To test whether any observed changes in prices and/or margins come from a reduction in competition and increased market power, or from a better understanding of underlying fluctuations in wholesale prices and consumers' demand elasticity, we look separately at stations that are monopolists and stations that are not.[64] If adoption does not change competition but benefits station operations in other ways, we might expect to see effects for monopolist adopters. If adoption also affects competition, we should expect to see additional non-zero effects for non-monopolist adopters on top of the effects for monopolist adopters. If adoption *only* affects competition, we should expect to see zero effects for monopolist stations and non-zero effects for non-monopolists.

Table 7: IV Station-Level Estimates by Market Structure

Outcome:	(1) Mean Margin	(2) Mean Price	(3) Mean Margin	(4) Mean Price
	Sample: Monopolist Stations		Sample: Non-Monopolist Stations	
Adopter × post-Adoption	0.004	-0.004	0.012***	0.013***
	(0.012)	(0.009)	(0.002)	(0.002)
Non-Adopter Mean Outcome	0.0850	1.363	0.0825	1.361
Station FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES
Observations	18,556	18,556	429,181	429,181

Notes: Sample includes gas station/month observations from January 2016 until December 2018, split up into two subsamples: one subsample only includes stations that have no competitors in their market. The other subsample includes only stations that have one or more competitors in their market. Margins are computed above wholesale gasoline prices at a regional terminal nearest to station i. Mean Margin/Price is the monthly average pump price for station i in month t. "Adopter" is a dummy equal to 1 in month t if the gas station ever experienced a structural break in at least 2 of 4 relevant measures, and "post-Adoption" is a dummy equal to 1 for adopter stations after we label them as adopters. "Share Brand Adopters" is the excluded instrument used in the 2SLS regression. It is equal to the share of stations that belong to the brand of station i that adopted by period t. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the number of stations belonging to station i's brand in month t, for the number of competitors in the market and for the number of adopting competitors in the market. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Results of our IV regression for the two subsamples are presented in Table 7. We find that non-monopolist stations are driving the margin increase, with mean margins increasing for non-monopolist adopters by 1.2 cents post-adoption (15%), compared to a small and non-statistically

[64] Markets are defined according to a clustering algorithm based on driving time between stations (Appendix C). As a robustness check, we use an alternative definition based on ZIP codes. See Appendix G.2.

significant change for monopolist adopters. Price effects are similar: mean monthly prices increase by 1.3 cents per litre for adopting non-monopolist stations, but not at all for monopolist stations.[65]

The null estimated effects of adoption for monopolist stations naturally lead to questions about why they would have adopted in the first place. Of course, the null effect only shows that there is no change on the *mean*. There may well be substantial changes in prices at different points during the day, reflecting a monopolist station's ability to better price discriminate (as in Dubé and Misra 2023). These changes could average out to a daily null effect. We test for this by using an alternative set of outcomes: mean monthly prices at different points during the day. We calculate the price of each station at 9am, 12pm, 5pm and 7pm at each non-holiday weekday and then average out across a month. As mentioned above, gas prices in Germany follow a decreasing pattern throughout the day with high prices in the morning that gradually fall until the evening. A comprehensive discussion of price cycles in the German gasoline retail market and the effects of algorithms on these cycles requires formal modelling of both pre-algorithmic and post-algorithmic pricing behaviour and so is outside the scope of this paper. Nonetheless, our results suggest that, on average, non-monopolist AP adopters increase their prices during the day such that the price pattern becomes flatter and average daily prices increase.[66] Monopolist AP adopters show a different pattern, likely reflecting an improved ability to temporally price discriminate and improve overall profits.

IV regression results at different times of the day are presented in Table 8. As in Table 7, the sample is split between monopolist and non-monopolist stations (aggregate results are presented in Table E4 in the appendix), and, consistent with the findings for mean daily prices, estimates show substantial differences in the effects of AP adoption between monopolists and non-monopolists. For non-monopolist adopters, prices do not change in the morning (relative to non-adopters) but then increase progressively throughout the day, with the highest price increase at 5pm, generating the flatter pattern we just mentioned. For monopolist adopters, prices fall on average at 9am relative to non-adopters, and increase on average at 5pm. The monopolist results hint at the potential welfare improving effects of AP adoption through better price-targeting across demand conditions. AP software may learn that morning prices are "too high" and that reducing them will increase monopolist station profits (in addition to consumer welfare). Although human- or rule-based pricing allowed for multiple price changes throughout the day, price adjustments were likely more costly than with AI-powered algorithms, permitting an additional price change (Garcia et al. 2021). AI-powered

[65] In Appendix G.7 we estimate a pooled version of this regression and results are unchanged.

[66] It is possible that tacit collusion was more difficult to sustain later in the day when demand was elevated and so this is when the benefits from adoption are greatest.

Table 8: IV Station-Level Estimates by Market Structure - Time Specific Prices

Outcome:	(1) Mean 9am Price	(2) Mean 12pm Price	(3) Mean 5pm Price	(4) Mean 7pm Price
	Sample: Monopolist Stations			
Adopter × post-Adoption	-0.029**	0.019	0.041**	0.012
	(0.013)	(0.014)	(0.017)	(0.010)
Observations	18,554	18,556	18,556	18,556
Non-Adopter Mean Outcome	1.382	1.358	1.347	1.344
	Sample: Non-Monopolist Stations			
Adopter × post-Adoption	-0.002	0.033***	0.050***	0.024***
	(0.003)	(0.003)	(0.004)	(0.003)
Observations	429,094	429,160	429,181	429,181
Non-Adopter Mean Outcome	1.381	1.356	1.345	1.341
Station FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES

Notes: Sample includes gas station/month observations from January 2016 until December 2018, split up into two subsamples: one subsample only includes stations that have no competitors in their market. The other subsample includes only stations that have one or more competitors in their market. Mean Price is the monthly average pump price for station i in month t at a particular time. " Adopter × post-Adoption " is a dummy equal to 1 in month t if the gas station experienced a structural break in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. "Share Brand Adopters" is the excluded instrument used in the 2SLS regression. It is equal to the share of stations that belong to the brand of station i that adopted by period t. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the number of stations belonging to station i's brand in month t, for the number of competitors in the market and for the number of adopting competitors in the market. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

algorithms may also help monopolists price discriminate better, which is again profit increasing though not necessarily consumer welfare decreasing. Unfortunately we do not observe intra-day regional wholesale prices and so cannot compute margins.

