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

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

January 17, 2022

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, Michigan, Montreal, 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. 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 - Toulouse School of Economics, Université Toulouse 1 Capitole, 1 Esplanade de l'Université, 31080 Toulouse, France, Email: daniel.ershov@tse-fr.eu., [d]Lei Xu - Queen's University, Dunning Hall, 94 University Avenue, Kingston, ON; 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 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. Algorithmic pricing 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 the fields of 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; Brown and MacKay 2021; Asker, Fershtman and Pakes 2021). 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 algorithmic pricing software. We focus on the German retail gasoline market, where, according to trade publications, algorithmic pricing 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 algorithmic pricing 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 be related to the use 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 station's price to local weather shocks. We focus on these measures since they capture the promised impacts of algorithmic software in the retail gasoline market. Leading algorithmic pricing software providers explain that their software perform high frequency analysis to "rapidly, continuously and intelligently" react to market conditions. We use a Quandt-Likelihood Ratio (QLR) test (Quandt 1960) to look for the best-candidate break date. For each of the four pricing behaviour measures, we test for structural breaks at each station for each week in a large window around the time of supposed adoption. 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 algorithmic-pricing adopter if it experiences a best-candidate structural break in at least two out of four pricing behaviours 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 data set experience best-candidate breaks in multiple pricing behaviour measures within a four week window. The majority of these breaks occur in mid-2017, just as algorithmic pricing software supposedly became widely available in Germany. We provide evidence that adopting stations experience noticeably different trends in all four measures, confirming that our data-driven approach for identifying adoption captures meaningful changes and is therefore valid.

After having identified adopters, we examine the impact of their adoption on retail prices and margins. For retail gasoline, margins are a clear indictor of profitability and market power: the ability of stations to mark-up retail prices over wholesale prices. Previous studies on coordination and collusion in this market use margins 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 2021). 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 show evidence of selection bias and diverging outcomes between non-adopters and adopters before their adoption date that attenuates 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 the gasoline retail market, brands can facilitate adoption by their stations. "Adopting" brands provide support/subsidies/training to individual stations, reducing adoption costs.[2] 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 AI, it is likely that the brand itself adopted and facilitated adoption by the stations. As a robustness check, we also use an alternative instrument: an annual measure of broadband internet availability in the local area around each station. Most algorithmic pricing software are "cloud" based and require constant downloading and uploading of information. Without high-speed internet, adoption will not be particularly useful. Conditional on local demographic characteristics broadband quality should not depend on station-specific unobservables, but stations should be more likely to adopt algorithmic pricing software once their local area has access to reliable high-speed internet.

Using brand-adoption as an IV we find that, following adoption, mean station-level prices and margins increase by approximately 1.3 cents per litre, or roughly 15% for margins.[3] The magnitude of these estimates is similar to claimed increases in gross profits achieved by stations employing algorithmic pricing software in Brazil and Denmark. These findings provide evidence of the causal impact of adoption of algorithmic pricing software on prices and margins. However, it is not clear whether these higher margins can be attributed to changes in the degree of competition intensity rather than to other factors, such as an improved ability to detect underlying fluctuations in wholesale prices or better predict demand.

To isolate the effects of adoption on competition we focus on the role of market structure, comparing adoption effects in monopoly (one station) markets and non-monopoly markets. If adoption does not influence competition, effects should be similar for monopolists and non-monopolists. We also perform a more direct test of theoretical predictions by focusing on small markets – two- and three-station markets. We compare market-level margins in markets where no stations adopted, mar-

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

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

kets where a subset of stations adopted and markets where all stations adopted. In the first market type, competition is between human- or rule-based algorithms. In the second it is between human- or rule-based algorithms and AI-powered algorithms, while in the last it is only between AI-powered algorithms. By comparing all three market types we are able to identify the effect of algorithmic pricing on competition.

We observe heterogeneity in outcomes based on market structure suggesting that AI-powered algorithmic pricing software affects margins and prices by changing competition. Adopting stations *with no* competitors in their local markets (i.e. monopolists) see no statistically significant change in their mean margins or prices. In contrast, adopting stations *with* competitors in their market see a statistically significant mean margin increase of 1.3 cents per litre.[4] Our market-level results suggest that relative to markets where no station adopts, markets where all do see a mean margin increase of 3.2 cents per litre, or roughly 38%. Mean prices increase by 6 cents per litre. Markets where only a subset of stations adopts see no change in mean margins or prices. These results show that market-wide algorithmic-pricing adoption raises margins, suggesting that algorithms soften competition. The magnitudes of margin increases are consistent with previous estimates of the effects of coordination in the retail gasoline market (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., how to tacitly collude) by testing the *timing* of price and margin changes. Updating algorithms operating in fluctuating markets should experience a relatively long adjustment period, as they learn and explore the state space and the set of possible outcomes. As a result, convergence to stable strategies can take as long as several years. Asker, Fershtman and Pakes (2021) show that their less-sophisticated *asynchronous* algorithm converges to something close to the monopoly price, but takes many periods to do so. Similarly, results in Calvano et al (2020) suggest that it can take time for algorithms to train and converge to stable strategies. Algorithms may learn to punish competitors for reducing prices or other tacitly-collusive strategies. We find evidence consistent with these results. Margins do not start to increase until about a year after market-wide adoption, suggesting that algorithms in this market learn tacitly-collusive strategies. We also examine the pricing behaviour that emerges in markets where both duopolists are algorithmic adopters. We show that in a market where both duopolists adopt, a station is more likely to respond to a rival's price decrease with an immediate price decrease of their own. There is no comparable change in the propensity to respond to price

[4] 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.

increases by a rival. The timing of these effects is consistent with the timing of the price and margin increases. Altogether, these findings provide further evidence that adoption affects competition and they suggest that the algorithms learn that undercutting will not be profitable, since the undercutter will always be followed to the lower price by its rival.

Our results have important policy implications. Antitrust authorities around the world 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, competition authorities expend substantial resources pursuing hard-core cartels on an individual basis. In so doing they may overlook what may be a much broader set of collusion-facilitating devices that do not even require a conspiracy. Algorithmic pricing may be one such mechanism. Communication via earnings calls is another (see Aryal et al 2021). 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 relevant players in the German market. Section 4 discusses the data and methodology we use to identify AP adopters. Section 5 shows our results on the impacts of AP adoption on station and market outcomes. We also conduct a number of robustness checks. In Section 6 we provide evidence to support the idea that outcome 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 algorithmic pricing and collusion. Theoretical and experimental results remain ambiguous. Several papers have shown that when algorithmic-pricing competition is modelled in a repeated game framework collusive outcomes are inevitable 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 studies find that these repeated games will converge to collusive outcomes including supra-competitive pricing and profits, as well as punishment of competitor deviation.[5] Asker, Fershtman and Pakes (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 (2021) develop a model where firms compete in pricing algorithms (rather than prices) and show that prices may increase even without collusion. 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.[6]

The question as to whether algorithm usage may result in coordinated behaviour has been studied in fields outside of economics such as law and computer science. There are several papers in the computer science literature studying coordination of algorithms in repeated games. 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 express more certainty that the use of algorithmic pricing can lead to collusive behaviour. 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 the retail gasoline market. There is a literature on collusion in gasoline markets. Earlier work includes Borenstein and Shepard (1996), as well as Slade (1987, 1992). More recently Wang (2008, 2009), Erutku and Hildebrand (2010), Clark and Houde (2013, 2014), and Byrne and de Roos (2019) have all studied anti-competitive behaviour in the retail gasoline industry. There have been a small number of papers looking specifically at the German retail gasoline market (Dewenter and Schwalbe 2016, Boehnke 2017, Cabral et al 2018, Montag and Winter 2019).

[5] Johnson, Rhodes and Wildenbeest (2021) propose simple market-design policies that have the potential to disrupt algorithmic collusive strategies in platform settings.

[6] 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 in this market is done through algorithms, which leads to parallels with the algorithmic pricing literature and 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 (Harrington 2018b). Unlike this paper, the primary focus of Decarolis and Rovigatti (2021) is on increasing intermediary concentration rather than on algorithmic pricing software behaviour and the mechanism through which bidding decisions are made.

A related area of literature studies the impact of technological advancements on price discrimination. A consequence of the rapid expansion of Big data and AI driven market analysis by firms is that personalized pricing strategies may become increasingly feasible and sophisticated. As technology advances, it can be better used to learn more about consumer tastes as well as to more accurately price products as a function of these tastes. In particular, authors have noted that Big data may facilitate first-degree price discrimination, which has generally been seen as challenging to implement in many markets (Ezrachi and Stucke 2016). It is possible that more accurate determination of optimal personalized pricing can increase firm revenues (Shiller and Waldfogel 2011; Shiller 2014). Kehoe, Larsen, and Pastorino (2018) find that firm profit, as well as consumer surplus, may increase or decrease under personalized pricing depending on consumers certainty regarding their product tastes. They also find that in every case, total welfare is higher under discriminatory pricing in comparison to uniform pricing. Dubé and Misra (2021) show through experiments that personalized pricing improves firm profits and that a majority of consumers benefit.

3 Background

3.1 The German Retail Gasoline Market

Similar to other retail gasoline markets around the world, distinct retail brands play an important role in Germany. Most stations in the market are affiliated with brands.[7] ARAL and Shell are the dominant brands, together making up over 25 percent of stations in Germany.[8] There are a number of other large brands with over 350 stations each: Esso, Total, Avia, Jet, Star, BFT, Agip, Raiffeisen, and Hem. In terms of market shares, ARAL, Shell, Jet, BFT, Total and Esso together account for 84 percent of fuel sales in the German retail gas market.[9]

There are two notable features of competition in the German gasoline market that relate to our analysis: the presence of price transparency and a price-matching policy initiated by Shell in 2015. Price transparency was instituted in August 2013 in response to concerns about tacit collusion and high consumer prices by German regulatory authorities. As part of this initiative, stations that change

[7]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).

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

[9]2019 fuel sales market shares for each brand are 21 percent for ARAL, 20 percent for Shell, 16 percent for BFT, 10.5 percent for Jet, 9.5 percent for Total, and 7 percent for Esso (bft.de).

their price must report their new prices "in real time" to the German Market Transparency Unit for Fuels (www.bundeskartellamt.de). Price changes are shared with consumer-facing information service providers. They are integrated into websites and mobile applications as well as into car GPS systems like TomTom.[10] The stated use of these data 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). There is conflicting evidence on the effects of this policy on prices and margins in Germany (Dewenter, Heimeshoff and Luth 2017, Montag and Winter 2019).[11]

The second major competition-related event is Shell's 2015 price matching guarantee. Under this policy, each Shell station had to match the lowest price of the 10 stations nearest to them within a 30 minute period. This policy did not apply to all consumers but only to those with Shell loyalty cards. Dewenter and Schwalbe (2016) and Cabral et al (2018) study this price matching guarantee and find that it very quickly increased average retail gasoline prices. The latter paper attributes the margin increase to the fact that an additional price jump was incorporated into the daily price cycles that characterize Germany's retail gasoline market. Prior to the policy, prices at stations throughout the country started the day at some high level and then gradually decreased starting at around 8am before rising sharply again in the evening. Shortly after the policy was implemented, a price jump at noon, followed by reversion, emerged. Overall, stations in this market featured considerable price variability throughout the day.[12] As we discuss in the following subsection, rule-based algorithms existed prior to the arrival of AI-powered algorithms leading to fairly sophisticated pricing even before adoption.

Our paper takes this competitive environment as a given. Our data set begins in January 2016, so we study the *additional* effects of algorithmic pricing software in a market with price transparency, daily price variation, and with the Shell price matching policy. We perform several robustness checks to confirm that Shell stations (or stations directly competing with Shell stations) are not driving the main results. We find that excluding them from the analysis does not change our main findings (see Appendix F.1).

[10] 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.

[11] See Luco (2019) for an analysis of a similar transparency program in Chile.

[12] 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 secretive about their pricing technology. AI-powered algorithmic pricing software providers are mostly privately-owned companies that are similarly secretive about their customer base. The structure of the "upstream" algorithmic pricing software market is unknown and there is no way to gauge the market share of any given software provider. 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). A few 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 artificial intelligence based pricing 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).[13] 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 gas station owners in the Netherlands and Belgium in 2015 (servicestationmagazine.be) and adopted by a number of Shell stations in the Netherlands by 2017 (WSJ.com).

In Germany, the December 2017 issue of *Tankstop*, a trade publication for Germany's retail gasoline sector, notes that *a2i*'s software had been available to gas station operators within Germany since that summer (see Figure A1).[14] Kalibrate's website explicitly refers to German markets as benefitting from "agile" (i.e., algorithmic) pricing (Kalibrate.com). Kalibrate has had contracts with German brands Orlen and Tamoil/HEM (Kalibrate, businesswire.com).

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 sta-

[13] It is possible that providers sold algorithmic pricing software in Germany before 2016 (the start of our sample). We should not be observing 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.

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

tion gross profits by approximately 10% (towardsdatascience.com). a2i similarly estimated that its software could increase station profits by at least 5% (a2i.com).

3.2.2 How does algorithmic pricing software work?

Most software providers reveal few details about their algorithms. Promotional materials generally describe their pricing software as based on "machine learning" or "artificial intelligence," with references to "neural networks" and "deep learning" (Kalibrate.com, PDISoftware, a2i.com). They describe software that can 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).

All providers stress the ability of their 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).[15] 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.[16]

a2i's algorithm works in three repeating steps. The first is "observation," where the gas station agent collects data from the environment and forms its "beliefs." As mentioned previously, these data include own prices, sales, traffic and environmental factors. Competitor behaviour is not explicitly modelled but the competitor station prices are included as inputs in this step. In the second step, "learning," the gas station agent uses an Artificial Neural Network (ANN) to map inputs into outcomes.[17] The outcomes are not explicitly outlined in Derakhshan et al (2016), but they likely correspond to sales, revenues and/or profits.[18] These are the "desires"/payoffs in the BDI model. The last step is "adaptation," where the gas station agent sets prices to achieve their "desires"/maximize

[15] 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.

[16] 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).

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

[18] 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).

the objective function.

Another algorithm by Aprix, a software provider operating in Brazil, is similar. 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).

A simple interpretation of the effects of adopting AI-powered algorithmic pricing software is that it makes stations substantially more sensitive to the state of the market. Figure A2, reprinted from *The Wall Street Journal*, presents a general summary of the functioning of gasoline pricing algorithms. Both this figure and the more detailed descriptions above describe pricing agents who are constantly learning about the state of the market. In retail gasoline this means continuously collecting consumer demand related information such as weather patterns and traffic patterns that can change driving behaviour and the probability consumers stop for gas, cost related information such as crude oil price fluctuations, or other relevant information such as competitor prices. This information can now be collected more or less continuously by scraping various websites (e.g., weather websites or Google maps for traffic). The algorithm then relates all this information to outcomes and decides on the best price to set conditional on the state. When the state changes, the algorithm sets new prices. This is not necessarily fundamentally different than what human pricing agents do. Human gas station operators also collect information about what competitors and consumers do and set their prices in response (Time.com). The differences are that the algorithm collects and processes more information than any human could. The algorithm can also respond faster to changes in the state of the market or respond to more subtle changes in the state. This is consistent with evidence from hotel markets that shows human pricing agents exhibit substantially more inertia and higher price adjustment costs as compared to algorithms (Garcia, Tolvanen and Wagner 2021).

Many questions remain about how this algorithm or other algorithms of this type operate in practice. Derakhshan et al (2016) does not explicitly state whether the "desires" and "intentions" (or the objective function and strategies) in the model are static or dynamic and whether the algorithm only sets current or both current and future prices. It is also not clear how many past prices it considers in each optimization round (the memory), or how it learns. The learning method is important since Milgrom and Roberts (1990) show that agents characterized by "passive/adaptive learning" (based on past rival responses) and who optimize their static best response cannot reach collusive equilibria. This is not the case for "reinforcement learning" algorithms, such as Q-learning, that can experiment with temporarily sub-optimal strategies to maximize the overall net present value of future payoffs. Reinforcement learning has been the focus of existing simulations-based evidence of the possibility of algorithmic collusion (Calvano et al 2020, Klein 2021).

