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

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

July 2020

Abstract

Economic theory provides ambiguous and conflicting predictions about the association between algorithmic pricing and competition. In this paper we provide the first empirical analysis of this relationship. We study Germany's retail gasoline market where algorithmic-pricing software became widely available by mid-2017, and for which we have access to comprehensive, high-frequency price data. Since dates of adoption are unknown, we identify stations that adopt algorithmic-pricing software by testing for best-candidate structural breaks in markers associated with algorithmic pricing. We find a large number of station-level structural breaks around the suspected time of large-scale adoption. Using this information we then investigate the impact of adoption on outcomes linked to competition. Since station-level adoption is endogenous, we use brand headquarter-level adoption decisions as instruments. Our IV results show that adoption increases margins by 12%, but only in non-monopoly markets. Furthermore, restricting attention to duopoly markets, we find that market-level margins do not change when only one of the two stations adopts, but increase by nearly 30% in markets where both do. These results suggest that AI adoption has a significant effect on competition.

*Preliminary and incomplete. The views in this paper do not reflect those of the Bank of Canada. Helpful comments were provided by Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, Joe Harrington, Alex MacKay, Jeanine Miklós-Thal and seminar participants at the University of Bologna. Daniel Ershov would like to acknowledge support received from ANR under grant ANR-17-EUR-0010 and from ANITI. Correspondence to: [a]Stephanie Assad - Queen's University, Dunning Hall, 94 University Avenue, Kingston, 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, 1 Esplanade de l'Université, 31080 Toulouse, France, Email: daniel.ershov@tse-fr.eu., [d]Lei Xu - Bank of Canada, 234 Wellington St Ottawa, 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 vast quantities of data and interact, learn, and make decisions with unprecedented speed and sophistication. The evolution of algorithmic-pricing software has raised concerns regarding the potential impact on firm behaviour and competition. In particular, the potential for the use of algorithms as a means 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). Since the goal of reinforcement learning-based algorithms is to converge to an optimal policy, the concern arises that AI agents will learn to play a collusive strategy to achieve a joint-profit maximizing outcome. Whether these strategies are learned or programmed-in explicitly by users, the employment of algorithmic software can facilitate collusion through increased ease of monitoring and speed of detection and punishment of possible deviations.

The literature on algorithmic collusion has been expanding, with contributions from the fields of economics, law, and computer science. Despite this growing attention, no consensus has yet been reached as to whether, from a theoretical perspective, algorithms facilitate tacit collusion. There are also questions about whether algorithmic collusion can arise in practice since 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 the 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 mass adoption allegedly occurred beginning in 2017. We use a high-frequency database on German retail gasoline prices, containing information on prices and characteristics for every retail gas station in the country.[1]

We investigate the impact of algorithmic-pricing software adoption on competition by comparing the retail margins of adopting and non-adopting stations. However, there are many channels, other than through competition, that adoption of algorithmic-pricing software can increase margins: an algorithm can better detect underlying fluctuations in wholesale prices or better predict demand. 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.

isolate the effects of algorithmic-pricing software adoption on competition we focus on the role of market structure, comparing adoption effects in monopoly (one station) markets and non-monopoly markets.[2] We also perform a more direct test of theoretical predictions by focusing on isolated duopoly (two station) markets. We compare market-level average margins in markets where no stations adopted, markets where one station adopted and markets where both stations adopted. In the first market type, competition is between human price setters. In the second it is between a human price setter and an algorithm, while in the last it is between two algorithms. As a consequence we are able to identify a pure effect of algorithmic pricing on competition.

An obstacle we face is that the decision to adopt algorithmic-pricing software is not observed. To overcome this challenge we test for structural breaks in pricing behaviours that are thought to be related to the use of sophisticated pricing software. Since we do not have an exact date at which we expect adoption to have occurred, we use a Quandt-Likelihood Ratio (QLR) test (Quandt 1960) to test for the best candidate break date when the break date is unknown. We test for structural breaks at each station for each time period in a large window around the time of supposed adoption. The best candidate structural break for a given station is identified as the time period with the highest resulting F-statistic of each of these tests. Using this method, we test for structural breaks for each station in the following measures: (i) the number of price changes made in a day, (ii) the average size of price changes, and (iii) the response time of a station's price update given a rival's price change. We focus on these measures since they capture the promised impacts of algorithmic software in the retail gasoline market.[3] As described on a software provider's website, their pricing software collects data and performs high frequency analysis to "rapidly, continuously and intelligently" react to market conditions. To minimize false positives, we classify a station as an algorithmic-pricing adopter if it experiences a structural break in at least two out of three measures within a short time period (which we take to be eight weeks, but is robust to alternative specifications). We find that approximately 30% of stations in our data set structurally break in multiple measures of pricing behaviour within an eight week window, with the majority of these breaks occurring in mid-2017, just after AI pricing software becomes widely available.