For non-monopolist adopters, the potential strategic effects arising from AP could work against the beneficial effects and against targeted price discrimination. Adopting stations may not want to adjust prices to better target consumers if they know competing algorithms are responsive to price changes. This is consistent with our findings that prices are more stable during the day for AP adopting non-monopolists. Our findings may also suggest that the value of deviation and punishment are changing throughout the day. The benefits to adopters are stronger later in the day, roughly around the time of the evening commute, possibly implying that, without algorithmic pricing, tacit collusion was harder to sustain at this time (although the benefits of adoption are higher even for monopoly stations at 5pm).

Our timing results can be related to the theory literature on temporal price discrimination. Conlisk et al.. (1984) explain how a monopolist can periodically lower its price to discriminate between low- and high-valuation shoppers, the former being more inclined to wait for lower prices than the latter. The monopolist will charge high prices most of the time, selling only to high-valuation shoppers, and will periodically lower its price to attract low-valuation consumers. Sobel's (1984) model extends Conlisk et al. (1984) to allow for multiple sellers that compete for the low-valuation shoppers, and finds similar results. Pesendorfer (2002) and Chevalier and Kashyap (2018) develop related models. The pricing patterns in the German retail gasoline market are consistent with these predictions. Consumers with different driving habits, and therefore different inventory costs, can be thought of as having different valuations and more or less incentive to time the price fluctuations. Our monopoly results imply that AP technologies may make it easier for firms to better coordinate their sales along the daily path.

5.2.2 Impact of Adoption on Duopoly/Triopoly Markets

In a more direct test of theoretical predictions about the effects of AP adoption on competition, we compare outcomes between adopting and non-adopting oligopoly *markets*.[67] We focus on duopoly and triopoly markets, since most theoretical analysis is done for cases with few firms (i.e., Calvano et al. 2020, Miklós-Thal and Tucker 2019). As with our station-level estimates, we choose to use an IV specification in order to avoid endogeneity concerns. The second stage regression for market m in month t is as follows:

$$y_{mt} = \alpha_m + \alpha_t + \beta_1 \text{Not All Stations Adopted}_{mt} + \beta_2 \text{All Stations Adopted}_{mt} + \gamma X_{mt} + \epsilon_{mt}, \qquad (4)$$

where y_{mt} is the outcome variable for market m at time t, α_m and α_t are, respectively, market and time fixed-effects. The dummy "Not All Stations Adopted" is a variable equal to one if at least one, but not all, stations in a market are labelled as an adopter at time t. It is equal to zero if all stations are labelled as adopters or if no stations are labelled as adopters.[68] The variable "All Stations

[67] This analysis is done using our main market definition (i.e., clusters). As a robustness check, we use an alternative market definition based on ZIP codes. See Section 7 and Appendix G.2 for additional discussion.

[68] In duopoly markets, this variable can be expressed as $(\text{Adopter} \times \text{post-Adoption})_{1mt}(1 - (\text{Adopter} \times \text{post-Adoption})_{2mt}) + (\text{Adopter} \times \text{post-Adoption})_{2mt}(1 - (\text{Adopter} \times \text{post-Adoption})_{1mt})$, where 1 and 2 are the stations

Adopted" is equal to one in market m in month t if all stations in this market are adopters.[69] The two key coefficients in this regression are β_1 and β_2. β_1 captures the effects of AP adoption by *some of the firms* in a duopoly/triopoly market and β_2 captures the effects of market-wide AP adoption.[70]

Since there are two endogenous variables, we have two first stage regressions. Following the logic of our main station-level instruments, we construct two time-varying market-level IVs using brand-level adoption decisions.[71] In a duopoly market, the two instruments are functions of the brand-level adoption decisions for the brands of stations in market m:

$$IV^1_{mt} = \text{Share Brand Adopters}_{1mt}(1 - \text{Share Brand Adopters}_{2mt}) + \text{Share Brand Adopters}_{2mt}(1 - \text{Share Brand Adopters}_{1mt}) \tag{5}$$

$$IV^2_{mt} = \text{Share Brand Adopters}_{1mt}\text{Share Brand Adopters}_{2mt},$$

where Share Brand Adopters$_{1mt}$ is the share of other stations belonging to market m station 1's brand identified as AP adopters in month t. Share Brand Adopters$_{2mt}$ is similarly defined.[72]

The first-stage regressions are as follows:

$$\begin{aligned} \text{Not All Stations Adopted}_{mt} &= \alpha_m^{1st,1} + \alpha_t^{1st,1} + \pi_1^{1st,1} IV^1_{mt} + \pi_2^{1st,1} IV^2_{mt} + \kappa^{1st,1} X_{mt} + \mu_{mt} \\ \text{All Stations Adopted}_{mt} &= \alpha_m^{1st,2} + \alpha_t^{1st,2} + \pi_1^{1st,2} IV^1_{mt} + \pi_2^{1st,2} IV^2_{mt} + \kappa^{1st,2} X_{mt} + \mu_{mt}. \end{aligned} \tag{6}$$

In each we include market and time fixed effects and all time-varying controls.

Table 9 presents estimates of Equation (4) using the instruments defined in Equation (5) and

in market m and "(Adopter $\times$ post-Adoption)" is a dummy equal to one for adopting stations after adoption. The definition for triopoly markets is similar, but with a combination of three stations.

[69] In duopoly markets this variable can be expressed as:(Adopter $\times$ post-Adoption)$_{1mt}$(Adopter $\times$ post-Adoption)$_{2mt}$, where 1 and 2 are the stations in market m and "(Adopter $\times$ post-Adoption)" is a dummy equal to one for adopting stations after adoption. The definition for triopoly markets is (Adopter $\times$ post-Adoption)$_{1mt}$(Adopter $\times$ post-Adoption)$_{2mt}$(Adopter $\times$ post-Adoption)$_{3mt}$.

[70] This distinction is natural for duopoly markets, but triopoly markets can also be separated into those markets where fewer than 50% of stations adopted and markets where more than 50% of stations adopted (e.g., two stations are adopters and one stations not an adopter). We focus on market-wide adoption for two reasons: first, it allows us to aggregate effects across duopoly and triopoly markets. Second, and more importantly, there is substantial evidence that it is harder to sustain supra-competitive prices in markets with asymmetric firms (e.g., with "mavericks").