Derakhshan et al (2016) implies that station agents have dynamic objective functions and set both current and future prices. In an illustrative example they mention that their algorithm can "predict the volume through the day (24 h) at the start of the day." Elsewhere, objectives for algorithmic pricing software are described dynamically (i.e., "maintain market shares"). Derakhshan et al (2016) does not mention using a Q-learning algorithm or any algorithmic "exploration"/"experimentation." The broad description of the algorithm appears to be closer to the "passive learning" approach. However, it also cites reinforcement learning literature (e.g., Shoham et al 2003). More generally, the BDI model provides an attractive setting for reinforcement learning (Guerra-Hernandez et al 2005), and the combination of BDI and reinforcement learning has been an active field of research in computer science in the last 20 years (Albrecht and Stone 2018). We should also mention that there is no detailed information about algorithms available from other providers. They may very well be based on Q-learning or other reinforcement-learning mechanisms.

Even with passive/adaptive learning, as long as algorithms set a sequence of prices rather than simply optimize the static best response, the introduction of algorithms in many gasoline retail markets could lead to increased cooperation between stations. This is because of price disclosure initiatives that have been introduced in many countries. In Germany, France, Spain, Chile, Argentina, and other countries, gas stations must report price changes within minutes of changing their prices at the pump.[19] Price information is then immediately and publicly displayed on price comparison websites. This policy creates a market with *perfect monitoring*. Pricing algorithms can process information and react faster than humans to changes in rival behaviour. Derakhshan et al (2016) presents an illustrative example where the algorithm of one station detects changes in the pricing of another station and responds rapidly. Algorithms, therefore, increase the speed of interaction. In a setting with perfect monitoring, increases in the speed of interaction facilitate cooperation, since it is easier to detect and punish deviations from tacitly-collusive equilibria (Abreu et al 1991).

[19] In Germany, stations must do it within 1 minute. In France and Chile, stations must report within 10 and 15 minutes of changing prices at the pump, respectively.

3.3 Algorithmic Pricing Software Adoption

As in other cases of corporate technology adoption (e.g., Tucker 2008), 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 the technology they would like their stations to use. They provide stations with employee training, technical support and maintenance and subsidies. Individual station owners make adoption decisions specific to their stations. This involves incurring investment costs such as pump and Point of Sale (PoS) terminal upgrades. The costs can be substantial and are not necessarily fully subsidized by the brand.

An example is the adoption of electronic payment systems in the 1990s. Analogous to algorithmic pricing software, this is a technology that clearly benefits brands and that brands would want their stations to adopt, but that some stations may not want to adopt because of the costs involved. BusinessWeek reports that as part of a brand-wide roll-out of a contactless electronic payment system in 1997 by Exxon Mobil (Esso's US parent company), individual station owners "have to install new pumps costing up to $17,000–minus a $1,000 rebate from Mobil for each pump" (BusinessWeek). Partial investment subsidies by brands help explain staggered or delayed technology adoption in this market. We provide additional evidence for staggered technology adoption in the gasoline retail market in Appendix E. 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 of availability, fewer than 50% of stations owned by leading brands adopted the technology (Figure E1).

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. For example, in Denmark *a2i* directly entered into a partnership with the large 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 December 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 Point of Sale (PoS) terminals and self-checkout solutions (DieboldNixdorf.com).

However, if a brand decides to "adopt," or enter into a partnership with an AI pricing software provider, its stations do not necessarily automatically and instantaneously adopt. There are many reasons why not every one of a brand's stations would adopt this technology. Cloud-based AI-pricing software potentially requires substantial infrastructure investments and not all station owners are in a position to incur these costs right when the technology becomes available, or possibly ever. For example, high-speed internet *and* high-speed internet enabled PoS terminals and pumps are likely required for the software to work. In Germany, many areas do not have access to stable high speed internet connections.[20] Equipment upgrades of this sort are expensive, costing thousands to tens of thousands of Euros (mobiletransaction.org).[21] 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 we use 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. It includes all price changes for the most commonly used fuel types (E5, E10, diesel) for over 16,000 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.[22] Our sample covers January 2016 to December 2018.[23] We focus on E5 fuel, which has over 80% market share in Germany (bdbe.de).[24]

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

[20] Reports suggest that many areas and regions in Germany receive sub-par services and speeds that are compared to the "old dial-up days" (NPR.org). We use broadband internet availability as an alternative instrument. See discussion in Appendix F.4.

[21] 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).

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

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

[24] Super E10 is an ethanol based fuel with 10% ethanol and 90% unleaded petrol. Super E5 is an ethanol based fuel with 5% ethanol and 95% unleaded petrol. In Assad et al (2020), we find similar results using E10 fuel.

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.[25]

4.1 Station-Level Descriptive Statistics

Table 1 shows summary statistics, including the number of stations per brand, the number of stations per ZIP code and the average distance between stations. Out of the 16,027 stations in our data set, single-operating stations account for approximately 11 percent. With our IV strategy, these stations are not part of our final estimating sample. The remaining stations are affiliated with brands. 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). There are 258 distinct brands in the data, of which 239 have between 2 and 100 stations and 19 have more than 100. The top 5 brands account for 43 percent of stations and the 19 largest brands (those with more than 100 stations) account for 71 percent of total stations (11,752 stations total).

Table 1: Brand and Market Summary Statistics

Variable	**Observations**	**Mean**	**Std. Dev.**	**Min**	**25%**	**75%**	**Max**
Stations per Brand	258	57.6	227.0	2	2	19	2417
Stations per Market	4,138	3.87	2.35	1	2	5	21
Stations per 5 Digit ZIP Code	5,781	2.77	2.15	1	1	4	17
Distance to nearest station (KM)	16,027	1.40	1.77	0	0.30	1.69	17.19

The market definition we use in the main text 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 B. Using this approach there are 4,138 markets of which 489 have a single station (are monopoly markets), 738 have two stations (are duopolies) and 877 have three stations (are triopolies). The full distribution is presented in the appendix. The mean number of stations per market is around 3.9. Only 107 markets have more

[25] 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.

than 10 stations.[26] As we explain in further detail below, in our robustness checks we also consider a market definition based on ZIP-codes, and our results are robust to this alternative.

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 the price changes in the raw data, we calculate an average weekday (non-weekend or holiday) price from 7am to 9pm for each station. To construct margins, we merge these with regional wholesale prices that are average daily ex-terminal prices in eight major German refinery and storage areas. We calculate the distance between each gas station and all refinery and storage areas and use wholesale prices from the nearest refinery.[27] Prior to subtracting wholesale price, we also subtract German

[26] 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 5,781 5-digit ZIP codes in our data of which 2,094 have a single station (are monopoly markets), and 1,307 have two stations (are duopolies).

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

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 E5 price at 9am, noon, 5pm and 7pm. Once again, we take a monthly mean for our station-month level analysis.

Table 2 shows summary statistics at the station/month level, including prices, margins and regional demographics and weather. The average price that a station charges is 1.36 Euros per E5 litre, but the mean monthly margin that the average station earns over wholesale regional price (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 local weather stations. We compute the average distance between each gas station and all local weather stations and use weather 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. That is, we exploit the fact that we observe exactly when stations adjust their prices, 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. Algorithms continuously observe the "state" by scraping information from the internet and other sources. Once the state changes, the algorithms re-optimize prices. Human- or rule-based price setting would operate in a similar manner, but would be worse at observing and conditioning on state variables than AI-powered algorithms. Therefore, 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. The 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 this out across each week.

2. The 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 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 out response times over each week.

3. The 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.[28] 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. The 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 gas-station.[29] We separate fluctuations in temperature from the moving average. We define a local weather shock as large deviations from the moving average.[30] We define a response to a local weather shock as a price change within 5 minutes of the shock. We calculate the number of local shocks, the number of responses and response probability for each station and week.

These measures are consistent with the promises of AP software providers. *a2i*'s website states that their software "rapidly, continuously, and intelligently react[s]" to market conditions. Similar measures have also been used previously in the literature to identify heterogeneity in pricing technology. Brown and MacKay (2021) use the number of price changes by retailers in a given period and the speed of reaction to identify new pricing technology. Aparicio, Metzman and Rigobon (2021) similarly 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

[28]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.

[29]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.

[30]As for oil shocks, they are defined as deviations from the moving average that are above the 90th percentile of all deviations in a given year-month.

correlation of user pricing with certain target prices, such as the lowest price of that given product in the Marketplace.

Formally, we look for structural breaks in the above 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 suggested and used in previous work involving collusive behaviour (Harrington 2008; Clark and Houde 2014; Boswijk et al 2018; Crede 2019; Byrne and de Roos 2019). We conduct a QLR test for each station in our data set and for each measure and station. Further details on these tests can be found in Appendix C.

An example of stations with and without structural breaks is in Figure 1, which shows the average number of price changes per day for two stations: Willer Station 131 in Kiel and Tankstelle Lehmann in Aalen. Both stations change their prices approximately five times per day in 2016 and up to the middle of 2017. Tankstelle Lehmann continues to change their prices at the same rate throughout the rest of the sample, but Willer Station 131 starts changing their prices over 10 times per day starting in the middle of 2017. This variation in the number of price changes per day is abrupt. A QLR test for this measure picks up no structural breaks in the number of price changes per day for Tankstelle Lehmann, and a structural break in the middle of 2017 for Willer Station 131.

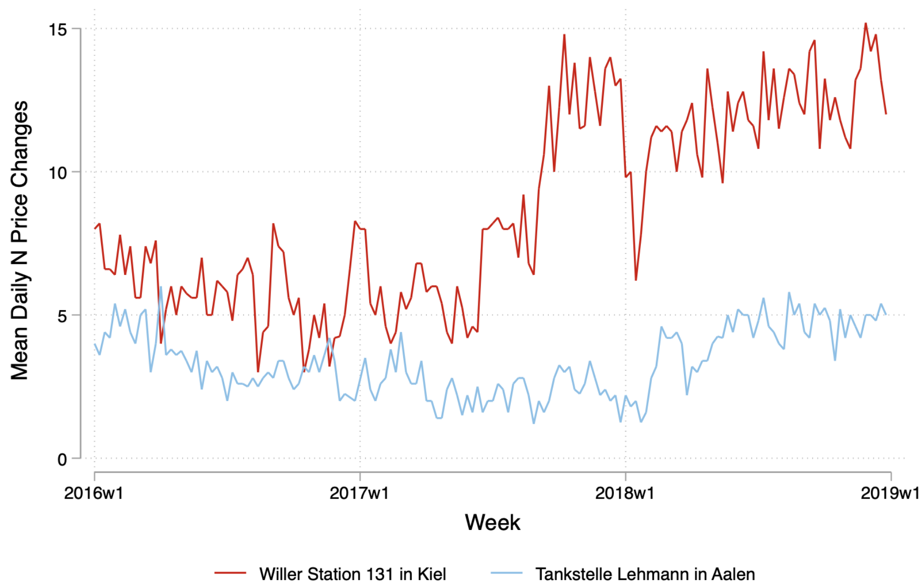
Appendix C.2 shows the distribution of structural breaks for each measure.[31] We find a large number of statistically significant breaks in the data. Most importantly, many breaks occur in the middle of 2017, when we believe AP technology became available to stations. Nearly 50% of best-candidate breaks in the number of price changes are in the spring of 2017. Similarly, 40% of best-candidate breaks in the responsiveness to local weather shocks, 20% of best-candidate breaks in rival response time and nearly 20% of best-candidate breaks in responsiveness to oil price shocks happen around that time as well.[32]

We also find that the structural breaks capture quantitatively important changes in pricing behaviour. Figure 2 compares outcomes between stations with breaks and stations without breaks throughout our sample period. We find large differences in all four outcomes. Stations with structural breaks in the number of price changes per day have 20% more price changes on average than

[31] One concern is that other structural breaks may occur at significantly different dates if we considered F-statistics that are not the maximum, but close to it. We find that generally F-statistic distributions are unimodal and stations do not have significantly different dates that may be identified as a structural break. Examples of F-statistics distributions are in Figure C5.

[32] 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.

Figure 1: Evolution of Number of Price Changes per Day in Two Stations



Notes: Each time series shows the weekly average of the number of daily price changes in each non-holiday weekday for Willer Station 131 in Kiel and Tankstelle Lehmann in Aalen.

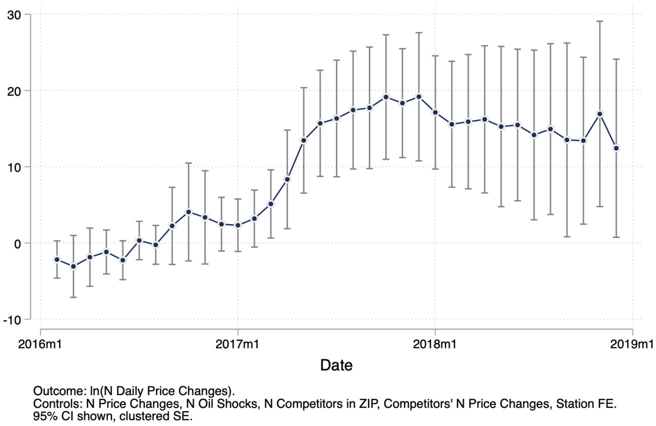
stations without structural breaks in the number of price changes. Stations with structural breaks in response time to rivals, respond 10% faster to rivals' price changes than stations without structural breaks in that measure. Stations with structural breaks in responsiveness to weather shocks respond 20% more frequently than stations without best-candidate breaks in that measure. Stations with breaks in responsiveness to oil price shocks respond 5% more frequently than those without breaks. Importantly, the change between stations with breaks and without breaks appears very rapidly around the middle of 2017. Overall, these graphs suggest that our structural break measures are picking up rapid and substantial changes in pricing technology.

We find that on average, stations without structural breaks in the number of price changes change their prices approximately 5 times per day during the sample period. Approximately, this means they change their prices once every 3 hours, assuming average opening hours from 7am to 9pm.[33] Stations with structural breaks change their prices approximately 8 times per day during the sample period (9.3 times per day in 2018). This means they change their prices approximately once every hour and a half. This frequency of price changes is similar to the most rapid (hourly) average price change frequencies identified in online markets by Brown and MacKay (2021) and Aparicio, Metzman

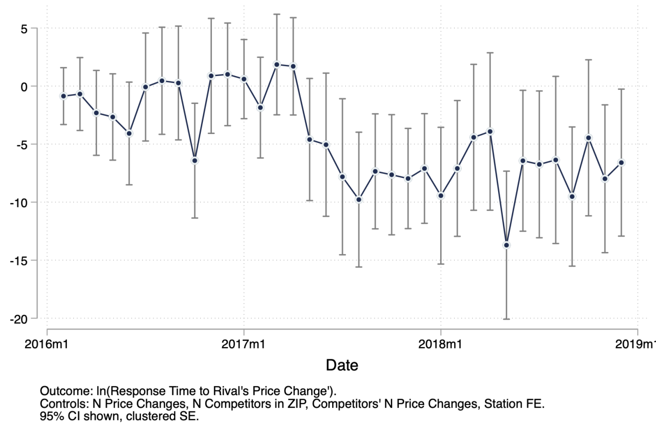
[33]The number of price changes increases slightly throughout the sample period, from 4.8 to 5.3.