Having identified adopters, we compare outcomes linked to competition between adopters and non-adopters. We focus on the impact of adoption on mean monthly margins (above crude oil prices).

[2] We use two geographic definitions for markets: drawing 1KM radii around stations, or using 5-digit ZIP codes.

[3] Using rival responsiveness as a marker may over-state AI adoption if non-adopters who react to their AI-using competitors' frequent price changes are automatically labelled as adopters. This does not appear to be the case in our data. When examining duopoly (two station) markets, we find asymmetric adoption (where one duopolist adopts and the other does not) in a large share of markets. More details are in Section 6.1.

In the gasoline retail market, mean margins are a clear indictor of profitability and market power: the ability of stations to mark-up retail prices above 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). Theory papers on algorithmic competition also make clear predictions related to margins (Calvano et al 2019, Brown and MacKay 2020). In addition to mean margins, we investigate the effects of adoption on minimum and maximum monthly margins to learn about the impact of adoption along the margin distribution.[4]

Although we control for time and station-specific effects, as well as time-varying market level demographics, adoption is endogenous. Individual station adoption decisions are likely correlated with station/time specific unobservables (managerial skills, changing local market conditions, etc). As a result, OLS estimates are attenuated. We instrument a station's adoption decision using the adoption decision by the station's *brand* (i.e., by brand headquarters). As demonstrated by previous technology-adoption episodes, brands can facilitate adoption by their stations. "Adopting" brands provide support/subsidies/training to individual stations, reducing adoption costs.[5] 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 AI pricing. If only a very small fraction of a brand's stations adopts AI, it is unlikely that the brand itself decided to adopt. If a large fraction adopts, it is likely that the brand itself adopted and facilitated adoption by the stations.

Using this instrument for adoption we find that mean margins for adopting stations increase by 1.5 cents per litre, or roughly 12%, after adoption. We observe heterogeneity in outcomes based on market structure. Stations that have no competitors in their ZIP code (monopolists) see no statistically significant change in their mean margins. Adopting stations with competitors in their ZIP code see a mean margin increase of 1.6 cents per litre and the entire distribution of their margins shifts right.[6]

To further investigate whether algorithmic pricing affects competition and test theoretical predictions, we focus our attention on market-level adoption in duopoly (two station) markets. IV estimates suggest that relative to markets where no stations adopt, markets where both stations adopt see a

[4] We take the monthly average, minimum and maximum of daily margins above crude oil, also subtracting German gasoline taxes and VAT.

[5] This does not necessarily fully covers station adoption costs, so brand adoption does not imply that all its stations adopt immediately. See Appendix B for an example from the adoption of electronic payments by gas stations in the 1990s.

[6] Monopolist stations that adopt increase their minimum margins and reduce their maximum margins.

mean margin increase of 3.4 cents per litre, or roughly 30%. Markets where only one of the two stations adopts see no change in mean margins. These results show that market-wide algorithmic pricing adoption raises margins (*ceteris paribus*), suggesting that algorithms reduce 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).

Finally, we explore the mechanism underlying algorithmic pricing and competition: whether algorithms are unable to learn how to compete effectively, or whether algorithms actively learn how not to compete (i.e., tacitly collude). To do so, we test whether margin changes happen immediately after adoption or arise gradually. If the former effect dominates, we should see immediate increases in margins. If the latter effect dominates, it should take a longer time for the algorithm to train and converge to tacitly-collusive strategies (Calvano et al 2019). We find evidence that margins do not start to increase until about year after adoption, suggesting that algorithms in this market learn tacitly-collusive strategies. These results are in line with the findings in Calvano et al (2019).

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 and 5 discuss the data and methodology we use in our analysis respectively. Sections 6 and 7 discuss, respectively, our results regarding (i) identifying adoption and (ii) the impacts of AI adoption on station and market outcomes. We also conduct a number of robustness checks. In Section 8 we provide evidence to support the idea that outcome results are driven by algorithms learning to tacitly collude. In Section 9, we discuss algorithmic pricing as a means to facilitate collusion, antitrust concerns over algorithmic collusion, and how these relate to German and European competition law. We also present policy recommendations drawing on our results. We conclude in Section 10.