[71] As a robustness check for station-level estimates, we propose an alternative instrument: the availability of broadband internet in the local area around a gas station. This instrument would only work for market level data if the duopolists/ triopolists are in the same market but also have different broadband access/quality conditions. Our broadband access data is calculated at a coarse geographical level (NUTS2), so we are unable to use these instruments for market level data. See additional discussion in Appendix G.4.

[72] In a triopoly market, if we label Share Brand Adopters$_{imt}$ as B_{imt}, then $IV^2_{mt} = B_{1mt}B_{2mt}B_{3mt}$, and $IV^1_{mt} = B_{1mt}(1-B_{2mt})(1-B_{3mt}) + B_{2mt}(1-B_{1mt})(1-B_{3mt}) + B_{3mt}(1-B_{1mt})(1-B_{2mt}) + B_{1mt}B_{2mt}(1-B_{3mt}) + B_{1mt}B_{3mt}(1-B_{2mt}) + B_{2mt}B_{3mt}(1-B_{1mt})$.

market-level margins and prices as the outcome variables of interest.[73] 2SLS estimates are in Columns (1) and (2), first-stage estimates are in Columns (3) and (4), and reduced form estimates of regressing the instruments directly on the outcome of interest are in Columns (5) and (6). As was the case with the station-level instruments, the partial correlation between market-level instruments and the endogenous variables is strong.

Table 9: IV Duopoly and Triopoly Market Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	1st Stage	1st Stage	Reduced Form	Reduced Form
Outcome:	Mean Mkt Margin	Mean Mkt Price	Not all Stations Adopted	All Stations Adopted	Mean Mkt Margin	Mean Mkt Price
Not all Stations Adopted	-0.002	0.002				
	(0.005)	(0.005)				
All Stations Adopted	0.031*	0.061**				
	(0.018)	(0.024)				
IV^1			0.632***	-0.017	-0.002	0.001
			(0.088)	(0.025)	(0.003)	(0.003)
IV^2			-1.818***	1.230***	0.041**	0.070***
			(0.476)	(0.309)	(0.017)	(0.019)
Zero Adopter Mean Outcome	0.0857	1.355			0.0857	1.355
Market FE	YES	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES	YES	YES
Observations	49,431	49,431	49,431	49,431	49,431	49,431

Notes: The sample includes duopoly and triopoly market/month observations from January 2016 until December 2018. Outcome variable Mean Market Margin is the monthly average of mean market daily differences of pump prices for stations in market m in month t from wholesale price. Outcome variable Mean Market Price is the monthly average of mean market daily pump prices for stations in market m in month t . “Not all Stations Adopted ” is a dummy equal to 1 in month t if at least one station but not all stations in the market experienced a structural break in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. “All Stations Adopted” is a dummy equal to 1 in month t if *all* stations in the market experienced a structural break in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. Instruments for adoption are functions of the “share of brand adopters” of the stations in the market. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the sizes of the brands of the stations at month t. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Our estimates suggest that AP adoption by only some stations in duopoly and triopoly markets does not affect average market-level margins or prices relative to similar market where no stations adopted. However, market-wide AP adoption does affect market-level margins and prices. Mean market-level margins increase by 3.1 cents per litre after market-wide AP adoption. This is a substantial increase of 36% relative to the baseline. Similar effects are observed for market-level prices after market-wide adoption, with mean market prices increasing by 6 cents per litre.

One explanation for why we do not observe a change in mean market-level margins after incom-

[73]See Appendix G.2for additional discussion of alternative market definitions.

plete adoption is that the adopter's margins increase and the non-adopter's margins fall, cancelling out on average. We test this hypothesis by looking at non-adopter stations and comparing margins and prices before and after their rivals adopt (as before, we instrument for rivals' adoption with the rivals' brand adoption shares). Results are in Table E3 in Appendix E. There are no statistically significant changes in margins or prices following rivals' AP adoption, ruling out this explanation.

These results serve as a direct test of theoretical hypotheses about the effects of AP adoption on market outcomes (Calvano et al. 2020, Miklós-Thal and Tucker 2019).[74] We cannot be sure what type of algorithms stations are using and whether they fully delegate to them pricing decisions. Nonetheless, lack of margin changes from partial/asymmetric adoption and substantial increases in margins and prices after complete adoption indicate that algorithms facilitate tacit-collusion. The magnitude of margin increases in duopoly and triopoly markets is consistent with previous findings on coordination in retail gasoline markets (Clark and Houde 2013, 2014; Byrne and De Roos 2019).

6 Mechanism

In this section we use data from duopoly and triopoly markets to provide evidence of the mechanism through which algorithmic competition increases prices and margins. We first examine the time it takes for prices to converge to higher, possibly collusive, levels following adoption. Updating algorithms operating in fluctuating markets should experience a relatively long adjustment period, as they "learn" and explore the state space, such that convergence to stable strategies can take as long as several years. Asker et al. (2021) show that their less-sophisticated *asynchronous* algorithm converges to something close to the monopoly price, but takes considerable time to do so. Calvano et al. (2020) show that convergence to tacitly-collusive "punishment" strategies takes time. Alternative explanations of supra-competitive pricing by algorithms do not imply similar temporal patterns.[75]

[74] There is also a possibility that multiple stations in a market turn over their pricing decisions to a common algorithmic software provider. Algorithms in this case serve as the "hubs" of a hub-and-spoke cartel (Garrod et al. 2021). If multiple stations in a market turn over their pricing decisions to a common algorithmic software provider, our results are in line with the findings of Decarolis and Rovigatti (2021).

[75] There are at least two alternative explanations for why algorithms could reach margins above competitive levels. First, they could *fail to learn to compete effectively* (Cooper et al. 2015, Hansen et al. 2021). Algorithms may not fully incorporate rivals' prices or best-respond to them. In this case though, if margins were high, they would remain so initially and then might decrease over time as the algorithms learned to compete. Second, according to Brown and MacKay (2023), adoption of AP software changes the game firms play from a simultaneous Bertrand pricing game to a stage game, thereby increasing prices. We test a key prediction from their model: the bigger the asymmetry in pricing technology, the higher market prices and margins should be. We observe a large number of duopoly markets that feature asymmetric adoption of algorithmic pricing technology. Table E3 shows results from a regression of a non-adopting stations' margins on a dummy variable for whether its rival has adopted AP technology (instrumented by