Figure 2: % Difference Between Stations with and without Structural Breaks

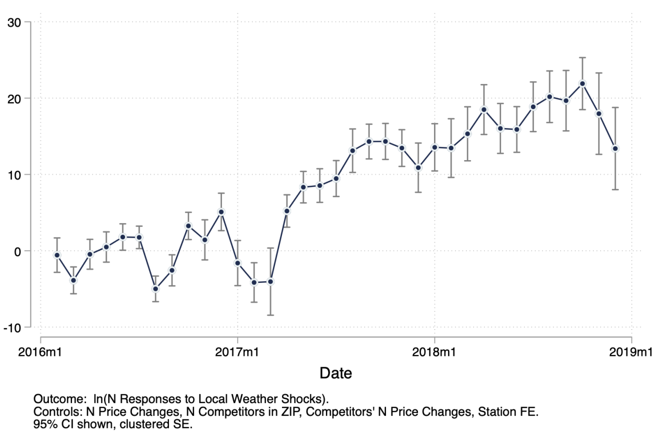
(a) Outcome: ln(N Price Changes per Day)



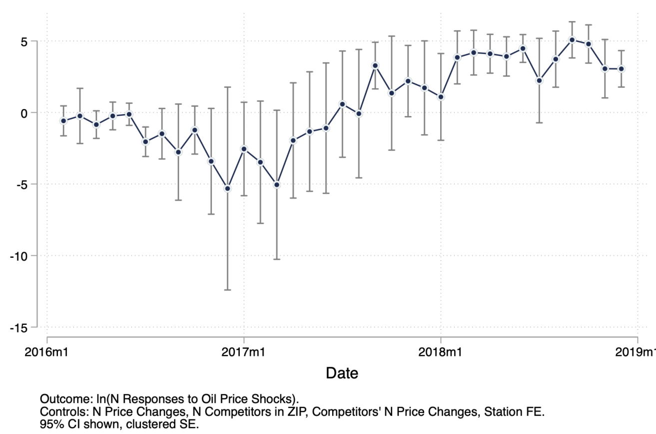
(b) Outcome: ln(Response Time)



(c) Outcome: ln(N Responses to Weather Shocks)



(d) Outcome: ln(N Responses to Crude Oil Shocks)



Notes: Each panel shows γ_t^2 estimates and their 95% confidence intervals from a station-month level regression of: $y_{it} = \sum_t \gamma_t^1 D_t + \sum_t \gamma_t^2 D_t \times \text{Break}_i + \delta_i + \epsilon_{it}$, 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 outcome y between a station with a best-candidate break and a station without a best-candidate break in month t.

and Rigobon (2021).

For response time to rivals, we find that on average stations with structural breaks respond to rivals within 50 minutes after a break. This is at least as fast as the responsiveness identified in Brown and MacKay (2021) and Aparicio, Metzman and Rigobon (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). Average response time for stations without structural breaks is over 84 minutes.[34] Online pricing in a multiproduct markets may be more complex than offline pricing in the more relatively homogenous retail gasoline market, but online retailers should have data that is at least as good as German gas station data. Online retailers can easily and continuously scrape competitors' price data and demand proxies such as sales ranks. Online pricing technology should be at least as advanced as offline retail pricing 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.[35]

For responsiveness to weather shocks, according to our definition of a weather shock (see above and in the Data Appendix), we identify approximately 3.7 weather shocks per week. On average, we identify 0.9 responses for stations with structural breaks after their break date. By comparison, we identify 0.5 responses for stations without structural breaks, meaning that stations with superior pricing technology are twice as likely to respond to a weather shock. For crude oil price shocks, we identify an average of 12 oil price shocks per week. A station with a structural break responds to an oil shock 1.2 times per week after their break date, as compared to 1 time per week for a station without a structural break.[36]

Classification: There are many factors that 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.[37] Our results are robust to stricter alternative definitions of adoption.[38]

We classify 2,728 stations as adopters. Figure 3 shows the distribution of the average break date for all adopters, defined as is the average year-week between best-candidate break dates of the two or three measures in which a station experiences a significant break.[39] Over 50% of these average

[34]The difference in raw averages is larger than the difference in Fig. 2 since Fig. 2 accounts for station fixed effects that absorb some of the heterogeneity.

[35]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.

[36]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.

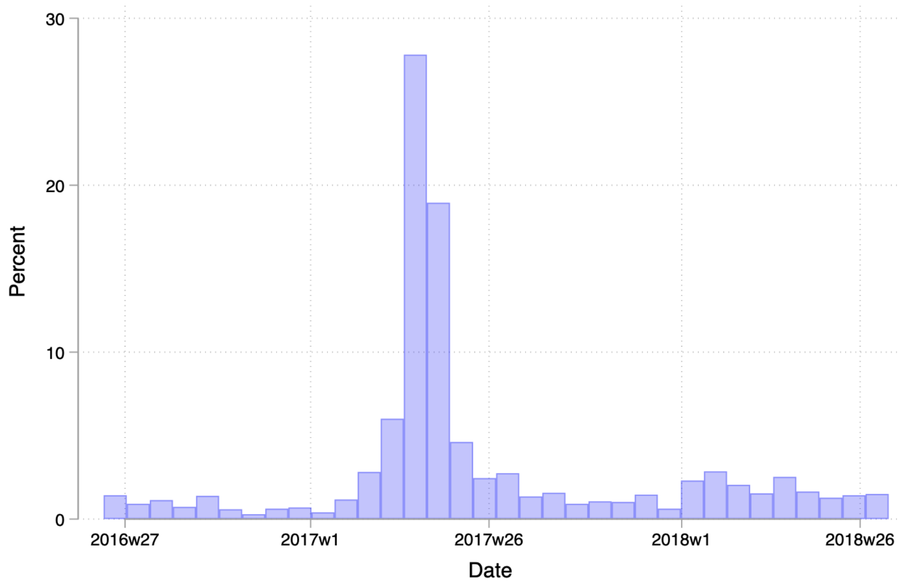
[37]Any combination of two measures will result in a station being classified as an adopter.

[38]In Appendix F.3 we require stations to experience best-candidate breaks in at least two of the four 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.

[39]This is a conservative approach. We may be "missing" some adopters, either due to measurement errors in our

break dates occur in the middle of 2017. This is consistent with the supposed increased availability of algorithmic pricing software in the middle of 2017 in Germany (see Section 3.3).

Figure 3: 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.

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 4 compares outcomes between adopter and non-adopter stations throughout our sample period.

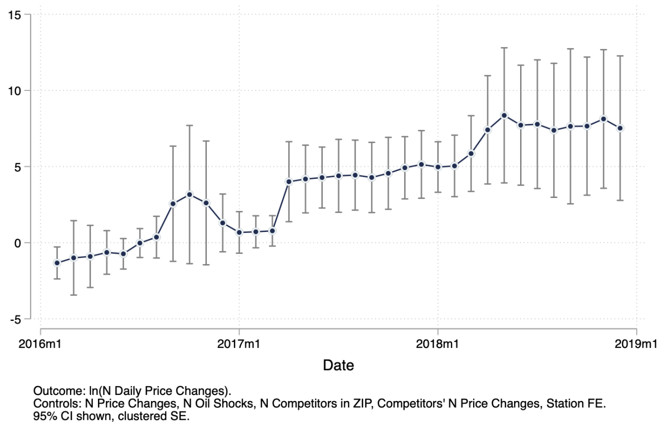
We also find ex-ante average differences in local demographic characteristics and local markets between stations labelled as adopters and stations not labelled as adopters.[40] In Table 3 we find statistically significant differences in market characteristics between adopter and non-adopter stations before any adoption takes place (in 2016). Adopter stations are located in denser areas with different demographic profiles. Adopter stations also face more competition. This could suggest that adoption decisions are likely endogeneous, with stations choosing to adopt in response to observable and unobservable market conditions.

measures or due to other signals of adoption that we did not consider. In practice, this means that some of the adopters are labelled as non-adopters. This would bias our station-level estimates towards zero and under-state the true effects of adoption.

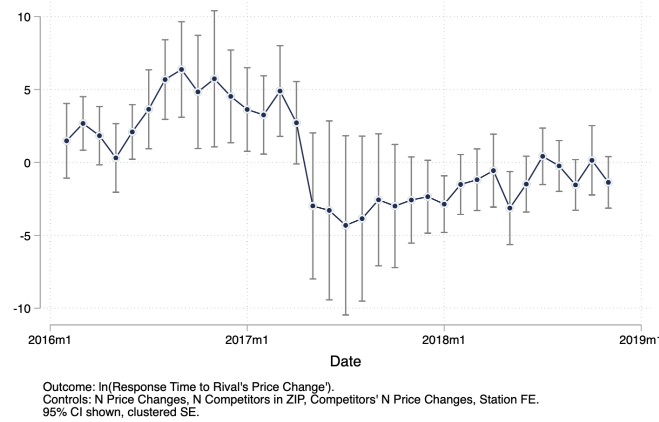
[40]We do not observe individual station characteristics.

Figure 4: % Difference Between Adopters and Non-Adopters

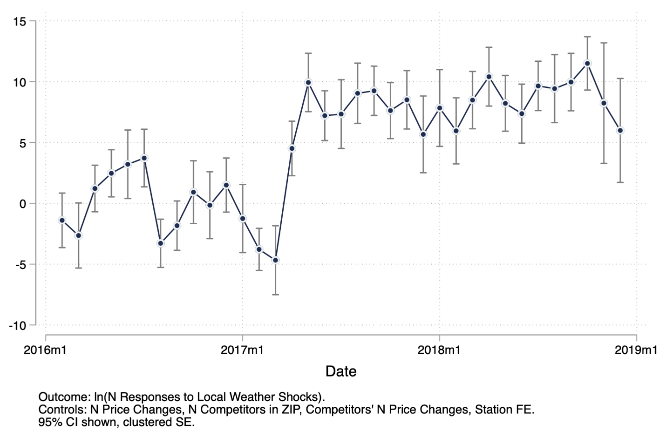
(a) Outcome: ln(N Price Changes per Day)



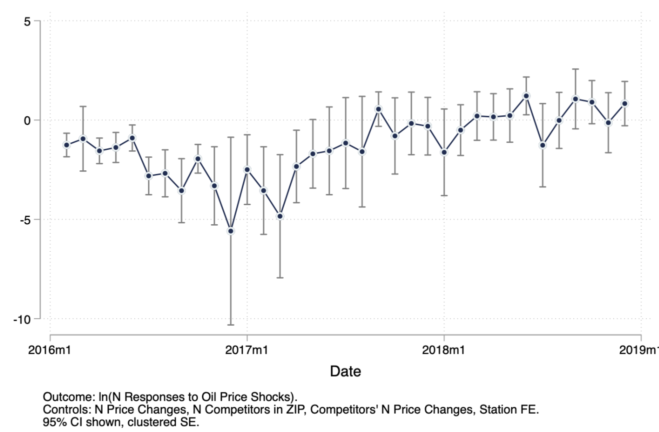
(b) Outcome: ln(Response Time)



(c) Outcome: ln(N Responses to Weather Shocks)



(d) Outcome: ln(N Responses to Crude Oil Shocks)



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.

A possible concern with our definition of adoption is that non-adopting stations may be mistakenly labelled as adopters because their response to an adopting rival's pricing makes them behave as though they also adopted. This does not appear be a regular occurrence. For example, 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 adopter station, and 31 had two adopters. More generally, Figure 5 shows the geographic distribution of adoption shares in markets with more than one adopting station in December 2018

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

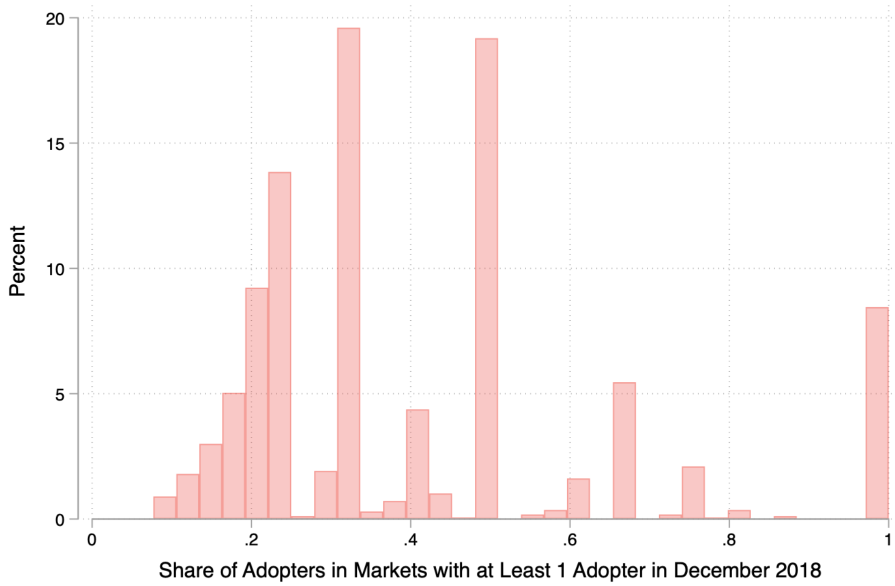
(the last month in our data). There are 1,700 such markets out of a total of more than 4,000. There are relatively few markets where adoption shares are higher than 50%.

4.3.2 Brand-Level Adoption

We do not observe an indicator for whether a brand HQ decided to enter into a 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 that have been classified as adopters by time t. This approach 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 time t is unlikely to be an adopter at time t, while a brand for which a large percentage of stations adopted is more likely to be an adopter. Alternative definitions could classify a brand as an adopter as soon as *any* one of its stations is classified as an adopter, or only after *all* of its stations are classified as adopters. These alternative approaches do not reflect technology adoption in this market.[41]

[41] Brand adoption is *not a necessary condition* for a station to adopt algorithmic pricing software. There are many providers of algorithmic pricing software that 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 their brand,

Figure 5: December 2018 Market-level Adoption Shares



Notes: This histogram shows the distribution of adopting station shares in markets with at least one adopting station in December 2018. Adoption shares are calculated as: $\frac{\text{N Adopting Stations in Market}}{\text{N Stations in Market}}$.

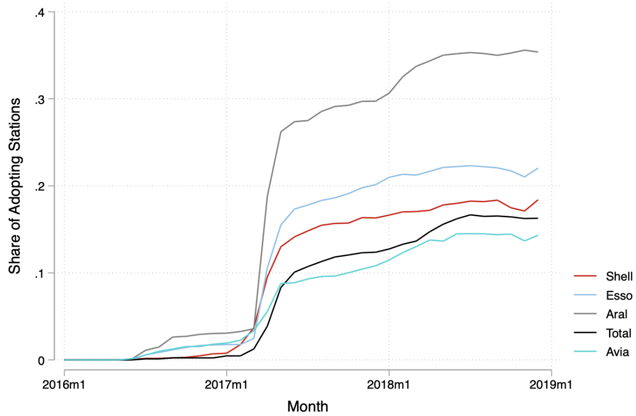
Figure 6 shows the evolution of the share of adopting stations in our data throughout our sample period. Panel (a) displays results for the top 5 brands (by station count). Adoption happens at a staggered rate that varies across brands. All brands experience an increase in adoption rate that happens 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 a steadier (albeit slower) pace compared to other brands. The heterogeneity in adoption rates across brands suggests that there is a brand-specific component to AP adoption, possibly reflecting that some brands were more likely to support the new technology (or adopt at the "HQ" level). None of these brands have adoption rates over 40% by the end of the sample period.

Turning to panels (b) and (c), which, respectively, show results for the top 20 brands and for all brands, we can see that the share of adopters for smaller brands is typically lower, and often occurs a bit later than for the large brands. While the mean adopter share for top 5 brands is

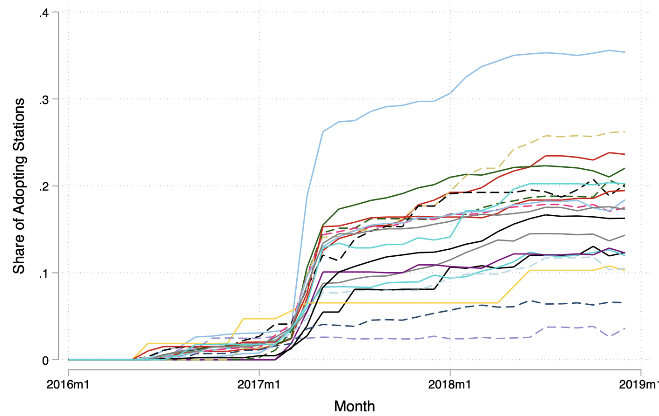
can adopt their technology. Defining a brand as an adopter if *any* one of its stations is classified as an adopter would be sensitive to outliers and amplify noise from our station-level adoption measure. The opposite approach, defining a brand as an adopter only if *all* of its stations is inconsistent with the history of technology adoption in gasoline retail markets. As explained in Section 3.3, brand subsidies to stations for technology adoption are often incomplete and technology adoption is highly staggered. In Appendix E we show that it took years for a substantial share of gasoline stations belonging to top brands to adopt electronic payments in the 1990s.