2 Related Literature

This paper most closely relates to the recent literature concerning the potential link between algorithmic pricing and collusion. Theoretical 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 2019; Klein 2019); 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 2019).

Klein (2019) and Calvano et al. (2019) 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. 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 (2019) come to similar conclusions. Brown and MacKay (2020) 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.[7]

The question as to whether algorithm usage may result in coordinated behaviour is of widespread interest and thus 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. A number of these papers, including Kaymak and Waltman (2006, 2008) and Moriyama (2007, 2008) have indicated 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 algorithmic usage can lead to collusive behaviour. Authors including 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 have been a small number of papers looking specifically at the German market. See in particular Dewenter and Schwalbe (2016), Boehnke (2017), and Cabral et al (2018). There is also 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.

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

[7] Decarolis and Rovigatti (2019) find that common biddings 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 interesting parallels with the algorithmic pricing literature and regulatory concerns about multiple competitors in a market adopting the same pricing algorithm. Unlike this paper, the primary focus of Decarolis and Rovigatti (2019) is on increasing intermediary concentration rather than on algorithmic pricing software behaviour and the mechanism through which bidding decisions are made.

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 faciliate 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) on the other hand find that firm profit, as well as consumer surplus, may increase or decrease under personalized pricing depending on consumers certainty regarding their product tastes. However, they also find that in every case, total welfare is higher under discriminatory pricing in comparison to uniform pricing.

3 Background

3.1 Algorithmic Pricing in the German Retail Gas Industry

The December 2017 issue of *Tankstop*, a trade publication for Germany's retail gasoline sector, notes that *a2i* systems, a Danish artificial intelligence company, had begun to offer services to petrol station operators within Germany (see Figure A1). *a2i* systems creates industry specific technology for retail gas markets, including an AI pricing technology software, *Pricecast Fuel*. The promise of Pricecast Fuel is to use BDI (belief-desire-intention) and Neural Network based algorithms to determine optimal pricing strategies for retail gas stations (Derakhshan et al 2016). In Figure A2 we reprint an intuitive explanation from *The Wall Street Journal* of how such pricing algorithms work. For a given station, the algorithm takes in historical data on transactions, competitors' information (i.e., competitor's prices) and other market conditions (i.e., weather) as inputs. Outputs are prices that maximize station profits.[8] Initial training for the algorithm is done on historical data, but the algorithm can also take in additional "real-time" information such as current weather and traffic patterns. The algorithm uses these inputs and sets prices. Transactions resulting from these prices are fed back into the system and are used to re-optimize the algorithm.

a2i began offering Pricecast Fuel to Danish retail fuel companies in 2011, noting on their website that they have established long-term partnerships with other European retail fuel stations since 2012 (*a2i* Systems 2020). *Tankstop*'s December 2017 issue further notes that *a2i*'s services are supported

[8]Individual station owners can set other goals, such as market share maintenance, or constraints such as minimum price.

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).

a2i's Pricecast Fuel software has many of the characteristics that legal experts and antitrust authorities express concern over in regards to the possibility of algorithmic collusion. For one, the software involves large-scale and high-frequency collection of, interaction with, and response to competitor data (*a2i* systems (2020)). This high-frequency interaction between competitors can make collusion easier to sustain due to increased ease of monitoring and quicker detection and punishment of deviations (Ezrachi and Stucke 2015; Mehra 2015). Additionally, although we do not know the exact specifications of the algorithms implemented by Pricecast Fuel, *a2i*'s website states it is based on "learning algorithms". Such learning algorithms have been the focus of many antitrust discussions and economic theory papers regarding algorithmic pricing and collusion (Calvano et al 2019), due to the potential for such algorithms to learn and converge to a collusive strategy. Finally, *a2i* is advertised to offer their software to multiple stations and brands within the German retail gas market (Tankstop 2017). The adoption of the same software by numerous brands may lead to a concern for a hub-and-spoke scenario or the parallel use of individual algorithms scenario, depending on how individualized *a2i*'s algorithms are for each customer.[9]

It is important to note that we focus on discussing *a2i* since they have been public about the contents of their algorithm and since we have direct anecdotal evidence of their availability and adoption in the German context. There are, however, several similar products available in the market. For example, *Kalibrate*, a UK company, also offers an AI powered pricing software for gasoline retail (Kalibrate.com). Their list of clients include Nordic gasoline retail chains ST1 and Preem, as well as the Polish chain Orlen which is also active in Germany (Kalibrate.com). Based on a limited description of their algorithm on the provider's websites, competing algorithms work similarly to *a2i*'s algorithm. Our discussion above is representative of the general type of algorithms that are available in the retail gasoline market.