We provide evidence in favour of this slow convergence to higher margins by examining the timing of adoption effects. Columns (1) and (2) in Table 10 show estimates of time-specific effects of incomplete and complete adoption on mean market margins and prices, in a regression that includes the controls from Table 9 and market and time FE. The time-specific adoption variables are instrumented by time-specific versions of IV_{mt}^1 and IV_{mt}^2 from Equation (5). We bin the timing effects into three periods: the first six months after adoption, the second six months after adoption, and a year or longer after adoption. We use these bins since there is only a small number of markets we observe for a very long period of time after adoption.[76]

Consistent with simulation results in Calvano et al. (2020) and Asker et al. (2021), we find that for roughly the first year after market-wide AP adoption there are no statistically significant changes in mean market margins at the 95% confidence level.[77] The magnitude of estimated coefficients for this time period is also quite small relative to our estimates in Table 9, which arise only a year after both stations adopt. For prices, we find similar results. Market-wide prices do increase in the first year after market-wide adoption, but once again the mean effects we estimate in Table 9 appear more than a year after market-wide adoption. We find no similar changes in prices or in margins following incomplete adoption. These results are similar to previous findings on transitions to collusive strategies in other markets. For instance, Byrne and de Roos (2019) document a three-year transition towards coordinated prices in the Australian retail gasoline market.[78] It is possible that the lagged effects might arise because algorithms take time to learn how to predict and respond to demand and cost shocks. However, unlike the competitive response, these effects are already visible in the first six months after adoption. The monopoly effects at 9am that we document in column 1 of Table 8 arise immediately, although they do become a bit stronger over time.

Next, we provide additional suggestive evidence of how algorithmic competition operates differ-

the rival brand's adoption share). We find no statistically significant changes in margins following a rival's adoption. Although the Brown and MacKay (2023) model appears to fit well certain settings (such as cold medicine markets), in our context it does not seem to apply.

[76]More generally, we have a relatively small number of markets with either partial or complete adoption, which restricts the heterogeneity in effects we can look for in the data.

[77]Figure 10 in Calvano et al. (2020) shows that profit margins for algorithms do not substantially change for over 500,000 simulation "periods." Under the assumption that a simulation period lasts for a few minutes, Calvano et al. (2020) suggest this would correspond to at least a year. This transition speed is also similar to previous evaluations of algorithmic learning in other settings. For hiring algorithms, Li et al. (2021) find that various algorithms require approximately a year to converge to new stable strategies after perturbations in the underlying data.

[78]Igami and Sugaya (2022) show that 1990s Vitamin cartels took several years to increase their prices and margins, while Clark et al. (2023) find a lengthy adjustment period to high prices for a Canadian bread cartel. However, these both involve explicit collusion.

Table 10: IV Duopoly and Triopoly Additional Price Effects

	(1)	(2)	(3)	(4)
Outcome:	Mean Market Wholesale Margin	Mean Market Price	Prob. Response to Price Decrease	Prob. Response to Price Increase
1-6 months since at Least One Station Adopted	-0.001	-0.000	0.015	-0.009
	(0.001)	(0.002)	(0.021)	(0.019)
7-12 months since at Least One Station Adopted	-0.002	0.000	0.012	-0.008
	(0.001)	(0.002)	(0.034)	(0.024)
12+ months since at Least One Station Adopted	-0.001	0.002	0.021	-0.001
	(0.002)	(0.003)	(0.078)	(0.052)
1-6 months since All Stations Adopted	0.010*	0.015***	0.103***	0.008
	(0.005)	(0.005)	(0.022)	(0.044)
7-12 months since All Stations Adopted	0.013*	0.022***	0.102***	-0.004
	(0.007)	(0.006)	(0.022)	(0.048)
12+ months since All Stations Adopted	0.045**	0.080***	0.350***	-0.046
	(0.021)	(0.018)	(0.072)	(0.166)
Zero Adopter Mean Outcome	0.0857	1.355	0.109	0.136
Market FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES
Observations	49,431	49,431	17,337	16,644

Notes: The sample includes duopoly and triopoly market/month observations from January 2016 until December 2018. "X months since at Least One Station Adopted " is a dummy equal to 1 in month t if at least one but not all stations in the market has become an adopter in the previous X months and zero otherwise. "X months since All Stations Adopted" is a dummy equal to 1 in month t if *all* stations in the market become adopters in the previous X months and zero otherwise. Instruments for both "X months since at Least Station Adopted " and "X months since All Stations Adopted" include measures of the "share of brand adopters" of the stations interacted with timing dummies. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. We also control for the sizes of the brands of the stations at month t. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

ently from non-algorithmic competition.[79] We empirically evaluate changes in pricing behaviour and the timing of these changes coming directly from duopoly algorithms competing against one another. We focus on pricing patterns that generally characterize AI-powered algorithmic pricing behaviour. We know that the algorithms are better than human- or rule-based algorithms at conditioning their behaviour on the state of the market and on competitor actions. We test for whether the conditioning behaviour evolves differently in markets with and without full AP adoption, and whether there is heterogeneity in responsiveness to the direction of competitor price changes. Our two key variables are (i) the market-level probability that if one station reduces its price, another station also reduces

[79]There are no clear conduct measures that can be identified in a reduced-form without an underlying model. In our context, developing such a model is not straightforward, since it would require making assumptions about how stations were competing prior to AP introduction, and about how algorithms operate. Many algorithms, including Q-learning as in Calvano et al. (2020), are not designed to play mixed strategies, while others are able to (including humans). There are many possible asymmetric equilibria, and characterising them without further information is not feasible. We leave this for future research.

its price within 5 minutes, and (ii) the market-level probability that if one station increases its price, another station also increases its price within 5 minutes.[80]

Columns (3) and (4) show estimates for the two probabilities. We find that after market-wide adoption, there is an immediate increase in the probability of responding to a rival station's price decrease within 5 minutes. We also find that this propensity is increasing over time, again suggesting there is gradual learning of new strategies by the algorithms. The magnitude of increased responsiveness is substantial. At the zero-adopter baseline, a station has 11% probability of responding to its rival price decrease with a price decrease of its own within 5 minutes. 12+ months after market-wide adoption, the propensity of responding within 5 minutes to a price decrease grows to 50%. The same pattern does not occur in markets where not all stations are AP adopters. Coefficient estimates for markets with incomplete adoption are positive but small and noisy. Notably, this is also not the case for responsiveness to price *increases*. Column (4) shows no evidence of decreases in stations' propensity to respond to rival price increases after algorithmic adoption.