Figure 6: Share of Brand Stations that are AP Adopters

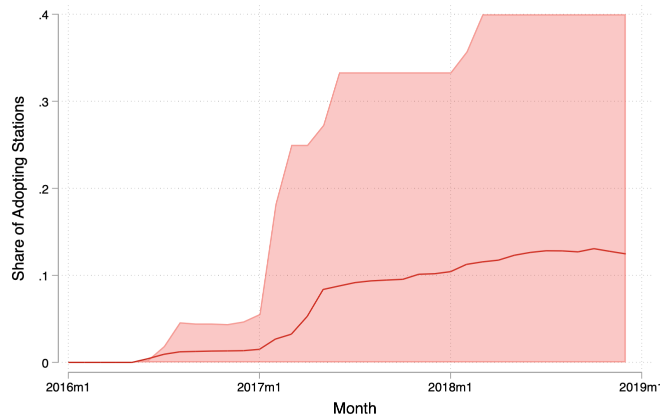
(a) Top 5 Brand Shares



(b) Top 20 Brand Shares



(c) All Brand Share Distribution



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 share of brand stations that are AP adopters in each month for the top twenty brands in our data (by count of stations). Panel (c) shows the distribution of brand adoption shares for all 258 brands in our data. The solid red line in Panel (c) shows the average brand adoption rate in a month. The shaded area in Panel (c) shows the distribution of brand adoption shares between the 5th and 95th percentiles (e.g., for each month, we calculate the 5th and 95th percentile brand adoption share and plot it).

21% by December 2018, for non top 5 it is 12%. This likely reflects the better support that larger brands can provide to their stations, which would reduce their cost of adoption. In Panel (c) we present the mean for all brands (the red line), but also the area captured by the 5th and 95th percentiles. The difference between the 5th and 95th percentile grows over time, suggesting that over time heterogeneity in within-brand adoption rates is growing.

The pattern in Figure 6 is similar to the staggered-adoption that was observed for electronic payment adoption by Canadian gasoline retail stations in the 1990s (see Figure E1 in Appendix E). Despite the differences in time, geography and technology, we also find a staggered pattern of technology adoption that appears to be highly brand specific. This suggests that our AP adoption classification captures technology adoption.

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:

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

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 both stations that need to 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.[42] 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 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 show the time-varying β coefficients. They show statistically significant differences in mean prices and mean margins between adopter and non-adopter stations in the time before adoption. Adopters had margins that were 0.2 cpl higher than margins 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 that is made by stations that need to adopt and that can afford to. These stations likely had better unobservable shocks (and higher margins

[42] The presence of such endogeneity would be mitigated with 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 D.1.

and prices) some time before adoption and may expect worse shocks in the future.

Table 4: OLS Station-Level Estimates

Outcome:	(1) Mean Margin	(2) Mean Price	(3) Mean Margin	(4) 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.[43] As explained in the previous section, we

[43] 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

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.[44] 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.[45]

Station-level IV estimates are presented in Table 5. Column (1) shows the first stage of the IV 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 (excepting station i) that adopt by period t increases the probability that i adopts by period t by 66%. This is consistent with our intuition that adoption of algorithmic pricing 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 cents per litre on average after AP adoption. This is an increase of about 15% relative to the average non-adopter margin

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 F6 for results and Appendix F.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 F.4.

[44] 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.

[45] Table C1 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 null effects 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 also in Appendix F.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 station 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

of 8.2 cents.[46] Prices also increase by 1.2 cpl for adopters 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 10 times larger than OLS estimates, consistent with our intuition that the endogeneity of the adoption decision is attenuating the OLS estimates towards zero. The magnitude of effects is consistent 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).

[46] 2SLS regressions using alternative instruments based on broadband availability and quality also show that mean margins and mean prices increase after adoption (see Table F6). See Appendix F.4 for additional discussion of these instruments and results.

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

We confirm that our IV approach resolves concerns about diverging parallel pre-trends between adopters and non-adopters in two ways. 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 following reduced form regression:

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

where this is the same regression as Equation (2) but with the instrument in place of the endogenous "treatment" variable. As before, 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 $\times$ 13+ Months Pre)$_{it}$ is instrumented with (Share Brand Adopters $\times$ 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 adoption on station-level prices and margins. Algorithmic pricing can increase station 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 price elasticity of demand changes over the day or the week and adjust prices accordingly. In this section we describe how we isolate the effects of adoption on competition. We do this by separating the sample by whether or not the station is a monopolist (e.g., without any nearby competitors). We then evaluate how AP adoption affects strategic interaction between stations by focusing on duopoly and triopoly markets and testing whether the adoption of only one or of many competitors triggers changes in outcomes.

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 market monopolists and stations that are not monopolists.[47] If the adoption of algorithmic pricing software does not change competition but benefits station operations in other ways, we should 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 (0.012)	-0.004 (0.009)	0.012*** (0.002)	0.013*** (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

[47] Markets are defined according to a clustering algorithm based on driving time between stations (see Appendix B). As a robustness check, we use an alternative market definition based on ZIP codes. See Appendix F.2 for additional discussion and results.

non-monopolist stations are driving the increase in mean margins, with mean margins increasing for non-monopolist adopters by 1.2 cents post-adoption (15%). By comparison, monopolist adopters have a small and non-statistically significant changes in mean margins. Average price effects are similar to margins. Mean monthly prices for non-monopolist stations increase by 1.3 cents per litre. Average prices of monopolist adopter stations do not change.

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 2021). 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, noon, 5pm and 7pm at each non-holiday weekday and then average those 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. Monopolist AP adopters show a different pattern, likely reflecting an improved ability to 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 D3 in the appendix). As was the case 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 estimates 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, Tolvanen and Wagner 2021).

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

	(1)	(2)	(3)	(4)
Outcome:	Mean 9am Price	Mean 12pm Price	Mean 5pm Price	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

AI-powered 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, AP technology likely similarly helps with better pricing but could also generate strategic effects. Promotional materials by several AP software providers reference how their software could help avoid price wars or other price decreasing strategies (Kantify, Derakshan et al 2016). The strategic effects of AP adoption on competition could work against the beneficial effects and against targeted price discrimination. Adopting stations may not want to fluctuate their prices to better target consumer demand if they know that competing algorithms are highly responsive to their price changes. This is consistent with our findings that prices are more stable during the day for AP adopting non-monopolists. In the next section, we more carefully examine the role of strategic interactions between AP adopters in duopoly and triopoly markets.

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 *markets*.[48] 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.[49] The variable "All Stations

[48]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 F.2 for additional discussion.

[49]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 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.

Adopted" is equal to one in market m in month t if all stations in this market are adopters.[50] 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.[51]

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.[52] 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 that have been identified as AP adopters in month t. Share Brand Adopters_{2mt} is the share of other stations belonging to market m station 2's brand that have been identified as AP adopters in month t.[53]

The first stage regressions are as follows:

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

where we include market and time fixed effects (αs) in each first stage regression as well as all time varying controls.

[50] In duopoly markets this variable can be expressed as: $(\text{Adopter} \times \text{post-Adoption})_{1mt}(\text{Adopter} \times \text{post-Adoption})_{2mt}$, 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 $(\text{Adopter} \times \text{post-Adoption})_{1mt}(\text{Adopter} \times \text{post-Adoption})_{2mt}(\text{Adopter} \times \text{post-Adoption})_{3mt}$.

[51] 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").

[52] 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 F.4.

[53] 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 $IV^1_{mt} = B_{1mt}(1-B_{2mt})(1-B_{3mt}) + B_{2mt}(1-B_{1mt})(1-B_{3mt}) + B_{3mt}(1-B_{1mt})(1-B_{2mt}) + B_{1mt}B_{2mt}(1-B_{3mt}) + B_{1mt}B_{3mt}(1-B_{2mt}) + B_{2mt}B_{3mt}(1-B_{1mt})$.

Table 9 presents estimates of Equation (4) using the instruments defined in Equation (5) and market-level margins and prices as the outcome variables of interest.[54] 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.

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.

A possible explanation for not seeing changes in mean market-level margins after incomplete adoption could be because 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 from these regressions are in Table D2 in Appendix D. We do not see any statistically significant changes in margins and 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. Theoretical literature suggests that it is possible for algorithms to facilitate collusion (Calvano et al 2020, Miklós-Thal and Tucker 2019).[55] We cannot be sure what type of algorithms station-owners are using and whether they fully turn over pricing decisions to algorithms. Nonetheless, lack of margin changes from partial/asymmetric adoption and substantial increases in margins and prices after complete adoption is suggestive of algorithms facilitating 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). We present additional evidence on the mechanism through which algorithmic competition

[54] Analogous results for ZIP code markets are in Table F3. See Appendix F.2 for additional discussion of alternative market definitions.

[55] 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 (Harrington 2018b). 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).

affects margins in Section 6.

Table 9: IV Duopoly and Triopoly Market Estimates

Outcome:	(1) 2SLS Mean Mkt Margin	(2) 2SLS Mean Mkt Price	(3) 1st Stage Not all Stations Adopted	(4) 1st Stage All Stations Adopted	(5) Reduced Form Mean Mkt Margin	(6) Reduced Form 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.

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, Fershtman and Pakes (2021) show that their less-sophisticated *asynchronous* algorithm converges to something close to the monopoly price, but takes considerable time to do so. In their paper demonstrating that simple AI algorithms can behave according to tacitly-collusive "punishment" strategies, Calvano et al (2020) show that convergence to these strategies takes time. Alternative explanations of supra-competitive pricing by algorithms do not show similar temporal patterns.[56]

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^1_{mt} and IV^2_{mt} 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.[57]

Consistent with simulation results in Calvano et al (2020) and in Asker, Fershtman and Pakes (2021), we find that for roughly the first year after market-wide AP adoption there are no statistically significant changes in average market margins at the 95% confidence level.[58] The magnitude of estimated coefficients for this time period is also quite small relative to our average estimates in Table 9. The average effects we estimate in Table 9 come in 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 average 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. Igami and Sugaya (2019) show that 1990s Vitamin cartels took several years to increase

[56] There are at least two alternative explanations for why algorithms could reach margins above competitive levels. First, pricing algorithms could *fail to learn to compete effectively* (Cooper et al 2015, Hansen, Misra and Pai 2020). For example, algorithms may not fully incorporate rivals' prices or may not best respond to these prices. In this case though, if margins were high, they would remain so initially and then possibly decrease over time as the algorithms learned to compete. Second, according to Brown and MacKay (2021), adoption of algorithmic software changes the game that firms play from a standard simultaneous Bertrand pricing game to a stage game. This increases prices and margins relative to a simultaneous Bertrand-Nash equilibrium. 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 D2 shows results from a regression of a non-adopting stations' margins on a dummy variable of whether its rival has adopted algorithmic pricing technology (instrumented by the rival brand's adoption share). We find that there are no statistically significant changes in margins following a rival's adoption. Although the Brown and MacKay (2021) model appears to fit well certain settings (such as cold medicine markets), in our context it does not seem to apply.

[57] 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.

[58] 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 that 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, Raymond and Bergman (2021) find that various algorithms require approximately a year to converge to new stable strategies after perturbations in the underlying data.

their prices and margins. Clark et al (2020) also show a lengthy adjustment period to high prices for a Canadian bread cartel, as do Byrne and de Roos (2019) in the Australian retail gasoline market.

Table 10: IV Duopoly and Triopoly Additional Price Effects

Outcome:	(1) Mean Market Wholesale Margin	(2) Mean Market Price	(3) Prob. Response to Price Decrease	(4) 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.

We provide additional suggestive evidence of how algorithmic competition operates differently from non-algorithmic competition. There are no clear measures of conduct that can be identified in a reduced form setting without an underlying model. In our setting, developing such a model is not straightforward. It requires us to make assumptions about how gas stations are competing before the availability of algorithms, as well as assumptions about how the algorithms operate, and we do not know the precise algorithms used by firms in this market.[59] Nonetheless, we can empirically attempt to evaluate changes in pricing behaviour and the timing of these changes coming directly from duopoly algorithms competing against one another.

[59] Many price setting algorithms including the Q-learning algorithm in Calvano et al (2020) are not designed to play mixed strategies. Other algorithms, as well as humans, are able to play mixed strategies. There are many possible asymmetric equilibria and characterising them without further information is not feasible. We leave this question for future research.

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 what their competitors do. We test for whether the conditioning behaviour evolves differently in markets with full AP adoption and markets 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 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.[60]

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, once again suggesting that there is gradual learning of new strategies by the algorithms. The magnitude of increased responsiveness after market-wide adoption 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 the propensity of stations to respond to price increases by their rivals 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.

[60] Although these regressions share some similarity with one of our measures of adoption (response time to a rival's price changes), we believe that they are sufficiently distinct. Here we explicitly allow for stations to both have immediate changes in responsiveness (which would identify the initial structural break) as well as longer term evolution in responsiveness that identifies changes in competitive strategy. We also separately responsiveness to rival price increases and price decreases.

7 Robustness

In this section we briefly outline a series of checks that confirm the robustness of our results to alternative samples, market definitions, adoption classifications and instruments. Results along with further details are in Appendix F. In every case results on the impact of adoption on margins are robust to the proposed check.

1. **Alternative Estimation Samples (Appendix F.1):** We re-estimate the main regressions with alternative samples of stations. We address possible contamination from the Shell price matching promotion from 2015 by (i) dropping observations from markets containing Shell stations, and (ii) dropping all observations from 2016. We also address potential concerns about the entry/exit of stations from the sample by looking at 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 F.2):** We use an alternative market definition, using 5-digit ZIP codes to define a market. A ZIP code is a well-defined unit of population in space, such that rural ZIP codes are bigger geographically. We find similar results for this market definition.

3. **Alternative Adoption Definitions (Appendix F.3):** We test the robustness of our "adopter" definition by using alternative classifications. 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 also test a classification that drops responsiveness to crude oil price shocks. We also consider an alternative definition regarding the time between structural breaks. While the baseline model classifies a station as an adopter if they experience a structural break in at least two out of four measures within an 4 week period, in our alternative definition we label stations as adopters if they experience structural breaks in at least two out of three measures within a period of 2 weeks. Last, we consider an alternative definitions where a station is classified as an adopter if they experience structural breaks in their pricing of Diesel fuel, or structural breaks in both E5 and Diesel.

4. **Alternative Instruments (Appendix F.4):** We propose using the availability of broadband access in station i's region as an instrument for adoption. Intuitively, if a station has access to high speed internet and/or reliable internet signals, it should be more likely to adopt algorithmic

pricing technology. We measure whether the local area around the gas-station has widespread access to high speed internet in a particular year. We also introduce a "placebo" instrument. Rather than using the share of stations of station i's brand that adopted as an IV, we use the share of stations by *another* brand (i.e., the brand of some station k in the market of station i). We expect that there should 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:** We use E5 gasoline since it has the highest market share (80%) in Germany. In Assad et al (2020) we use E10 gasoline instead of E5 gasoline. Results are qualitatively and quantitatively similar to the ones in this paper.

8 Policy Discussion and Conclusions

We investigate potential links between algorithmic pricing and competition by looking at the widespread introduction of AP software in the German retail gas market. First, we identify which stations have adopted this pricing software through structural break tests in various measures of behaviour during a sample period of 2016-2018. We then analyze the impact of algorithmic-pricing adoption by comparing competition measures for adopting stations vs. non-adopting stations.

To identify algorithmic-pricing adoption, we focus on stations that experience structural breaks in at least two out of four measures of pricing behaviour within a 4 week period. The measures capture the responsiveness of a station's pricing behaviour to the state of the market. We expect algorithmic pricing to increase this responsiveness. Comparing breaks in (i) the number of price changes, (ii) rival response time, (iii) responsiveness to crude oil price shocks, and (iv) responsiveness to local weather shocks, we find that the vast majority of breaks occur in mid-2017, the time at which the AP software became widely available.