[9]Decarolis and Rovigatti (2019) find that the adoption of common bidding intermediaries by multiple competitors (who then presumably use the same bidding software) generates anti-competitive effects in online advertising markets.

4 Data

The data we use in our analysis are taken from the German Market Transparency Unit for Fuels. The data set contains price data for the most commonly used fuel types, Super E5, Super E10, and Diesel and for every German petrol station in 5 minute intervals from 2014 to the end of 2019.[10] We also have characteristic information on each station, including their exact address (street address, 5 digit ZIP code), latitude and longitude coordinates, station name, and associated brand. In total, there is information on 16,661 stations; however, 634 of these stations are missing location data.

We combine this price data with additional data sources. To compute station-level margins, we take the average daily Brent Crude oil price per barrel, and convert it into price per litre (to match up with station prices). We also subtract German gasoline retail taxes (0.65 cents per litre) and VAT (0.247 cents per litre). We compute station-level daily margins and take the average, minimum or maximum values of these daily margins for our main analysis (which is done at the station-month level).

We also merge in annual regional demographics from Eurostat from 2016 to 2018. We include data on total population, population density, median age, employment (as a share of total population) and regional GDP. This data is 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 county, but is larger than a 5-digit ZIP code.

4.1 Station-Level Descriptive Statistics

Gas stations in our data set belong to 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 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 2,058 brands in our data. 87 per cent of these brands are single-operating stations, 239 brands have between 2 and 100 stations and 19 brands have more than 100 stations. Although single-station brands are a majority, they only reflect a small share of stations. Out of the 16,661 stations in our data set, single-operating stations account for approximately 11 per cent of all stations

[10]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. E10 and E5 are similar to 95 and 98 octane rating fuels commonly used in North America.

and about 18 per cent of stations belong to brands of size 2 to 100 stations (3,109 stations belong to brands of this size). The top 5 brands account for 43 percent of stations and the 19 largest brands (those with more than 100 stations) account for 71 per cent of total stations (11,752 stations total). ARAL and Shell are the dominant brands in the German retail gas market, with the largest number of stations, 2,417 and 1,852 respectively, together making up over 25 per cent of stations in the German retail gas market. There are a number of other large brands with over 350 stations each: Esso, Total, Avia, Jet, Star, BFT, Agip, Raiffeisen, and Hem.[11] In terms of market shares, ARAL, Shell, Jet, BFT, Total and Esso together account for 84 per cent of fuel sales in the German retail gas market.[12]

There are 5,781 5-digit ZIP codes in our data. 2,094 ZIPs have a single station (are monopoly markets). 1,307 ZIPs have two stations (are duopolies). The mean number of stations per ZIP code is around 3 and the median is 2. Only 81 ZIP codes have more than 10 stations.[13] The majority of stations are within 5KM of their closest competitors (about 94 per cent) and the average distance of a station to its closest competitor is 1.4KM.

Table 1: Station Summary Statistics

Variable	**Observations**	**Mean**	**Std. Dev.**	**Min**	**25%**	**75%**	**Max**
Stations per Brand	2,058	8.10	82.40	1	1	1	2417
Stations per 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
# of other stations within 1KM	16,027	1.09	1.34	0	0	2	17
# of other stations within 3KM	16,027	5.14	5.07	0	2	8	36
# of other stations within 5KM	16,027	10.91	10.96	0	3	17	64

4.2 Station/Month-Level Descriptive Statistics

The average price that a station charges per litre of E10 fuel is 1.34 Euros, but the mean monthly margin that the average station earns over the price of crude oil (after taking away gasoline taxes and VAT) is 13 cents per litre. The maximum daily margin a station earns on average in a month is

[11]These brands operate 1,091, 933, 804, 689, 563, 430, 450, 440, and 387 stations respectively

[12]As of 2019, taken from https://www.bft.de/daten-und-fakten/kraftstoffmarktanteile. Fuel market sales for each brand are 21 per cent for ARAL, 20 per cent for Shell, 16 per cent for BFT, 10.5 per cent for Jet, 9.5 per cent for Total, and 7 per cent for Esso.

[13]ZIP codes reflect population patterns, so urban ZIP codes are much smaller in terms of area than rural ZIP codes.