Together these results are striking and suggest a simple mechanism through which algorithmic competition maintains high prices and margins. Effectively, the algorithms meet any price decrease with an immediate price decrease of their own, teaching each other that undercutting is not be profitable since the undercutter will always be followed to the lower price by the other station.

7 Robustness

We perform a series of checks to confirm robustness of our results to alternative samples, market definitions, adoption classifications and instruments. Results and with further details are in Appendix G. In every case results on the impact of adoption on margins are robust to the proposed check.

1. Alternative Estimation Samples (Appendix G.1): We address possible contamination from Shell's 2015 price matching promotion by (i) dropping observations from markets with Shell stations, (ii) dropping all observations from 2016. We also address concerns about entry/exit of stations by using a balanced sample of stations and a balanced sample of stations and markets, dropping any market where the number of stations changes over time.
2. Alternative Market Definitions (Appendix G.2): We define a market as 5-digit ZIP code, a

[80] These regressions are related to our adoption marker based on response time to rivals' price changes, but we believe they are distinct. Here we explicitly allow stations to have both immediate changes in responsiveness (which would identify the initial structural break) and longer term evolution in responsiveness that identifies changes in competitive strategy. We also separate responsiveness to rival price increases and price decreases.

well-defined unit of population in space (rural ZIP codes are bigger geographically).

3. Alternative Adoption Definitions (Appendix G.3): We classify adopters based only on measures that do not rely on the presence of a nearby rival, since this could be important for our comparison of monopoly and non-monopoly markets. We test a classification that drops responsiveness to crude oil price shocks, and another that requires a station to experience a break in the number of price changes. We consider a definition altering the time between structural breaks (in at least two out of four measures within two weeks rather than four). We classify a station as an adopter if they break in Diesel pricing, or in both E5 and Diesel.
4. Alternative Instruments (Appendix G.4): We use broadband access in station i's region as an instrument for adoption. If a station has access to high-speed internet and reliable signals, it should be more likely to adopt AP. We measure whether the local area around the gas-station has widespread access to high speed internet in a given year. We also introduce a "placebo" instrument. Rather than the share of stations of station i's brand that adopted as an IV, we use the share of stations of *another* brand (i.e., the brand of some station k in the market of station i). We expect there to be no correlation between the propensity of station i to adopt and average adoption by other brands, since they do not directly affect station i's costs.
5. Alternative Fuel Types (Appendix G.5): We replicate the analysis, including the definition of adoption, for E10 and Diesel fuels.
6. Alternative Specifications (Appendix G.7): We pool together data for monopolist and non-monopolist stations when estimating the impact of adoption on prices and margins.

8 Policy Discussion and Conclusions

Our findings suggest regulators should be concerned about mass-adoption of AP software. Reports released by antitrust authorities and economic organizations agree that explicit algorithmic collusion would not require change to existing competition law, but would affect how authorities monitor and investigate collusive practices. Increased tacit collusion through algorithms could change the legal status of such forms of collusion. Currently, tacit collusion is difficult to prove and prosecute as it does not rely on explicit communication. The UK Digital Competition Expert Panel states that with "further evidence...of pricing algorithms tacitly co-ordinating of their own accord, a change in the legal approach may become necessary" (p.110, 2019).

While our evidence is particular to German retail gasoline markets, similar AP software is being adopted in gasoline markets around the world. Our results suggest that authorities should undertake a census of retail-gasoline pricing software to understand the structure of the AP software market and the extent of adoption. Such a census can help separate whether the main effect of AP software on competition comes from multiple stations in a market adopting *the same* or *different* algorithms. We do not directly observe which algorithm competitors adopt and the two possibilities have different implications for regulators and policy-makers.[81]

Our focus is on the retail gasoline market, but custom-made and "off-the-shelf" algorithmic pricing software is widely available to use for online and offline retailers. Adoption of such algorithms is growing: Brown and MacKay (2023) present evidence of algorithmic pricing by pharmaceutical drug retailers online. Our results suggest that competition authorities should investigate the relationship between algorithmic pricing software adoption and competition in these and other contexts.

Finally, as mentioned in the Introduction, our findings suggest that competition authorities may be focusing their time and resources on the wrong things. Rather than pursuing hard-core cartels on an individual basis, it might be more effective to concentrate on collusion-facilitating devices that do not even require a conspiracy, such as algorithmic pricing and communication via earnings calls (see Aryal et al. 2022). In a platform setting, Johnson et al. (2023) propose simple market design features that can disrupt algorithmic price-increasing strategies, and such features may have wider applicability in other markets.

Data Availability

Code, publicly available data, and information about obtaining the proprietary data required for replicating the tables and figures in this article can be found in Ershov, Daniel; Clark, Robert; Assad, Stephanie; Xu, Lei, 2023, "Replication Data for: Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market", https://doi.org/10.7910/DVN/X4MSWW, Harvard Dataverse.

References

1. Abreu, D., Milgrom, P., & Pearce, D. (1991). "Information and timing in repeated partnerships." *Econometrica* 59, 1713-1733.

[81] If multiple stations in a market assign pricing decisions to a common algorithmic software provider, our results are in line with the findings of Decarolis and Rovigatti (2021). Algorithms serve as the "hubs" of a hub-and-spoke cartel (Garrod et al. 2021, Clark et al. 2023).

2. Agrawal, A., J. Gans and A. Goldfarb (2019). *The Economics of Artificial Intelligence: An Agenda.* University of Chicago Press.

3. Albrecht, S. V., & Stone, P. (2018). "Autonomous agents modelling other agents: A comprehensive survey and open problems." *Artificial Intelligence* 258, 66-95.

4. Andrews, D. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61, 821-56.

5. Aparicio, D., Z. Metzman & R. Rigobon (2021). "The Pricing Strategies of Online Grocery Retailers." NBER Working Paper 28639.

6. Aryal, G., Ciliberto, F., and B. Leyden (2022). "Coordinated Capacity Reduction and Public Communication in the Airline Industry." *Review of Economic Studies* 89, 3055-3084.

7. Asker, J., C. Fershtman and A. Pakes (2021). "Artificial Intelligence and Pricing: The Impact of Algorithm Design." NBER Working Paper 28535.

8. Assad, S., R. Clark, D. Ershov and L. Xu (2020). "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market." CESifo Working Paper No. 8521, Available at SSRN: https://ssrn.com/abstract=3682021.

9. Assad, S., R. Clark, D. Ershov and L. Xu. 2023. "Replication Data for: Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market," Harvard Dataverse, https://doi.org/10.7910/DVN/X4MSWW.