Having identified adopting stations we investigate the effects of algorithmic adoption on the mean and the distribution of daily margins and prices. Due to the potential endogeneity of station-level adoption decisions, we instrument for station i's adoption using the share of stations in i's brand that have adopted. Results indicate that, overall, AP-adopters with nearby competitors increase mean margins by 15% on average in comparison to pre-adoption levels. Mean prices also increase. In contrast, adopters that are a monopolist do not see changes in their mean margins or prices. Looking at duopoly and triopoly markets exclusively, we find that there is no difference in market-level margins between markets in which no stations adopted and markets in some but not all stations adopted. However, markets in which all stations adopted show a mean margin increase of nearly 36%. Mean prices increase by approximately 6 cents per litre. The magnitude of these estimates is consistent with what AP software providers describe in their marketing materials.

We investigate the mechanism behind the increases in margins by looking at the timing of effects. If algorithms *fail to learn to compete effectively* we should see immediate increases in margins after both stations in duopoly markets adopt AP. If algorithms *learn how not to compete*, we should see no initial effects followed by eventual convergence to high prices and margins. This is what we find in the data - margins in markets where all stations adopt do not substantially change for about a year after adoption and then increase. This is suggestive of algorithms learning tacitly-collusive strategies over time. Overall, the results indicate that the adoption of algorithmic pricing has affected competition and facilitated tacit-collusion in the German retail gas market.

Our findings suggest that regulators should be concerned about the mass-adoption of algorithmic pricing software. Multiple antitrust authorities and economic organizations (OECD 2017; Competition Bureau 2018; Autorité de la Concurrence and Bundeskartellamt 2019; UK Digital Competition Expert Panel 2019) have released reports on algorithmic collusion and competition law. The reports agree that explicit algorithmic collusion would not require any changes to existing competition laws, but would change how competition authorities monitor for and investigate collusive practices. Increased tacit collusion through algorithms could change the legal status of such forms of collusion (in addition to monitoring and investigative practices). Currently, tacitly collusive behaviour 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 retail gasoline markets in Germany (where extraordinarily detailed pricing data are available), the same algorithmic pricing software is adopted in gasoline retail markets around the world. At a minimum, our results suggest that competition authorities should undertake a census of retail-gasoline pricing software to understand the market structure of the algorithmic software market and the extent of adoption. Such a census can help separate whether the main effect of algorithmic pricing 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.[61]

[61] 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). Algorithms in this case serve as the "hubs" of a hub-and-spoke cartel (Harrington 2018b).

Our focus in this paper 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 (2021) present evidence of algorithmic pricing by pharmaceutical drug retailers online. *Vendavo*, an AI based retail pricing software reports over 300 global deployments in manufacturing, chemicals, distribution and high tech industries (Vendavo.com). 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 2021). In a platform setting, Johnson, Rhodes and Wildenbeest (2021) propose simple market design features that can disrupt algorithmic price-increasing strategies, and such features may have wider applicability in other markets.

References

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

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. Aryal, G., Ciliberto, F., and B. Leyden (2021). Coordinated Capacity Reduction and Public Communication in the Airline Industry, forthcoming *Review of Economic Studies*.

6. Asker, J., C. Fershtman and A. Pakes (2021). Artificial Intelligence and Pricing: The Impact of Algorithm Design.

7. 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

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

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

10. 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

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

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

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

14. Brown, Z. and A. MacKay (2021). Competition in Pricing Algorithms, forthcoming *American Economic Journal: Microeconomics*.

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

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

17. Cabral, L., N.S. Durr, D. Schober, and O. Woll (2018). Learning Collusion: Theory and Evidence from the Shell Price Matching Guarantee.

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

19. 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.

20. 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.

21. 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.

22. 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.

23. Clark, R., I. Horstmann and J-F. Houde (2020). Two-sided hub-and-spoke collusion: Evidence from the grocery supply chain. Mimeo.

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

25. 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.

26. Crede, C.J. (2019). A structural break cartel screen for dating and detecting collusion. *Review of Industrial Organization* 54, 543-574.

27. Decarolis, F. and G. Rovigatti. (2021). From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising, forthcoming *American Economic Review*.

28. 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.

29. 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.

30. 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(5), 302-305.

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

32. Dubois, P. and L. Lasio (2018). "Identifying Industry Margins with Price Constraints: Structural Estimation on Pharmaceuticals." American Economic Review, 108 (12): 3685-3724.

33. Dubé, J.P. and S. Misra (2021). Personalized Pricing and Customer Welfare, forthcoming *Journal of Political Economy*.

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

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

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

37. Ezrachi, A. and M. Stucke (2016). *Virtual Competition*, Harvard University Press, Harvard, MA.

38. 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*.

39. 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 (No. w22969). National Bureau of Economic Research.

40. Garcia, D., J. Tolvanen and A.K. Wagner (2021) Demand Estimation Using Managerial Responses to Automated Price Recommendations. Management Science, forthcoming.

41. 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.

42. 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.

43. Hansen, K., K. Misra, and M. Pai (2020). Algorithmic Collusion: Supra-competitive Prices via Independent Algorithms

44. Harrington, J. (2008). Detecting Cartels. in Handbook of Antitrust Economics (P. Buccirossi Ed.), MIT Press, Cambridge, MA.

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

46. Harrington, J. (2018b). How do hub-and-spoke cartels operate? Lessons from nine case studies. Mimeo.

47. Igami, M. and T. Sugaya (2019). Measuring the incentive to collude: The vitamin cartels, 1990-1999. Available at SSRN 2889837.

48. Johnson, J., A. Rhodes and M.R. Wildenbeest (2021) Platform Design when Sellers Use Pricing Algorithms. Available at SSRN 3691621.

49. 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.

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

51. Kehoe, P., B. Larsen, and E. Pastorino (2018). Dynamic Competition in the Era of Big data

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

53. Kühn K-U. and S. Tadelis (2018). The Economics of Algorithmic Pricing: Is collusion really inevitable?, mimeo.

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

55. Li, D., L. Raymond and P. Bergman (2021). Hiring as Exploration. Mimeo

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

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

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

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

60. Monpolkommission (2018), Algorithms and Collusion, *Biennial Report of the Monopolies Commission.*

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

62. Montag, F and C. Winter (2019) Price Transparency Against Market Power, mimeo.

63. 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.

64. 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

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

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

67. 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.

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

69. Shiller, B. (2014). First-Degree Price Discrimination Using Big Data

70. 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.

71. Shoham, Y., Powers, R., & Grenager, T. (2003). Multi-agent reinforcement learning: a critical survey. mimeo.

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

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

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

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

76. Varian, H. (2018). Artificial intelligence, economics, and industrial organization (No. w24839). National Bureau of Economic Research.

77. Wang, Z. (2008). Collusive Communication and Pricing Coordination in a Retail Gasoline Market. *Review of Industrial Organization* 32, 35-52.

78. 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.

Appendices: For online publication

A Background

Figure A1: December 2017 TANKSTOP Trade Magazine Cover and a2i Advertisements

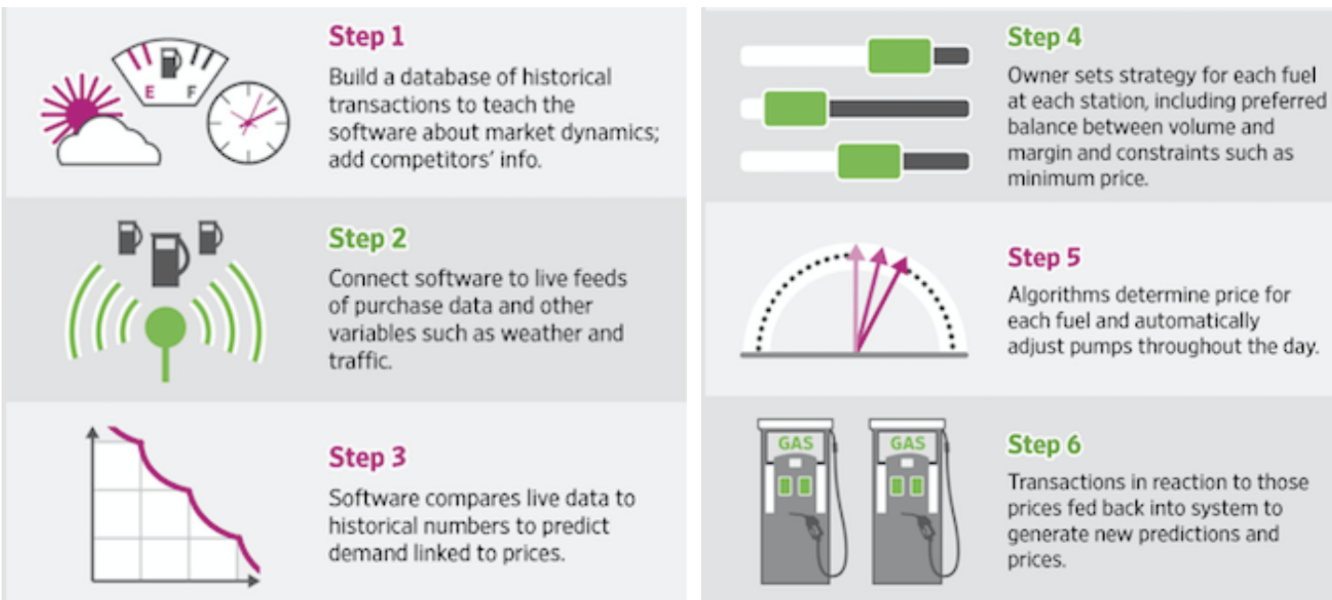


Figure A2: How Algorithms Work (wsj.com)

B Cluster Markets

Following Carranza, Clark and Houde (2015) and Lemus and Luco (2019), we define geographic markets using a hierarchical clustering algorithm based on driving time between gas stations. We calculate the travel time between each pair of stations in our sample with OpenStreetMap data using the Open Source Routing Machine. The algorithm begins with each station in its own market/cluster. It then creates a new larger cluster by linking the station-specific clusters with other nearby clusters, such that the average driving time between all stations in the larger cluster is equal. The algorithm then repeats this exercise, enlarging the market/cluster by merging additional clusters with stations that are further away. The algorithm thus creates a hierarchical tree, where each level of the tree groups each station into larger and larger markets / clusters based on driving time.

Additional details about the algorithm can be found in Appendix C of Lemus and Luco (2019), but the important feature of this approach is that the researchers do not set the total number of clusters / markets or the number of stations per cluster. Instead the researchers choose (1) where to "prune" the hierarchical tree, and (2) how to define a monopoly market (based on driving time). Pruning is done based on an "inconsistency" measure, which captures differences in the "height" of the cluster below a link in the tree.[62] Fewer clusters are formed when the inconsistency measure is larger. We choose an inconsistency threshold pruning the tree at the 80th percentile of the distribution of inconsistency. We also experimented with other pruning thresholds, generating similar results.[63] We chose to define monopoly stations as stations that are at least 20 minutes away from the nearest rival. This is a stricter bound than what is used by the German Federal Cartel Office (FCO), which uses a driving time of 30 minutes in cities and 60 minutes in rural areas to specify markets. However, many German academics and practitioners critiqued the FCO's approach as being unreasonably loose, grouping together gas stations that are located in cities very far apart and assuming that a consumer will drive as much as an hour to fill up their tank (Bantle, Muijs and Dewenter 2018).[64]

Some summary statistics about our cluster markets are in Figure B1. Panel (a) shows the distribution of the total number of markets by the number of stations in each market. There are

[62] For example, suppose are two stations within a minute of each other in a cluster. Adding another station that is within two minutes of driving time to each of these stations will increase the "height" of the tree by one. But merging this cluster with a cluster where stations are on average twenty minutes away from the original two stations will increase the "height" of the tree by substantially more, creating a very large difference in height between the smaller and larger clusters/ levels of the tree.

[63] Lemus and Luco (2019) also found similar robustness to varying the threshold.

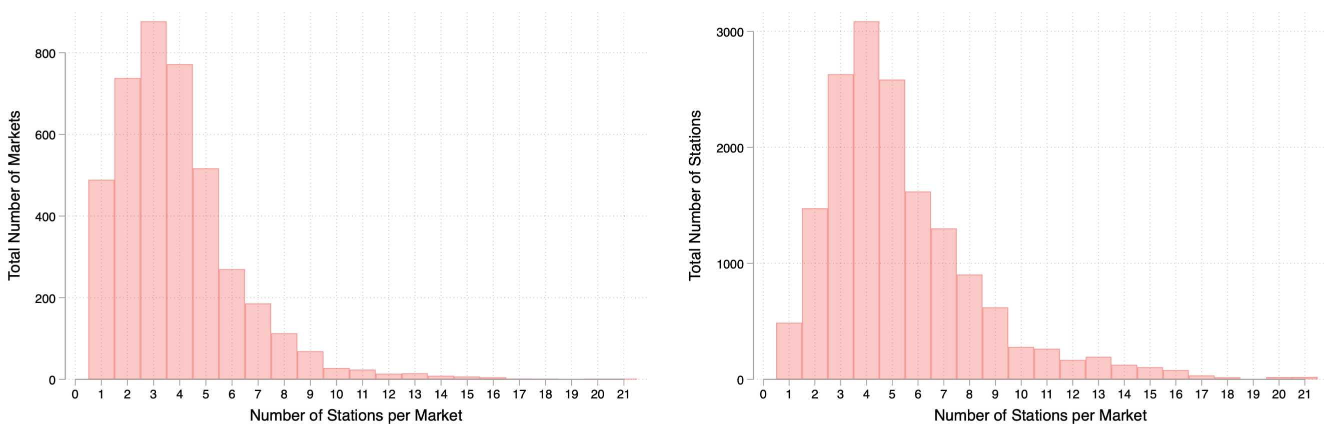
[64] Lemus and Luco (2019) use a 30 minute threshold in the Chile, which is very different geographically and in terms of population distribution than Germany. In their setting, a 25 or 20 minute monopoly definition does not change the number of monopolists substantially.

489 monopoly markets, 738 duopoly markets, and 877 triopoly markets, and only slightly more than 100 markets with more than 10 stations in the market. Panel (b) shows the distribution of the total number of stations, by the number of stations in each market. Out of 16,000 stations, there are 489 monopolists, 1,476 duopolists, 2,631 triopolists, and 3,088 stations that belong to 4 station markets. The average station belongs to a 5 station market.

Figure B1: Distribution of Stations and Markets

(a) Total Number of Markets

(b) Total Number of Stations



Notes: Panel (a) shows a histogram of the number of market by market size (in terms of the total number of stations in each market). Panel (b) shows a histogram of the number of stations by market size (in terms of the total number of stations in each market).

C QLR Estimation and Results

C.1 QLR Estimation

We estimate the following regression over a range of eligible break periods $\tau_0 \leq \tau \leq \tau_1$ (eligible break periods are measured by week):

$$y_{it} = \alpha_i + \beta_i D_t(\tau) + X_t \gamma_i + \epsilon_{it}, \tag{7}$$

where y_{it} is the variable of interest for station i in time t, $D_t(\tau)$ is a dummy variable equal to 0 if $t < \tau$ and 1 if $t \geq \tau$, and X_t is the crude oil price in time period t, which we use as a control variable. For each regression we test the null hypothesis $H_0 : \beta_i = 0$ and compute the F-statistic $F_i(\tau)$. The QLR statistic is the largest of these F-statistics over the range of eligible break dates:

$$QLR_i = max[F_i(\tau_0), F_i(\tau_0 + 1), ..., F_i(\tau_1)]. \tag{8}$$

The best candidate structural break period for station i is identified as the week τ^* that satisfies $QLR_i = F_i(\tau^*)$.[65] Structural breaks are identified as significant if they exceed a certain critical value.[66] We drop all stations from our data set that do not operate in every week in 2017 (i.e. we keep stations that have 52 observations of a given measure in 2017). We use 30% trimming for our test period, which is standard for QLR testing.[67]

[65] We refer to the QLR statistic as identifying the "best candidate" structural break period because if we look at a test for each time period τ individually, there may be multiple periods in which a structural break would be identified (i.e. has an F-statistic exceeding a certain critical value). The QLR statistic identifies the best candidate break period as it identifies the period with the most significant associated F-statistic.

[66] The distribution of the QLR statistic is non-standard so we cannot use the usual critical values for F-statistics to determine significance. Critical values for QLR statistics are taken from Andrews (2003). Using these values we measure a structural break as significant at the 10% level if $QLR_i \geq 7.12$, at 5% level if $QLR_i \geq 8.68$, and at the 1% level if $QLR_i \geq 12.16$.