16 cents, and the minimum daily margin a station earns on average in a month is 8 cents.

The average station is located in a fairly dense NUTS3 region, with population density of 738 persons per square-km. The median age of the population around a station is 46 years and 53 percent of the population is employed.

Table 2: Station/Month Summary statistics

Variable	Observations	Mean	Std. Dev.
Mean Monthly E10 Price (EUR/litre)	485,328	1.341	.089
Mean Monthly E10 Margin (EUR/litre)	485,146	.128	.05
Min. Monthly E10 Margin (EUR/litre)	485,146	.077	.13
Max. Monthly E10 Margin (EUR/litre)	485,146	.162	.054
ln(Total Regional Population)	482,884	12.406	.807
Regional Population Density (pop/km^2)	481,699	738	1,000
Regional Median Population Age	482,884	46.014	3.073
ln(Regional GDP)	480,517	9.072	.961
Regional Employed Share (employed/pop)	480,517	.526	.133

5 Methodology

Our empirical analysis to study the effect of algorithmic pricing on competition in the German retail gas market proceeds as follows. First, we identify AI-adopting stations by performing Quant Likelihood Ratio (QLR) tests to identify structural breaks in the time series of pricing strategies of gas stations in our sample period. Second, we investigate the impact of AI adoption on market outcomes related to competition such as margins over the price of crude oil. We compare outcomes for adopting stations vs. non-adopting stations, before and after adoption takes place. Selection bias coming from the differences between adopters and non-adopters, as well as endogeneity due to the timing of adoption would attenuate OLS estimates down from true effect of adoption. We use a brand-level instrument to deal with these endogeneity concerns. We allow for heterogeneity of effects across different market structures (monopoly stations vs. non-monopolists). We also focus on duopoly markets (geographic markets with two stations) and compare market-level outcomes between markets where no stations adopted, markets where one station adopted and markets where both stations adopted.

5.1 Identifying Adoption Decisions

We assume that there are two levels of decision-making when it comes to adoption of AI technology: at the brand HQ (headquarters) level and at the individual station level. We address each in turn. There is substantial evidence that brands have entered into long-term strategic partnerships with AI pricing and analytics providers, either directly or indirectly. In Denmark, *a2i* directly entered into a partnership with the "largest Danish retail fuel company OK Benzin" (a2isystems.com). On its website, *a2i* also claims that by 2012 it had "established long-term partnerships with a number of European Retail Fuel companies" (a2isystems.com). Preem, a Swedish gasoline retailer, entered into a similar strategic partnership with Kalibrate, a provider of AI-based pricing software along the lines of *a2i* (Kalibrate.com).

More indirectly, AI-pricing software providers enter into partnerships with IT companies that provide integrated services to brands. *a2i* 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). Wincor Nixdorf then enters into partnerships with brands. Also, as mentioned in Section 3, in Germany *a2i*'s software was at least partially implemented through WEAT - a technology company working with brands and providing technical and logistical support (WEAT.de).

However, as in other cases of corporate technology adoption (i.e., Tucker 2008), if a brand decides to "adopt," or enter into a partnership with an AI pricing software provider, this does not necessarily mean that all of its stations automatically and instantaneously adopt. Rather it implies simply that the technology exists for mass-adoption. There are many reasons why not every station in a brand will adopt right away. One possibility is that the brand does not want to immediately change its pricing strategy across all of its stations, which sometimes can number in the hundreds or thousands. Brands may wish to perform some experimental tests to make sure that the pricing software works or delivers the desired outcomes, or release the software gradually. This has been the case with the very public adoption of *a2i*'s software by Denmark's OK Benzin (a2isystems.com).

Even if a brand wanted all of its stations to adopt immediately, not all of them could or would. Cloud-based AI-pricing software potentially requires substantial infrastructure investments. For example, high-speed internet *and* high-speed internet enabled Point of Sale (PoS) terminals and pumps are likely required for the software to work.[14] Similarly, station operators also may require substan-

[14]Other upgrading decisions by gas station owners include allowing for chip cards or automated payment at the pump (Chicago Tribune).

tial training with the software to set its parameters and deal with potential errors. Brands may under-write or subsidise investments and training, but ultimately such costs fall on individual station owners. Not all station owners are in a position to incur these costs right when the technology becomes available, or possibly ever. This was the case for previous waves of gasoline retail technology adoption. For example, in the 1990s, Exxon Mobil (Esso's parent company) had a brand-wide roll-out for the Mobil Speedpass, a contactless electronic payment system. BusinessWeek reported that to actually adopt the technology individual station owners "have to install new pumps costing up to $17,000–minus a $1,000 rebate from Mobil for each pump" (BusinessWeek). In other words, individual station owners have discretion as to whether (and when) to invest in adopting AI software.[15]

It is important to note that in our dataset we do not have information on either brand-level or station-level adoption. Our approach is to take advantage of the high-frequency price data to identify these decisions.