10. A2i Systems (2020), https://www.a2isystems.com/a2i-systems/.

11. Autorité de la Concurrence and Bundeskartellamt (2019). Algorithms and Competition.

12. Bantle, M., M. Muijs, and R. Dewenter (2018). "A new price test in geographic market definition: an application to German retail gasoline markets." Working paper

13. Boehnke, J. (2017). "Pricing Strategies, Competition, and Consumer Welfare: Evidence from the German and Austrian Retail Gasoline Market." Working paper.

14. Borenstein, S. and A. Shepard (1996). "Dynamic Pricing in Retail Gasoline Markets." *The Rand Journal of Economics* 27, 429-451.

15. Boswijk, P., J.H. Bun., and M. Schinkel (2018). "Cartel Dating." *Journal of Applied Econometrics* 34, 26-42.

16. Brown, Z. and A. MacKay (2023). "Competition in Pricing Algorithms" *American Economic Journal: Microeconomics* 15, 109-156.

17. Bundeskartellamnt (2020). https://www.bundeskartellamt.de/EN/AboutUs/Bundeskartellamt/legislation/legislation_node.html.

18. Byrne, D. and N. de Roos (2019). "Learning to Coordinate: A Study in Retail Gasoline." *American Economic Review* 109, 591-619.

19. Cabral, L., N.S. Durr, D. Schober, and O. Woll (2021). "Learning Collusion: Theory and Evidence from the Shell Price Matching Guarantee." Working paper.

20. Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2020). "Artificial Intelligence, Algorithmic Pricing and Collusion." *American Economic Review* 110, 3267-3297.

21. Carranza, J-E., R. Clark, and J-F. Houde. (2014). "Price controls and market structure: Evidence from gasoline retail markets." *The Journal of Industrial Economics* 63, 152-198.

22. Chen, L., A. Mislove, and C. Wilson (2016). "An empirical analysis of algorithmic pricing on Amazon marketplace." *In Proceedings of the 25th international conference on world wide web.* International World Wide Web Conference Steering Committee.

23. Chevalier, J. and A. Kashyap (2018). "Best Prices: Price Discrimination and Consumer Substitution." *American Economic Journal: Economic Policy* 11, 126-159.

24. Clark, R. and J-F. Houde. (2013). "Collusion with asymmetric retailers: Evidence from a gasoline price-fixing case." *American Economic Journal: Microeconomics* 5, 97-123.

25. Clark, R. and J-F. Houde. (2014). "The Effect of Explicit Communication on pricing: Evidence from the Collapse of a Gasoline Cartel." *The Journal of Industrial Economics* 62, 191-228.

26. Clark, R., I. Horstmann and J-F. Houde (2023). "Hub-and-spoke collusion: Theory and evidence from the grocery industry." Working paper.

27. Competition Bureau (2018). "Big Data and Innovation: Implications for Competition Policy in Canada." Discussion paper.

28. Conlisk, J., E. Gerstner and J. Sobel (1984). "Cyclic Pricing by a Durable Goods Monopolist." *Quarterly Journal of Economics* 99, 489-505.

29. 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.

30. Corts, K. (1998). "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment." *RAND Journal of Economics* 29 306-323.

31. Decarolis, F. and G. Rovigatti. (2021). "From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising." *American Economic Review* 111, 3299-3327.

32. Derakhshan, A., F. Hammer, F., & H.H. Lund (2006). "Adapting playgrounds for children's play using ambient playware." In 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 5625-5630). IEEE.

33. Derakhshan, A., F. Hammer and Y. Demazeau (2016). "PriceCast Fuel: Agent Based Fuel Pricing." In International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 247-250). Springer, Cham.

34. Dewenter, R., Heimeshoff, U., and Lüth, H. (2017). "The impact of the market transparency unit for fuels on gasoline prices in Germany." *Applied Economics Letters*, 24, 302-305.

35. Dewenter, R. and U. Schwalbe (2016). "An Empirical Analysis of Price Guarantees: The Case of the German Petrol Market." Working paper.

36. Dubois, P. and L. Lasio (2018). "Identifying Industry Margins with Price Constraints: Structural Estimation on Pharmaceuticals." *American Economic Review*, 108, 3685-3724.

37. Dubé, J.P. and S. Misra (2023). "Personalized Pricing and Customer Welfare." *Journal of Political Economy* 131, 131-189.

38. Erutku, C. and V. Hildebrand (2010). "Conspiracy at the pump." *Journal of Law & Economics* 53, 223-237.

39. Ezrachi, A. and M. Stucke (2015). "Artificial Intelligence and Collusion: When Computers Inhibit Competition." University of Tennessee, Legal Studies Research Paper Series #267, 2015.

40. Ezrachi, A. and M. Stucke (2016). "How Pricing Bots Could Form Cartels and Make Things More Expensive." *Harvard Business Review*, October 2016.

41. Ezrachi, A. and M. Stucke (2017). "Two Artificial Neural Networks Meet in an Online Hub and Change the Future (of Competition, Market Dynamics and Society)." *SSRN Electronic Journal.*

42. Gelman, M., Y. Gorodnichenko, S. Kariv, D. Koustas, M. Shapiro, D. Silverman, and S. Tadelis (2016). "The response of consumer spending to changes in gasoline prices." NBER Working Paper 22969.

43. Garcia, D., J. Tolvanen and A.K. Wagner (2021). "Demand Estimation Using Managerial Responses to Automated Price Recommendations." *Management Science* 66, 7793-8514.

44. Garrod, L., J. Harrington, and M. Olczak (2021). *Hub-and-Spoke Cartels: Why they form, how they operate, and how to prosecute them.* MIT Press.

45. Guerra-Hernández, A., A. El Fallah-Seghrouchni, & H. Soldano (2004). "Learning in BDI multi-agent systems." In International Workshop on Computational Logic in Multi-Agent Systems (pp. 218-233). Springer, Berlin, Heidelberg.

46. Hammer, F., A. Derakhshan, Y. Demazeau, and H. H. Lund(2006). "A Multi-Agent Approach to Social Human Behaviour in Children's Play." In 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (pp. 403-406). IEEE.

47. Hansen, K., K. Misra, and M. Pai (2021). "Frontiers: Algorithmic Collusion: Supra-competitive Prices via Independent Algorithms." *Marketing Science* 40.