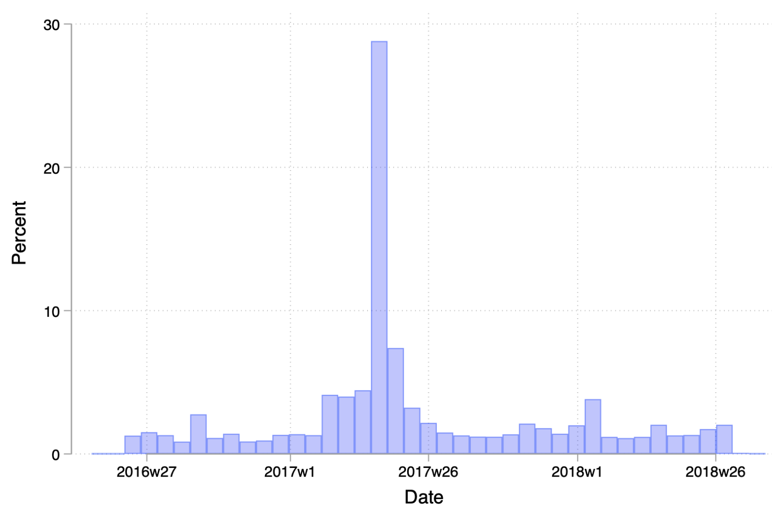
[67] We use as our first eligible break date the 15th percentile week in our sample period and as our last eligible break date the 85th percentile week in our sample period.

C.2 Structural Break Test Results

C.2.1 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.[68] 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. Figure C1 shows the overall distribution of best-candidate breaks.

Figure C1: Frequency of Best-Candidate Structural Breaks in Number of Price Changes (12,919 stations included)



Notes: this histogram shows the distribution of best-candidate QLR structural break weeks for the number of price changes.

C.2.2 Rival Response Time

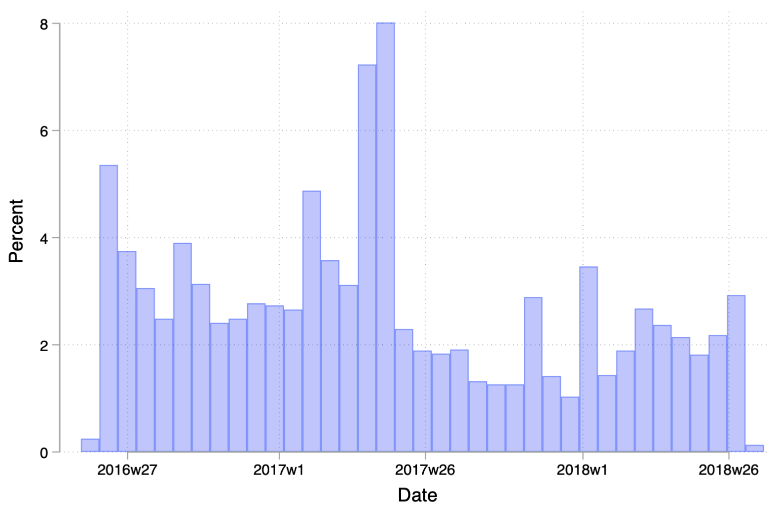
We define a rival for station i as the closest station j that is within a 1km radius of station i but that belongs to a different brand.[69] Rival response time for station i is calculated as the number of minutes between the time of a price change by rival j and the subsequent price change by station i. If station i changes its price more than once before station j makes a price change, rival response

[68] 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.

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

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. Figure C2 shows the overall distribution of best-candidate breaks.

Figure C2: Frequency of Best-Candidate Structural Breaks in Rival Response Time (5,227 stations included)



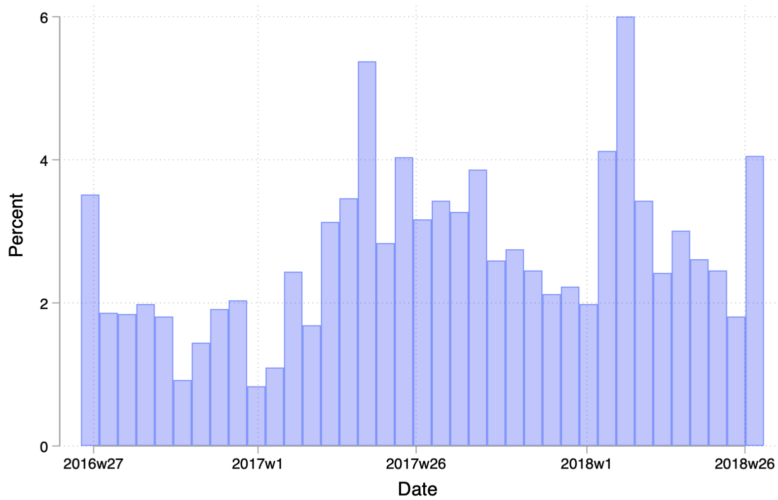
Notes: this histogram shows the distribution of best-candidate QLR structural break weeks for the response time to a rival's price changes.

C.2.3 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 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). Figure C3 shows the overall distribution of best-candidate breaks.

Figure C3: Frequency of Best-Candidate Structural Breaks in Responsiveness to Oil Price Shocks (5,747 stations included)



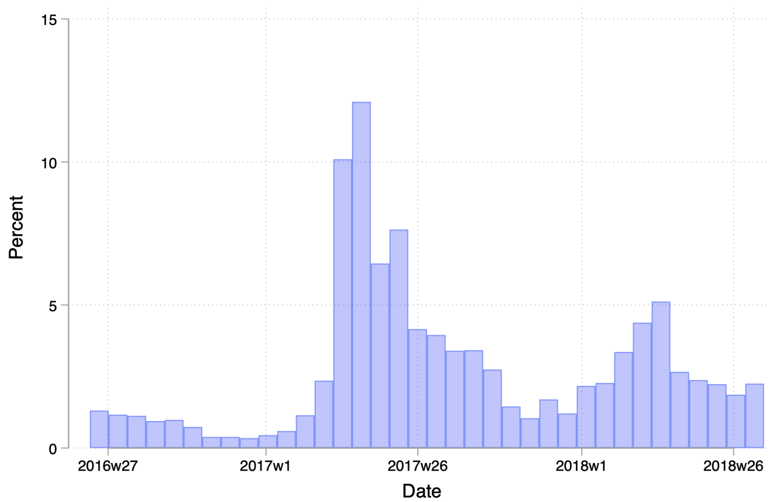
Notes: this histogram shows the distribution of best-candidate QLR structural break weeks for the number of responses to oil price shocks (conditional on the number of shocks and the number of station price changes).

C.2.4 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. Figure C4 shows the overall distribution of best-candidate breaks.

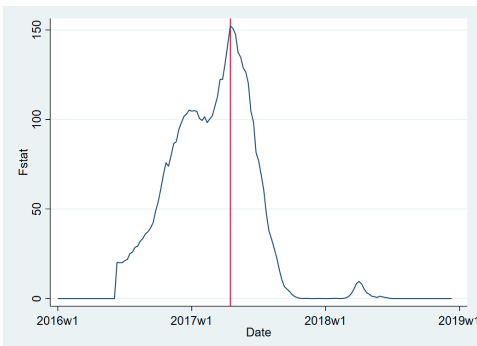
Figure C4: Frequency of Best-Candidate Structural Breaks in Responsiveness to Local Weather Shocks (4,892 stations included)



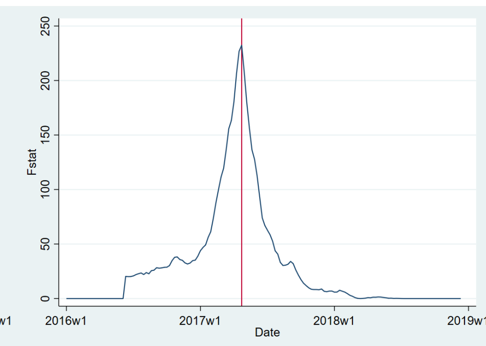
Notes: this histogram shows the distribution of best-candidate QLR structural break weeks for the number of responses to local weather shocks (conditional on the number of shocks and the number of station price changes).

C.3 Alternative Structural Breaks

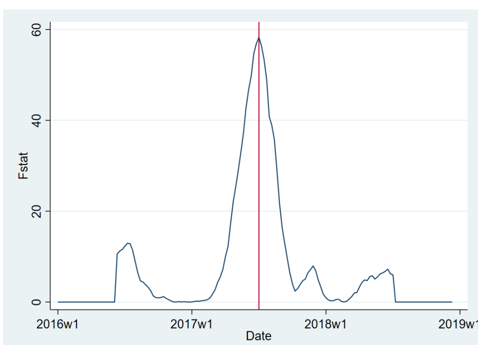
We look at the distribution of F-statistics for structural break tests in the number of price changes for stations over the sample period for a few representative stations. We find that generally, stations display a uni-modal distribution in their F-stastistics, meaning we are unlikely to find best-candidate breaks at a significantly different date if we were to, for example, take the second highest F-statistic rather than the maximum. A few examples are shown in C5 of what a typical distribution would look like for a station's F-statistics for structural break tests in the number of price changes.



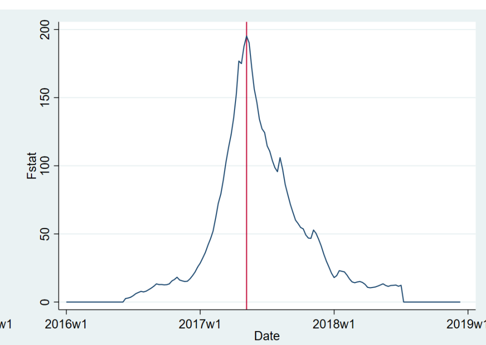
(a) Station 22 (Esso)



(b) Station 29 (Hem)



(c) Station 31 (Shell)



(d) Station 163 (ARAL)

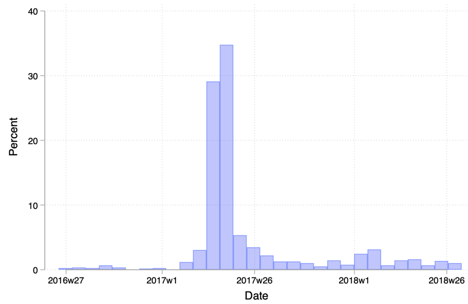
Figure C5: Distribution of F-statistics for Structural Break Tests in Number of Price Changes

To take a more systematic approach to test whether there may be significantly different alternative break dates, for each station, we look at the dates associated with the 2nd highest F-statistics for break tests in the number of price changes. We find that for 75% of stations, these dates are 1 week apart meaning that the next alternative break date would be occur either 1 week before or after the break associated with the highest F-statistic. We find only 10% of stations have difference of 3 or more weeks between the dates associated with the highest and 2nd highest F-statistic.

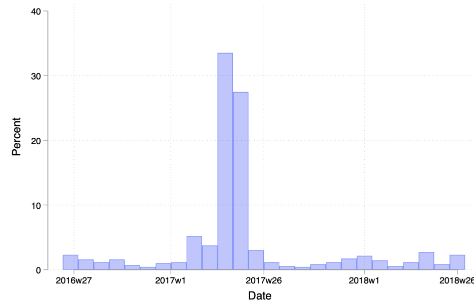
C.4 Distribution of Average Break Dates by Measure Combination

Figure C6: Frequency of Average Break Date for Measures Breaking Within 4 Weeks

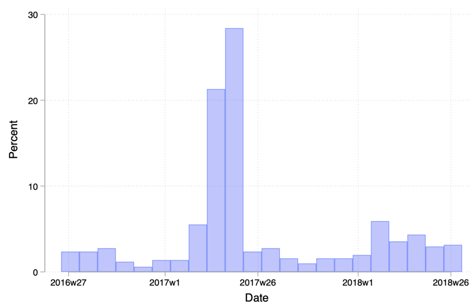
(a) Number of Price Changes and Weather (1,182 stations)



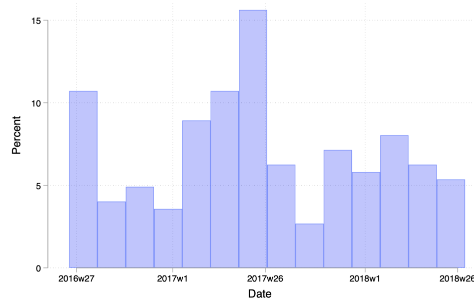
(b) Number of Price Changes and Rival Response Time (695 stations)



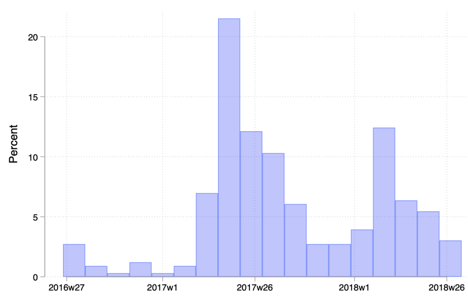
(c) Number of Price Changes and Oil (507 stations)



(d) Rival Response Time and Oil (224 stations))



(e) Weather and Oil (330 stations)



(f) Rival Response Time and Weather (234 stations))

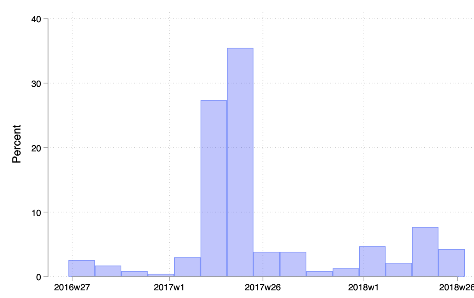
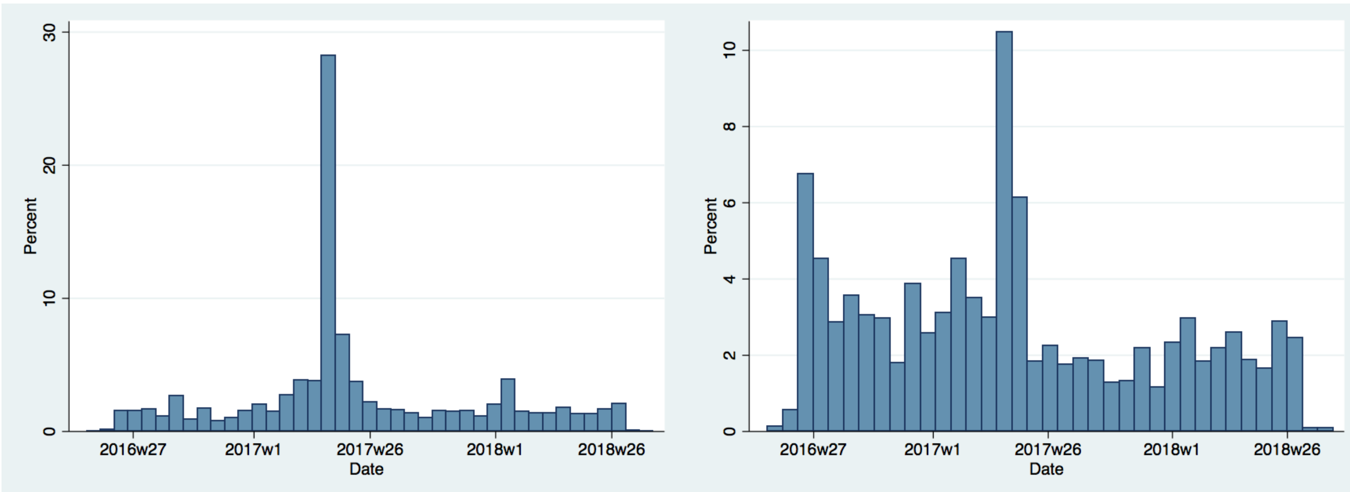


Figure C6 shows the distribution of the average break date for each combination of measures, where the average break date is the average year-week between each measure's best-candidate break date.

For each measure pair, the largest frequency of average break dates occur in mid-2017. Overall, we see the largest frequency of multiple measure breaks in mid-2017, the suspected period of large scale adoption, suggesting these measures accurately represent changes related to adoption of algorithmic pricing.