5.1.1 Identifying Station-Level Adoption

We currently focus on three measures of pricing behaviour (aggregated to a weekly level) to identify the adoption of algorithmic pricing at the station level: (i) number of price changes, (ii) average size of price changes, and (iii) rival response time. We focus on these measures as a means to capture the promised impacts of *a2i*'s pricing software, as well as to be consistent with the theoretical literature. As described on *a2i*'s website, their pricing software, PriceCast Fuel, uses the collection and high-frequency analysis of large quantities of data on consumers, competitors, and market dynamics to learn how previous pricing strategies have impacted consumer and competitor behaviour, and in turn station performance. The software is advertised to use these learned strategies to "rapidly, continuously, and intelligently react" to market conditions; automatically setting optimal prices in reaction to changes in demand or competitor behaviour or to maximize margins without eliciting a change in behaviour by consumers or competitors. Therefore we may expect that after AI-adoption, stations may make more frequent and smaller updates of their prices, due to quicker and more precise detection of demand fluctuations or changes in competitor behaviour, as well as detection of small

[15] We provide supporting evidence for staggered technology adoption in the gasoline market in Appendix B. We look at the adoption of electronic payments from 1991 to 2001 by Canadian gasoline retail stations. This is another technology that clearly benefits brands and that brands would want stations to adopt. However, some stations may not want to adopt. Other stations are incapable of adopting since adoption of this technology requires a stable phone connection. We show that it took years after the first appearance of this technology for a substantial fraction of a brand's stations to adopt.

changes that could be made without impacting consumer or competitor behaviour. Along these same lines, with quicker detection of, and response to, competitor behaviour, we would expect to see stations reacting more quickly to changes in competitors' prices.

These measures of pricing behaviour line up with what is described in the economic and legal literature discussing algorithmic adoption. Ezrachi & Stucke (2015) point out the ability for algorithmic software to increase the capacity of monitoring consumer activities and the speed of reaction to market fluctuations. Mehra (2015) points out the ability of AI pricing agents to more accurately detect changes in competitor behaviour and more quickly update prices accordingly. Brown and MacKay (2020) note that two significant features of pricing algorithms are their ability to *(i)* lower the cost of more frequent price updates and *(ii)* react quickly to price changes of other firms in the market. Our measure of rival response time follows a similar intuition to to the approach taken by Chen et al. (2016) who identify algorithmic pricing users in Amazon Marketplace by measuring the correlation of user pricing with certain target prices, such as the lowest price of that given product in the Marketplace.[16]

Structural breaks in pricing strategies are detected using Quandt-Likelihood Ratio (QLR) tests. This test, developed by Quandt (1960), with distributional properties established by Andrews (1993), tests for a structural break in a time-series measure for each period in some interval of time and takes the largest resulting test statistic. This test 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)). We conduct a QLR test for each station in our data set and for each variable of interest (number of price changes, average size of price changes, and rival response time) we run the following regression over a range of eligible break periods $\tau_0 \leq \tau \leq \tau_1$ (eligible break periods are measured by week):

$$y_{it} = \alpha_i + \beta_i D_t(\tau) + X_t \gamma_i + \epsilon_{it}, \tag{1}$$

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

[16]There are potential concerns that by looking at rival responses we may over-state adoption in the market if non-adopting firms will be automatically labelled as adopters if they react to the more frequent price changes of adopters. This does not appear to be the case in our data. When examining duopoly markets, we find asymmetric adoption in many markets where one duopolist adopts and the other does not. More details are in Section 6.1.