48. Harrington, J. (2008). "Detecting Cartels." In *Handbook of Antitrust Economics* (P. Buccirossi Ed.), MIT Press, Cambridge, MA.

49. Harrington, J. (2018). "Developing Competition Law for Collusion by Autonomous Artificial Agents." *Journal of Competition Law & Economics* 14, 331-363.

50. Harrington, J. (2022). "The Effect of Outsourcing Pricing Algorithms on Market Competition." *Management Science* 68, 6889–6906.

51. Holmes, T. (1989). "The Effects of Third-Degree Price Discrimination in Oligopoly." *American Economic Review* 79, 244-250.

52. Igami, M. and T. Sugaya (2022). "Measuring the incentive to collude: The vitamin cartels, 1990-1999." *Review of Economic Studies* 89, 1460-1494

53. Johnson, J., A. Rhodes and M.R. Wildenbeest (2023). "Platform Design when Sellers Use Pricing Algorithms." Forthcoming, *Econometrica*.

54. Kaymak, L. and U. Waltman (2006). "A Theoretical Analysis of Cooperative Behaviour in Multi-Agent Q-Learning." Erasmus Research Institute of Management (ERIM), ERS-2006-006-LIS.

55. Kaymak, L and U. Waltman (2008). "Q-learning agents in a Cournot oligopoly model." *Journal of Economic Dynamics and Control* 32, 3275-3293.

56. Kehoe, P., B. Larsen, and E. Pastorino (2020). "Dynamic Competition in the Era of Big data." Working paper.

57. Klein, T (2021). "Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing." *RAND Journal of Economics* 52, 538-558.

58. Kühn K-U. and S. Tadelis (2018). "The Economics of Algorithmic Pricing: Is collusion really inevitable?" Working paper.

59. Lamba, R. and S. Zhuk (2022). "Pricing with algorithms." Working paper.

60. Lemus, J. and F. Luco. (2019). "Price controls and market structure: Evidence from gasoline retail markets." *The Journal of Industrial Economics* 69, 305-337.

61. Li, D., L. Raymond and P. Bergman (2021). "Hiring as Exploration." Working paper.

62. Luco, F. (2019). "Who Benefits from Information Disclosure? The Case of Retail Gasoline." *American Economic Journal: Microeconomics* 11, 277-305.

63. MacKay, A., D. Svartback, & A. Ekholm (2022). "Dynamic Pricing and Demand Volatility: Evidence from Restaurant Food Delivery." Working paper.

64. Mehra, S.K. (2015). "Antitrust and the Robo-Seller: Competition in the Time of Algorithms." *Minnesota Law Review*, 100.

65. Miklós-Thal, J. and C. Tucker (2019). "Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers?" *Management Science* 65, 1552-1561.

66. Milgrom, P., & Roberts, J. (1990). "Rationalizability, learning, and equilibrium in games with strategic complementarities." *Econometrica* 58, 1255-1277.

67. Monpolkommission (2018). "Algorithms and Collusion." *Biennial Report of the Monopolies Commission.*

68. Monpolkommission (2020). https://www.monopolkommission.de/de/monopolkommission/ aufgaben.html.

69. Montag, F. and C. Winter (2019). "Price Transparency Against Market Power." Working paper.

70. Moriyama, K. (2007). "Utility Based Q-learning to Maintain Cooperation in Prisoner's Dilemma Games." In 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'07) (pp. 146-152). IEEE.

71. Moriyama, K. (2008). "Learning-rate adjusting q-learning for prisoner's dilemma games." In 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Vol. 2, pp. 322-325). IEEE.

72. Musolff, L. (2022). "Algorithmic Pricing Facilitates Tacit Collusion." Working paper.

73. O'Connor, J. and N. E. Wilson (2020). "Reduced Demand Uncertainty and the Sustainability of Collusion: How AI Could Affect Competition." Forthcoming, *Information Economics and Policy.*

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

75. Pesendorfer, M. (2002). "A Study of Pricing Behavior in Supermarkets." *The Journal of Business* 75 33-66.

76. Quandt, R. (1960). "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes." *Journal of the American Statistical Association* 55, 324-330.

77. Ryan, S. and C. Tucker (2012). "Heterogeneity and the Dynamics of Technology Adoption." *Quantitative Marketing and Economics* 10, 63-109.

78. Salcedo, B (2015). "Pricing Algorithms and Tacit Collusion." Working paper.

79. Shiller, B. (2020). "Approximating Purchase Propensities and Reservation Prices from Broad Consumer Tracking." *International Economic Review* 61 847-870.

80. Shiller, B. and J. Waldfogel (2011). "Music for a song: an empirical look at uniform pricing and its alternatives." *The Journal of Industrial Economics* 59, 630-660.

81. Shoham, Y., Powers, R., & Grenager, T. (2003). "Multi-agent reinforcement learning: a critical survey." Working paper.

82. Slade, M. (1987). "Interfirm Rivalry in a Repeated Game: An Empirical Test of Tacit Collusion." *Journal of Industrial Economics* 35, 499-516.

83. Slade, M. (1992). "Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies." *The Review of Economic Studies* 59, 257-276.

84. Sobel, J. (1984). "The Timing of Sales." *The Review of Economic Studies* 51, 353-368.

85. Thisse, J.-F. and X. Vives (1988). "On the strategic choice of spatial price policy." *The American Economic Review*, 78, 122-137.

86. Tucker, C. (2008). "Identifying formal and informal influence in technology adoption with network externalities." *Management Science* 54, 2024-2038.

87. United Kingdom Digital Competition Expert Panel (2019), Unlocking Digital Competition.

88. Varian, H. (2018). "Artificial intelligence, economics, and industrial organization." NBER Working Paper 24839.

89. Wang, Z. (2009). "(Mixed) Strategy in Oligopoly Pricing: Evidence from Gasoline Price Cycles Before and Under a Timing Regulation." *Journal of Political Economy* 117, 987-1030.

90. Wieting, M. and G. Sapi (2021). "Algorithms in the Marketplace: An Empirical Analysis of Automated Pricing in E-Commerce." Net Institute Working Paper 21-06.

A Structural Break Test Results

Number of Price Changes: For each station we construct a variable measuring the number of times it changes its price for each date in our sample period. For structural break testing, we aggregate this variable to the weekly level.[82] 12,919 stations experience a significant structural break in the number of price changes at the 5% confidence level. Out of the stations that experience significant breaks, almost 50% of the best-candidate breaks occur in the spring and summer of 2017. Panel (a) of Figure A1 shows the overall distribution of best-candidate breaks.