C.5 Diesel Gas Structural Breaks

Figure C7: Frequency of Significant Structural Breaks in Number of Price Changes and Rival Response Time (Diesel Gas)



C.6 Heterogeneity in Brand Level Adoption

We test whether the heterogeneity in brand-level adoption probability is explained by observable brand characteristics. Unlike station-level adoption, brand level adoption is not correlated with brand-level observables after controlling for brand size (the number of stations in the brand). Table C1 shows that conditional on the number of stations in the brand, the share of brand adopters is uncorrelated with average demographic characteristics of a brand's stations. It is also uncorrelated with the average number of competitors that a brand's stations have. This makes intuitive sense. Brands likely spread out their stations across different markets. Local characteristics will inevitably average out. Brands also make broad strategic decisions that should not be influenced by local market conditions. The only statistically significant correlate of adoption probability at the brand level seems to be brand size. Because of this, we control for brand size in the IV regressions.

Table C1: Correlates to Brand-Level Adoption Probability

	(1)
Outcome:	Share Brand Adopters
Mean Population Density	0.00003
	(0.00003)
Mean ln(Population)	-0.13789
	(0.11453)
Mean Median Age	0.00667
	(0.00763)
Mean Employment Share	-0.51591
	(0.37037)
Mean ln(region GDP)	0.13839
	(0.10112)
Mean N Competitors in ZIP	0.00147
	(0.00557)
N Brand Stations	0.00003**
	(0.00001)
Observations	6,853

Notes: The sample for this regression includes brand/month observations from January 2016 until December 2018 for brands with two stations or more. The outcome is the share of a brand's stations that are labelled as adopters by month t. Variable "Mean X" is a simple average of variable X across all brand b stations in month t. We include year-month fixed effects. Standard errors clustered at the brand level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

D Additional Estimates

D.1 Flat Specification

Instead of the station-month data in the main text, we use a "flat" two time period data structure where we calculate average daily outcomes and characteristics for each station i for the Jan 2016-May 2017 period and for the June 2017-December 2018 period. In this specification, $t \in \{\text{Pre June 2017, Post June 2017}\}$. Our OLS specification is as follows:

$$y_{it} = \beta(\text{Adopter} \times \text{Post June 2017})_{it} + \alpha_t + \text{Post June 2017}_t + \gamma X_{it} + \epsilon_{it}, \tag{9}$$

where y_{it} is the outcome variable for station i in time $t \in \{\text{Pre June 2017, Post June 2017}\}$, α_i are station fixed-effects and Post June 2017_t is a dummy equal to one for post June 2017 observations. Adopter_i is a dummy equal to one for all stations that are labelled as adopters. X_{it} are time-varying station specific controls (local demographics and weather).

The two period specification is potentially less subject to endogeneity concerns than the station-month specification, since it flattens out the differences between stations that adopt AP late and stations that adopt AP early. Those differences might be driven by underlying unobservable heterogeneity. The flat specification also helps with identification since it flattens the event study setting into a classic difference-in-differences setting with two time periods.[70]

OLS estimates for this specification are in Columns (1) and (2) of Table D1. They show that estimated OLS effects are larger in this specification than in the one in the main text. The estimated effect of adoption on mean margins is still 0.1 cpl, but the estimated effect on mean prices is three times bigger at 0.3 cpl. Nonetheless, we are still concerned about endogeneity and construct a version of the instrument from the main text: the share of brand adopters for each station by the end of our sample (excluding the focal station i), multiplied by a Post June 2017_t dummy. Estimates from the IV regression are in Columns (3) and (4) of Table D1. Once again, they show that endogeneity concerns are valid. Estimated effects are ten times larger in the 2SLS than in the OLS. The IV estimate of the effect of adoption on mean margins is 1.7 cpl, and the effect of adoption on prices is estimated to be 3.2 cpl. 2SLS effects estimated in this specification are larger than in the main text. We choose to primarily use the effects in the main text to be more conservative. As well, the full event study specification allows us to test for parallel pre-trends and for the timing of the effects.

[70] As Goodman-Bacon (2020) shows, identification in event studies may be challenging.

Table D1: "Flat" Regression Estimates

	(1)	(2)	(3)	(4)
	OLS	OLS	2SLS	2SLS
Outcome:	Mean Margin	Mean Price	Mean Margin	Mean Price
Adopter × Post June 2017	0.001***	0.003***	0.017***	0.032***
	(0.000)	(0.000)	(0.003)	(0.003)
Station FE	YES	YES	YES	YES
Post June 2017 Dummy	YES	YES	YES	YES
Pre/Post June 2017 Regional Demographics	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES
Observations	27,504	27,504	25,734	25,734

Notes: The sample includes two observations for all stations. One observation capturing the period between January 2016 and May 2017, and one observation capturing the period between June 2017 and December 2018. Mean Margin is an average of daily differences of pump price for station i in month t and wholesale price for each of the two periods. Mean price is an average of the daily pump price of station i for the two periods. "Adopter ×Post June 2017" is a dummy equal to 1 if the gas station experienced a structural break in at least 2 of 4 relevant measures at any point in time and the time period is after June 2017. "Share Brand Adopters", interacted with a "Post June 2017" dummy 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 we label as adopters (excluding station i). Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/ and pre/post June 2017 level. We also control for the number of stations belonging to station i's brand in each period, 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 each time period. Market level clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

D.2 Additional Estimates

Table D2: Rival Adoption Effects

	(1)	(2)
Outcome:	Mean Margin	Mean Price
One Rival Adopted	-0.000	0.001
	(0.001)	(0.001)
Two Rivals Adopted (in Triopoly Market)	0.000	0.002
	(0.002)	(0.002)
IVs	YES	YES
Station FE	YES	YES
Year-Month FE	YES	YES
Annual Regional Demographics	YES	YES
Weather Controls	YES	YES
Other Controls	YES	YES
Observations	131,275	131,283

Notes: The sample includes all station/month observations belonging to duopoly and triopoly markets from January 2016 until December 2018 where zero or one of the duopolists adopted AP, and zero, one or two of the triopolists adopted AP. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale price. "One Rival Adopted" is a dummy equal to 1 in month t if the duopoly rival of station i in market m (or one of the two triopolist rivals) experienced a structural break in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. "Two Rival Adopted" is a dummy equal to 1 in month t if two of the three triopolist rivals experienced structural breaks in at least 2 of 4 relevant measures in any previous month $\{1, ..., t-1\}$. Instruments for a rivals' adoption are the "share of brand adopters" of the rival in the market. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. 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

Table D3: IV Station-Level Estimates - Time Specific Prices

	(1)	(2)	(3)	(4)
Outcome:	Mean 9am Price	Mean 12pm Price	Mean 5pm Price	Mean 7pm Price
Adopter × post-Adoption	-0.003	0.033***	0.050***	0.024***
	(0.003)	(0.003)	(0.004)	(0.003)
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
Observations	448,131	448,200	448,221	448,221

Notes: Sample includes gas station/month observations from January 2016 until December 2018. 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

E Adoption of Electronic Payments Technology in 1990s

We use annual data from Kent Marketing, a leading survey company in the Canadian gasoline market.[71] It captures annual data from 1991 to 2001 for all retail gasoline stations in seven medium-sized markets in Ontario: Brantford, Cornwall, Guelph, Hamilton, Kingston, St. Catharines and Windsor. The 5 brands with most stations in this data are PetroCanada (98 stations), Esso (84 stations), Shell (61 stations), Sunoco (56 stations) and Pioneer (36 stations). The data includes station characteristics including whether the station accepts "electronic payments."

This is a good benchmark technology for AI adoption. Both could improve station performance as electronic payments allow for a wider set of consumers to purchase gasoline (and larger quantities of gasoline). As for AI, electronic payment companies also have HQ-level deals with retail gasoline brands, but individual station owners had to bear some of the costs of upgrading their equipment. For example, in 1997, Exxon Mobil (Esso's parent company) rolled out the Mobil Speedpass, a contactless electronic payment system. BusinessWeek reported that after the brand-wide rollout, individual Mobil station owners "have to install new pumps costing up to $17,000–minus a $1,000 rebate from Mobil for each pump" (BusinessWeek).

The first appearance of electronic payments at any gas station in the data is in 1993 (the third year of the dataset). Among the five largest brands, no one reached 50% adoption rates of this technology by 2001. The largest share of adopting stations is for Pioneer, where 46% of stations adopted by 2001. Figure E1 shows adoption rates by the top 5 brands (by the number of stations) in this data. It suggests that electronic payment adoption follows a highly staggered pattern. Of the 5 biggest brands, by 1998 (5 years after the technology became available) only two of the brands had *any* adoption. It is also brand specific. Some brands, such as Esso, appear to be continuously upgrading (or supporting the upgrade) of their stations. Stations by other brands, like Pioneer, adopt faster but later. This likely reflects brand-specific strategies.

[71] This is a subset of data used in Clark, Houde and Carranza (2015).

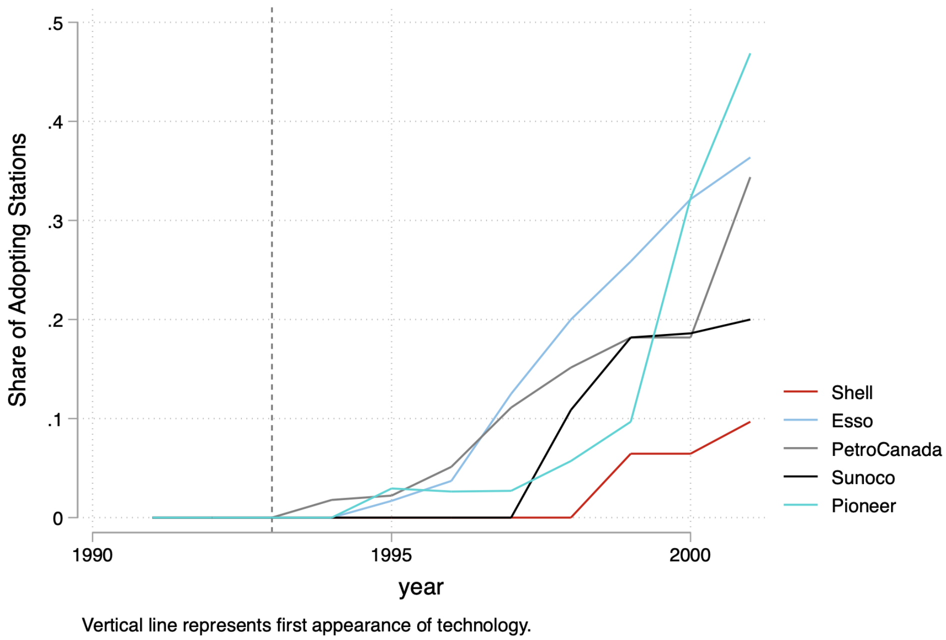


Figure E1: Share of Electronic Payment Adopters Among Top 5 Brands in Canada

F Robustness Checks

F.1 Alternative Estimation Samples

We perform a number of estimation-sample based robustness checks in Table F1. The first two robustness checks deal with concerns about the impact of Shell's 2015 price matching policy (see Section 3.1). The introduction of price matching in 2015 appears to have changed pricing strategies (Cabral et al 2021). These changes in strategies may still be ongoing in 2016. This would confound our results. Shell stations may be mistakenly labelled as algorithmic pricing software adopters. Shell stations and their competitors may also set higher prices due to the the price matching guarantee rather than due to the adoption of algorithmic pricing software.

Columns (1) and (2) in Table F1 deals with this concern by dropping all observations belonging to markets where the price matching guarantees would be relevant. This includes all Shell stations and stations that are in the same market as Shell stations. Results from this sample are quantitatively and qualitatively similar to the main estimates. Even without including any markets where Shell price matching guarantees would have an effect, we find that adoption of algorithmic pricing software increases average margins above wholesale prices by 1.4 cents. Column (2) drops all observations from 2016 (where the Shell effects would be most prominent). Results here are quantitatively similar to the main results.

Table F1: Sample Robustness Checks

	(1)	(2)	(3)	(4)
Sample:	No Shell Markets	Dropping 2016 Data	Balanced Sample	Market-Level Balanced Sample
Outcome:	Mean Margin	Mean Margin	Mean Margin	Mean Margin
	2SLS	2SLS	2SLS	2SLS
Adopter × post-Adoption	0.014***	0.011***	0.015***	0.010***
	(0.003)	(0.003)	(0.003)	(0.004)
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	238,124	300,528	293,496	132,394

Notes: All samples include only stations that are in markets with more than one competitor. Sample in Column (1) includes gas station/month observations from January 2016 until December 2018 that do not belong to a market where a station by a Shell brand is present. Sample in Column (2) includes gas station/month observations from January 2017 until December 2018 (dropping 2016 data). Column (3) includes all gas station/month observations belonging to gas stations that are present in every month of the sample. Column (4) includes all gas station/month observations belonging to gas stations that are present in every month of the sample and are in markets where the number of stations does not vary across the sample period. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale gasoline prices. "Adopter" 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 are clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We perform additional robustness checks to address concerns that our main results are driven by entry and exit of stations from the sample - either through the entry of high-quality and high-margin adopters, or through the exit of weak non-adopting stations. In Column (3) we look at a balanced sample of stations. We only include stations that are present in every month of the three year sample period. Results are qualitatively and quantitatively similar to our main estimates. But even these stations can be affected by entry and exit of other stations in their market. In Column (4), we look at a balanced sample of stations and markets that do not change over time. We only include stations that are present in every month of the three year sample period *and* we drop every market where the number of stations changes over time. Results from this subsample are also qualitatively and quantitatively similar to our main results.

F.2 Alternative Market Definition

There are many possible geographic definitions of "markets." In our main results, we define markets based on a hierarchical clustering algorithm that accounts for driving time between stations (Tables 7 and 9). Other commonly used definitions take advantage of existing geographic designations such as Census tracts, DMAs, or ZIP codes. In this section we use a 5-digit ZIP code as a market definition. Table F2 provides estimates of regressions similar to Table 7 but using the following definition of a monopoly: a station that has no competitors within its ZIP code. Non-monopoly stations then are those that have one ore more competitors in their ZIP code. Using this alternative definition yields qualitatively and quantitatively similar results to the results in Table 7. In Table F3 we replicate Table 9 using ZIP markets with two or three stations. We find comparable results to those in the main text.

Table F2: 2SLS Station-Level Results by ZIP Market Structure

	(1)	(2)	(3)	(4)
Outcome:	Mean Margin	Mean Price	Mean Margin	Mean Price
Sample:	Monopolist Stations		Non-Monopolist Stations	
Adopter × post-Adoption	0.007	-0.001	0.013***	0.014***
	(0.006)	(0.006)	(0.002)	(0.002)
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	67,300	67,300	380,826	380,826

Notes: Sample is 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 ZIP code. The other subsample includes only stations that have one or more competitors in their ZIP code. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale gasoline prices. Mean Price is the average pump price for station i in month t. "Adopter" 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. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. We also control for the number of competitors and for the number of adopting competitors in the ZIP. Standard errors clustered at the ZIP level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table F3: ZIP Duopoly Market Results

Outcome:	(1) Mean Mkt Margin 2SLS	(2) Mean Mkt Price 2SLS
Not all Stations Adopted	0.004	0.007
	(0.003)	(0.004)
All Stations Adopted	0.033*	0.070***
	(0.019)	(0.026)
IVs	YES	YES
Market FE	YES	YES
Year-Month FE	YES	YES
Annual Regional Demographics	YES	YES
Weather Controls	YES	YES
Other Controls	YES	YES
Observations	59,659	59,659

Notes: The sample includes duopoly and triopoly market/month observations from January 2016 until December 2018. A duopoly market is defined as a 5-digit ZIP code with only two stations. A triopoly market is defined as a 5-digit ZIP code with only three stations. Outcome variable Mean Market Margin is the average of mean market daily differences of pump prices for stations in market m in month t from wholesale price. "Not All Station Adopted " is a dummy equal to 1 in month t if one of the two duopolists, or one or two of the three triopolist 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 include 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. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. Standard errors clustered at the ZIP level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

F.3 Alternative Adoption Definitions

Tables F4 replicates Table 7 using alternative definitions of AP adoption. We consider six alternative definitions. Our baseline definition classifies a station as an adopter if, within a period of 4 weeks, it experienced a structural break in at least two out of four measures - number of price changes per day, the speed of response to a rival's price change, responsiveness to crude oil price shocks, and responsiveness to local weather shocks. A potential concern with this definition is that not all stations have a rival as we define it (another station within 1km). Such stations could then be less likely be defined as adopters. We address this concern by using only three of the four measures to define an adopter and an adoption date: the number of price changes per day, and responses to weather and oil shocks. Under this definition, a station is labelled as an adopter if it experiences a structural break in two or three of these measures within a period of 4 weeks. This is the definition we use in Column (1) of Tables F4. In Column (2), we repeat the same exercise, but drop the responsiveness to crude oil price shocks as a measure representing adoption. The distribution of structural breaks for this measure does not appear to be clearly defined, and the difference between stations with and without breaks is not as large as for the other measures. In Column (3), we label a station as an adopter if they experienced a structural break in any two out of four measures but within a period of 2 weeks. This is a stricter requirement for being labelled as an adopter. Results for these definitions are qualitatively similar to baseline results, and quantitatively even larger than the baseline results.