QLR statistic is the largest of these F-statistics over the range of eligible break dates:

$$QLR_i = max[F_i(\tau_0), F_i(\tau_0 + 1), ..., F_i(\tau_1)]. \tag{2}$$

The best candidate structural break period[17] for station i is identified as the week τ^* that satisfies $QLR_i = F_i(\tau^*)$. Structural breaks are identified as significant if they exceed a certain critical value.[18] For this analysis, we drop all stations from our data 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 also drop all dates before 2016 and after 2018, leaving us with a 3 year sample period.[19] We use 30% trimming for our test period, which is standard for QLR testing.[20]

5.1.2 Identifying Brand-Level Adoption

We do not observe an indicator for whether a brand HQ decided to enter into a strategic partnership with an AI-pricing software provider. However, using the methodology explained in the previous section, we identify station-level adoption decisions. Using station level adoption, there are several possible ways to proxy a brand's adoption. One possibility is that stations cannot adopt without their brands adopting. We could label a brand as an adopter as soon as any one of its stations is labelled as an adopter (has identified structural breaks in two out of three possible measures). The problem with this approach is that it is likely not *necessary* for the brand to make an adoption decision for a station to adopt the software. There are many providers of algorithmic pricing software that offer their products to small or medium enterprises.[21] *a2i*'s 2017 advertisements are clearly targeted at individual station owners and emphasize that all stations can adopt their technology. It is also possible that a brand would run an experimental trial on a station without fully adopting the software.

[17]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.

[18]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 $QLR_i \geq 12.16$.

[19]We drop 2015 to avoid confounding effects from Shell's price matching program that was introduced in Germany in that year. Dewenter and Schwalbe (2016) and Cabral et al (2018) have shown that this price matching policy had an effect on pricing strategies of stations in the German market, which we do not want to capture in our analysis.

[20]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.

[21]See among many others, Prisync.com or Comptera.net.

This measure would also be sensitive to outliers and would amplify any noise from our station-level adoption measure. For these reasons, we choose not to use this measure for brand level adoption.

The opposite approach would be to classify a brand as an adopter only if *all* of its stations are labelled as adopters. This would reduce any false-positive identification of adopting brands. However, it would not be consistent with the history of technology adoption in gasoline retail markets. There are substantial underlying differences between gas stations in their infrastructure, ownership and their contractual relationship with brands (i.e., whether they are directly owned by the brand or under a licensing agreement for the brand name). Since the adoption of any new technology is costly, it is very unlikely that every single station belonging to a sizeable brand would adopt simultaneously. Historical data on past gasoline retail technology adoption supports this statement. In Appendix B we provide data on the adoption of electronic payment systems in Canadian gasoline retail stations in the 1990s. We show that up to 10 years after the first appearance of this technology, less than 50% of stations belonging to leading brands adopted.

A third and intermediate approach is probabilistic. We could compute the probability that a brand adopted by using our classification of station-level adoption decisions. This approach would still capture the effects of the underlying brand-level decisions. As mentioned in Section 5.1, brand-level decisions should facilitate the adoption by individual stations. According to this approach, 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. This is the approach we take in this paper.

5.2 Evaluating the Impact of Adoption

5.2.1 Evaluating the Impact of Adoption on Station Outcomes

We want to capture the effects of station i adopting algorithmic pricing on the distribution of daily margins (above crude oil) in month t. We use three points in the margin distribution: (i) mean daily margin in month t, (ii) minimum daily margin in month t and (iii) maximum daily margin in month t. Our OLS specification is as follows:

$$y_{it} = \alpha_i + \alpha_t + \beta D_{it} + \gamma X_{it} + \epsilon_{it}, \tag{3}$$

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, D_{it} is a dummy variable equal to 1 if station i has adopted algorithmic pricing in time

t and 0 otherwise. X_{it} are time-varying station specific controls. Most importantly, X_{it} includes the number of other gas stations that are in the same postal code as station i. The key coefficient in this regression is β which captures the effect of AI adoption on y_{it}.

The OLS specification assumes that adoption is exogenous and as-good-as-random (conditional on observables). This is not necessarily the case. Algorithmic adoption could very much be endogenous and correlated with unobservable time-varying station characteristics. Stations with "high" unobservables (for example, better managed stations) could be more likely to adopt algorithmic pricing software and use it effectively. These stations could also have very different market outcomes. This would attenuate the adoption effect towards zero. Stations could also choose to adopt in response to unobservable station-specific shocks - once again, these would also affect both the adoption decision (D_{it}) and outcomes (y_{it}).