Rival Response Time: We define a rival for station i as the closest station j that is within a 1km radius of station i but that belongs to a different brand.[83] Rival response time for station i is calculated as the number of minutes between the time of a price change by rival j and the subsequent price change by station i. If station i changes its price more than once before station j makes a price change, rival response time is taken as the average of the time gaps between each of station j's price changes and station i's subsequent change. When testing for structural breaks in rival response time, we take into account the fact that changes in response time will be mechanically impacted by changes in number of price changes. To identify structural changes separately from this mechanical effect, we control for the number of price changes by both stations. 5,227 experience statistically significant structural breaks. Out of stations with significant breaks (at at least the 5% level), almost 29% have best-candidate breaks in the spring and summer of 2017. Panel (b) of Figure A1 shows the overall distribution of best-candidate breaks.

Responsiveness to Crude Oil Price Shocks: We observe an intra-day time series for crude oil prices. In each non-holiday weekday, we separate fluctuations in crude oil prices from the moving average. We define a crude oil price shock as large deviations from the moving average. More concretely they are defined as deviations from the moving average that are above the 90th percentile of all deviations in a given year-month. This helps us account for changing volatility of oil prices over time. We define a response to a crude oil price shock as a price change within 5 minutes of the shock.

The outcome variable in the QLR regressions is the average number of times a station responds to an oil price shock in a week. We control for the average number of price changes a station makes in a week as well as for the number of oil shocks that happen in a week. This helps to control for the

[82] Any stations that do not have a weekly observation for average number of price changes in every week of 2017 are dropped. See more details in the Data Appendix.

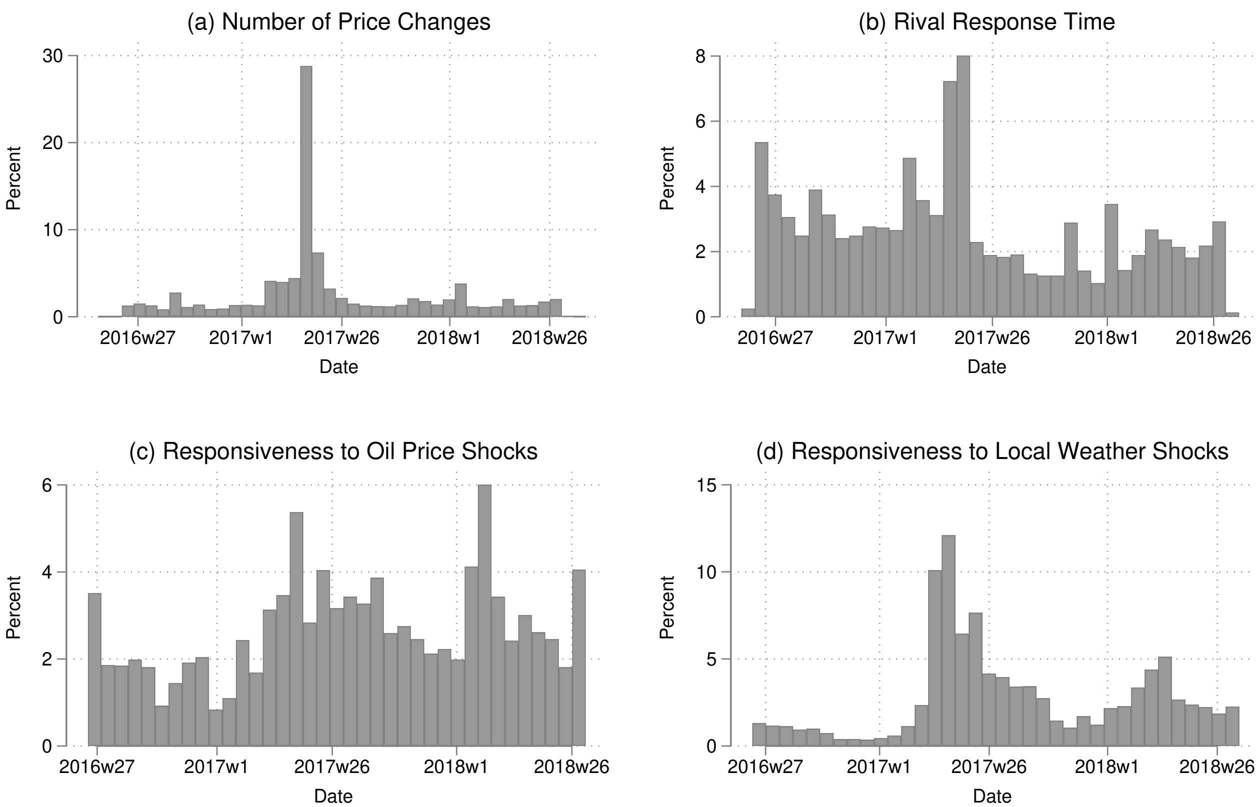
[83] This reflects the average distance of stations in the data.

fact that oil price volatility is changing throughout our sample. We find that there are 5,747 stations with statistically significant breaks (at the 5% confidence level). Panel (c) of Figure A1 shows the overall distribution of best-candidate breaks.

Responsiveness to Local Weather Shocks: Using data from the German Meteorological Service (DWD), we observe a high frequency time series of local air temperature around each gas-station. We separate fluctuations in temperature from the moving average in each non-holiday weekday. We define a local weather shock as large deviations from the moving average. They are defined as deviations from the moving average that are above the 90th percentile of deviations in a given year-month. We define a response to a local weather shock as a price change within 5 minutes of the shock.

The outcome variable in the QLR regressions is the average number of times a station responds to a local weather shock in a week. We control for the average number of price changes a station makes in a week as well as for the number of weather shocks that happen in a week, meaning that we are allowing for changes in responsiveness conditional on the weather volatility around the station. We find that there are 4,892 stations with statistically significant breaks. Panel (d) of Figure A1 shows the overall distribution of best-candidate breaks.

Figure A1: Frequency of Best-Candidate Structural Breaks



Notes: histograms show the distribution of best-candidate QLR structural break weeks for (a) the number of price changes (12,919 stations included), (b) the response time to a rival's price changes (5,227 stations included), (c) the number of responses to oil price shocks (conditional on the number of shocks and the number of station price changes)(5,747 stations included), and (d) the number of responses to local weather shocks (conditional on the number of shocks and the number of station price changes) (4,892 stations included).