Table F4: Station Level Results with Alternative "Adopter" Definitions

	(1)	(2)	(3)	(4)	(5)	(6)
Adopter Measure:	No Rival Response (≥2 out of 3)	No Oil Shock Response (≥2 out of 3)	≥2 out of 4 (2 weeks)	Diesel	E5 + Diesel	Big Breaks
Outcome:	Mean Margin	Mean Margin	Mean Margin	Mean Margin	Mean Margin	Mean Margin
			Sample: Monopolist Stations			
Adopter × post-Adoption	0.003	0.016	-0.001	-0.002	-0.004	0.030
	(0.011)	(0.011)	(0.032)	(0.012)	(0.011)	(0.035)
Station FE	YES	YES	YES	YES	YES	
Year-Month FE	YES	YES	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES	YES	YES
N Brand Stations Control	YES	YES	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES	YES	YES
Observations	18,556	18,556	18,556	18,556	18,556	18,556
			Sample: Non-Monopolist Stations			
Adopter × post-Adoption	0.017***	0.033***	0.023***	0.015***	0.014**	0.056***
	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)	(0.013)
Station 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	429,181	429,181	429,181	429,181	429,181	429,181

Notes: Sample is gas station/month observations from January 2016 until December 2018 that have one competitor or more in their market. Outcome variable Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale prices. In Column (1) "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in two out of the measures in the main text, but excluding the speed of rival response. In Column (2) "Adopter" is a dummy equal to 1 in month t if the station experienced a structural break in two out of the measures in the main text, but excluding the responsiveness to crude oil price shocks. In Column (2) "Adopter" 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 within *2* weeks in any previous period. In Column (4) "Adopter" 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 for *Diesel* gasoline within *4* weeks in any previous period. In Column (5) "Adopter" 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 for both *E5* and *Diesel* gasoline within *4* weeks in any previous period. In Column (6), "Adopter" is a dummy equal to 1 in month t if the gas station experienced "large" structural breaks in at least 2 out of 4 relevant measures for E5. A "large" break is defined as a break that is at least within the 25th percentile of break sizes and where stations experience no reversals in their measures for at least two months. "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, the number of competitors in the market, and 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 are clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We also consider adoption definitions that involves a station experiencing multiple structural breaks in different fuel types. In the first version, in Column (4), a station is defined as an adopter if it experiences structural breaks in our four measures, but using Diesel prices. In Column (5), we consider an event stricter version of this adoption definition where a station has to experience structural breaks in at least two out of our three adoption measures in **both** E5 and Diesel within a period of 4 weeks. As market structure and demand for E5 and Diesel are fundamentally different, if a station experiences changes in pricing strategy in both fuel types at the same time, it is highly likely to be driven by the adoption of new pricing software. We take the adoption date to be the average between the adoption date of E5 and the Diesel adoption date. Columns (4) and (5) in Table F4 present results using these definitions of adoption. We find that the results are qualitatively and quantiatively the same as the baseline results.

We also consider a stricter adoption definition that only defines stations as adopters if they experienced "large" structural breaks in the four measures. Large breaks are defined as those within at least the 25th percentile of all break sizes, and where the measures never "reverse" for at least two months after the date of the best-candidate break.[72] Using this stricter definition we are left with fewer than 900 stations labelled as adopters. IV estimates using this definition are in Column (6). They show qualitatively similar results to our less strict measure. The effect of adoption on average daily monopolist margins is a statistical zero. The effect of adoption on average daily non-monopolist margins are positive and statistically significant. Effect size for non-monopolists is larger than using our baseline definition, suggesting that our baseline effects may be conservative.

[72]For example, after a best-candidate structural break in the number of price changes, a station might switch to changing their prices 10 times per day rather than 5 times per day. However, at some point after the date of the best-candidate break, the station might return to changing their prices 5 times per day for a month and then return to changing their prices 10 times per day. The stricter definition does not consider this station as an adopter.

F.4 Alternative Instruments

F.4.1 Alternative Brand Adoption IVs

In the main text our instrument for station i's adoption is the probability that other stations in its brand adopted. We do this by calculating the share of station i's brand adopters excluding station i (the focal station). In this section we propose a set of brand adoption instruments that exclude not only station i, but also other stations in the same brand that are geographically close to station i. First, we exclude other stations in station i's brand if they are also in the same cluster market as station i.[73] Next, we exclude other stations in station i's brand if they are also in the same NUTS3 region as station i.[74] This is a geographic region that is close to in size to a US county (if not larger). NUTS3 regions are larger (spatially) than cluster markets - we have approximately 4,000 cluster markets, but only 400 NUTS3 regions. Last, we exclude other stations in station i's brand if they are also in the same NUTS2 region as station i.[75] NUTS2 regions are larger than NUTS3 regions. There are only 38 NUTS2 regions in Germany. These alternative versions of our instrument help to test for potential spatial correlation in the errors that could undermine our Hausman-Nevo-style IV approach. That is, we ensure that our main instruments are not capturing common local demand shocks that are included in station i's time-varying unobservables.

Results from regressions with these instruments are in Table F5. They are remarkably consistent to the estimates in the main text. Even the estimate where we exclude other stations in the same NUTS2 geography are within 0.1 cpl (one standard deviation) of the estimate in the main text. It suggests that our main instruments are actually picking up brand-wide shock costs rather than some other correlations across stations.

F.4.2 Broadband Availability

We propose an alternative set of instruments that correct for endogeneity in station adoption decisions without relying on unobservable brand HQ decisions. The instruments capture the quality of broadband access in station i's region. There is well documented heterogeneity in broadband access and quality in Germany, with some areas and regions receiving sub-par services and speeds that are compared to the "old dial-up days" (NPR.org). In 2017, the second year of our sample, 29% of

[73]Concretely, the instrument is equal to $\frac{\text{N brand adopters outside station } i\text{'s cluster market}_{it}}{\text{N brand stations outside station } i\text{'s cluster market}_{it}}$.
[74]Concretely, the instrument is equal to $\frac{\text{N brand adopters outside station } i\text{'s NUTS3 region}_{it}}{\text{N brand stations outside station } i\text{'s NUTS3 region}_{it}}$.
[75]Concretely, the instrument is equal to $\frac{\text{N brand adopters outside station } i\text{'s NUTS2 region}_{it}}{\text{N brand stations outside station } i\text{'s NUTS2 region}_{it}}$.

Table F5: Station Level Results with Alternative Geographic Instruments

	(1)	(2)	(3)
Excluded IV Geography	Cluster Market	NUTS3 Region	NUTS2 Region
Outcome:	Mean Margin	Mean Margin	Mean Margin
Adopter× post-Adoption	0.012***	0.013***	0.011***
	(0.002)	(0.002)	(0.002)
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
Weather Controls	YES	YES	YES
Other Controls	YES	YES	YES
Observations	447,564	445,354	438,014

Notes: Sample includes gas station/month observations from January 2016 until December 2018. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale gasoline prices. "Adopter" 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 $\{1, ..., t-1\}$. The instrument in Column (1) is constructed as the share of stations belonging to station i's brand that are outside of its cluster market that are labelled as adopters at time t. The instrument in Column (2) is constructed as the share of stations belonging to station i's brand that are outside of its NUTS3 region that are labelled as adopters at time t. The instrument in Column (3) is constructed as the share of stations belonging to station i's brand that are outside of its NUTS2 region that are labelled as adopters at time t. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. We also control for the number of stations in the market and the number of stations in the market who are adopters. Standard errors are clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

German users reported internet speeds less than half of those promised by providers (dw.com). A fast and reliable internet connection is a key requirement for the effective use of algorithmic pricing software. Computation is done in "the cloud," so gas stations need fast internet connections to access necessary price information in a timely manner. They also need reliable internet connections to upload their own data and feed and update the software.

Based on data obtained from the EU Commission's netBravo initiative, we construct a measure of broadband performance in the local area around each gas station: whether the local area around the gas-station has widespread access to high speed internet in a particular year. We use an indicator variable for high speed internet availability, capturing whether a connection faster than 10 Mb/s is widely available in the local area.[76] The intuition behind these instruments is that a gas station should be more likely to adopt algorithmic pricing software once its local area has access to high speed internet. It should also be more likely to adopt algorithmic pricing software if internet signals in its local region are reliable. The availability of internet in the area should not be correlated with station specific unobservables conditional on all other local demographics (income, population

[76] We define speed X to be widely available in an area if average speed-tests in that area in that year exceed that speed. As well, we assume that if an area has speed X widely available in a year, it also has the same speed widely available in every subsequent year. More details on the construction of these variables are in the Data Appendix.

density, etc).

There are two downsides to this identification strategy relative to our main approach. First, variation at the region-year level is relatively limited as compared to variation at the brand-month level. Second, because an important source of the variation comes from regional geographic conditions, it is difficult to extend these instruments from station-level analysis to duopoly market level analysis. Duopoly and triopoly markets, by definition, consist of stations that are close together in geographic space. There are no stations that we consider to be in the same market but that have different broadband conditions.

Table F6 presents results from regression using these instruments. Qualitatively, the results are similar to those derived using our primary identification strategy. IV estimates show that the adoption of algorithmic pricing software increases mean station margins above wholesale prices. Mean station prices also go up. We once again find that adoption by monopolist stations has no effect on mean margins.

Table F6: Station Level Results with Alternative Instrument

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample:	All Stations	All Stations	All Stations	Non-Monopolists	Non-Monopolists	Monopolists	Monopolists
Outcome:	Adopter	Mean Margin	Mean Price	Mean Margin	Mean Price	Mean Price	Mean Margin
Adopter × post-Adoption		0.081***	0.070***	0.091***	0.073***	0.008	-0.028
		(0.022)	(0.021)	(0.025)	(0.022)	(0.020)	(0.024)
10 Mb/s Internet Available Dummy	0.013*						
	(0.008)						
Station FE	YES	YES	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES	YES	YES	YES
Weather Controls	YES	YES	YES	YES	YES	YES	YES
Other Controls	YES	YES	YES	YES	YES	YES	YES
Observations	443,752	443,752	443,752	425,793	425,793	17,503	17,503

Notes: Samples in Columns (1)-(3) include gas station/month observations from January 2016 until December 2018. Columns (6) and (7) only include stations that have no competitors in their market. Columns (4) and (5) include only stations that have one or more competitors in their market. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale gasoline prices. "Adopter" 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 $\{1, ..., t-1\}$. The excluded instrument used in the 2SLS regressions in Columns (2)-(7) are a dummy for whether 10 Mbps internet was available in that year in that region. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. We also control for the number of stations in the market and the number of stations in the market who are adopters. Standard errors are clustered at market level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Quantitatively, point estimates of the effects of adoption are substantially larger than our main estimates. Results from the first stage suggest why this is the case. The instrument shifts the adoption variable in expected directions: the availability of 10 Mb/s broadband increases adoption.

But compared to the brand-level instruments the instruments do not shift adoption probabilities by as much as the brand level instruments. As mentioned previously, there is also less variation in these instruments than in our brand-month instruments.

F.4.3 Placebo IV - Other Brands' Adoption

The main assumption of our baseline instruments is that brand level adoption recovers something about the incentives that the brand provides for their stations to adopt - for example, subsidies for replacing equipment or training. Effectively, we should be capturing brand-specific time varying cost shocks. To test whether this is the case, or whether we are capturing some other set of brand-specific time varying changes, we propose a "placebo" instrument.

This "placebo" instrument for station i in month t is the share of adopting stations at time t by *a different* brand than station i's brand.[77] This instrument has some similar time variation to our baseline instrument (i.e., brand adoption in general is going up over time) but the cost correlation should not exist. Results from this regression are in Table F7. They show that (i) there is no correlation between the propensity of other brands to adopt algorithmic pricing technology and the adoption of station i, and (ii) 2SLS regressions using this instrument do not generate any statistically significant effects of adoption on mean margins or prices.

F.5 Alternative Standard Error Clustering

Standard errors in the main text are clustered at the cluster-market level (see Section B). These potentially do not sufficiently account for the correlation in error terms across observations. We test for alternative clustering approaches in this section. Estimates are in Table F8. First, we use a two-way cluster, where we allow for error correlation within a market and within the particular month (Column 1). Then we expand the geographic range of the spatial correlation in error terms we allow. In Column (2) we allow for error terms to be correlated within the 400 NUTS3 regions in Germany. In Column (4) we allow for error terms to be correlated within the 38 NUTS2 regions in Germany. In Column (3), we allow for error terms to be correlated *both* within a NUTS3 region and within a particular month. In Column (5) we allow for error terms to be correlated *both* within a NUTS2 region and within a particular month. In all cases, the standard errors do not change substantially and our estimates at the station level are still statistically significant at the 95% confidence level.

[77] In practice, we pick a random station in station i's market and use their adoption shares.

Table F7: IV Station Level Results with "Placebo" Instrument

	(1)	(2)	(3)
Outcome:	Adopter	Mean Margin	Mean Price
Adopter × post-Adoption		0.117	-0.027
		(0.151)	(0.059)
Share Random Non-station i Brand Adopters	-0.026		
	(0.032)		
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
Weather Controls	YES	YES	YES
Other Controls	YES	YES	YES
Observations	407,964	407,964	407,964

Notes: Sample included gas station/month observations from January 2016 until December 2018 with at least one competitor in their market. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale gasoline prices. Mean Price is the average retail price for station i in month t. "Adopter" 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 Non-station i Brand Adopters" is the excluded instrument used in the 2SLS regression. It is equal to the share of stations that belong to a brand present in the same market as 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. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. We also control for the number of other stations in the market and the number of stations in the market who are adopters at month t. Standard errors are clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table F8: IV Station Level Results with Alternative Standard Error Clusterings

	(1)	(2)	(3)	(4)	(5)
Outcome:	Mean Margin	Mean Margin	Mean Margin	Mean Margin	Mean Margin
Std. Err. Clustering	Market × Period	NUTS3 Region	NUTS3 × Period	NUTS2 Region	NUTS2 × Period
Adopter× post-Adoption	0.012**	0.012***	0.012**	0.012***	0.012**
	(0.005)	(0.003)	(0.006)	(0.004)	(0.006)
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	447,564	447,564	447,564	447,564	447,564

Notes: Sample included gas station/month observations from January 2016 until December 2018 with at least one competitor in their market. Mean Margin is the monthly average of daily differences of pump price for station i in month t and wholesale gasoline prices. Mean Price is the average retail price for station i in month t. "Adopter" 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 $\{1, ..., t-1\}$. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Weather controls include the mean and standard deviation of monthly temperature and precipitation near station i in month t. We also control for the number of other stations in the market and the number of stations in the market who are adopters at month t. In Column (1), standard errors are clustered at market and year-month level. In Column (2), standard errors are clustered at the NUTS3 level. In Column (3) standard errors are clustered at the NUTS3 and year-month level. In Column (4) standard errors are clustered at the NUTS2 level. In Column (5), standard errors are clustered at the NUTS2 and year-month level. *** p<0.01, ** p<0.05, * p<0.1