To address this issue we use an instrumental variables approach and instrument for D_{it}. We need an instrument that should be correlated with an individual station's adoption decision but is not affected by station-specific unobservable shocks. We propose *brand-HQ level adoption* as an instrument. As explained in the previous section, we can measure brand-level adoption by computing the share of stations belonging to each brand that have been identified as AI adopters by month t. For station i at time t our IV is therefore the share of stations in brand i's brand that adopted AI by time t. We exclude station i from this share. As discussed in Section 5.1, it seems reasonable that brand level decisions influence the adoption decisions of individual stations. 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. Such support is important for drastic technical changes such as AI adoption. At the same time, brand level decisions should not be influenced by station-level specific conditions.[22]

Finally, to test whether any observed change in prices and/or margins stems from a reduction in competition and increased market power as opposed to simply a better understanding of underlying fluctuations in crude/wholesale prices or better identifying consumers' price elasticity of demand, we consider differential responses based on market structure heterogeneity. We look separately at stations that are monopolists in their ZIP code, and stations that are not monopolists.

[22] We provide evidence for this in Table A5, showing that conditional on brand size, brand adoption shares are uncorrelated with market characteristics. See additional discussion in Section 6.5.

5.2.2 Evaluating the Impact of Adoption on Market-Level Outcomes

In a more direct test of theoretical predictions about the effects of AI on competition, we compare outcomes between adopting and non-adopting *markets*. We focus on duopoly station markets since most theoretical analysis is done for two firms (i.e., Calvano et al 2019, Miklos-Thal and Tucker 2019). We define duopoly markets in one of two ways. First, we draw a 1km radius circle around each station in our data. Stations are in the same market if their circles overlap. A duopoly market has only two stations within 1km of one another. Second, we use German ZIP codes and define duopoly markets as ZIP codes where there are only two stations. These two definitions do not necessarily capture the same things. There may be two stations in different ZIP codes within 1km of one another. There may also be two stations that are located more than 1km apart but are in the same ZIP code.

For market m in month t, we use the following OLS specification:

$$y_{mt} = \alpha_m + \alpha_t + \beta_1 T^1_{mt} + \beta_2 T^2_{mt} + \epsilon_{mt}, \tag{4}$$

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. T^1_{mt} is a variable equal to 1 if only one of the two stations in market m is labelled as an adopter.[23] T^2_{mt} is a variable equal to 1 if both stations in market m are labelled as adopters.[24] The two key coefficients in this regression are β_1 and β_2. β_1 captures the effects of AI adoption by *one of the two firms* in a duopoly market and β_2 captures the effects of AI adoption by *both firms* in a duopoly market.

As in the station-level regression, endogenous AI adoption by stations in response to market/time varying unobservables is a concern. Following the logic of our station-level instruments, we construct market-level IV using brand-level adoption decisions. The instruments for T^1_{mt} and T^2_{mt} are functions of the brand-level adoption decisions for the brands in market m:

$$\begin{aligned} IV^1_{mt} &= B_{1mt}(1 - B_{2mt}) + B_{2t}(1 - B_{1mt}) \\ IV^2_{mt} &= B_{1mt}B_{2mt}, \end{aligned} \tag{5}$$

where where B_{1mt} is share of other stations belonging to market m first station's brand that have been identified as AI adopters in month t. B_{2mt} is share of other stations belonging to market m

[23] $T^1_{mt} = D_{1mt}(1 - D_{2mt}) + D_{2mt}(1 - D_{1mt})$, where 1 and 2 are the stations in market m.
[24] $T^2_{mt} = D_{1mt}D_{2mt}$, where 1 and 2 are the stations in market m.

second station's brand that have been identified as AI adopters in month t.

6 Results – AI Adoption

In this section we present results regarding the identification of adopters. In the first subsection we discuss station-level adoption before then describing brand-level adoption in the second subsection. With these results in hand, in the Section 7 we will study the effect of adoption on outcomes related to competition.

6.1 Station-level adoption

As outlined in Section 5, we calculate QLR statistics for each station and each of the three adoption markers: (i) *number of price changes*, (ii) *average size of price changes*, and (iii) *rival response time*. We find that a large number of stations experienced structural breaks in each of the three markers. Out of 13,426 candidate stations, 12,402 experience a significant structural break in the number of price changes at the 5% confidence level.[25] Out of the stations that experience significant breaks, over 50% of the breaks occur in the spring and summer of 2017 (see Appendix A and Figure A3 for the full distribution and additional discussion).[26] We test 5,908 stations for structural breaks in *rival response time*, excluding all stations without rivals within 1km as well as stations that do not operate in each week of 2017. Out of the tested stations, 5,449 experience significant structural breaks. Out of stations with significant breaks (at at least the 5% level), over 25% break in the summer of 2017 (see Appendix A and Figure A4 for the full distribution and additional discussion). For the *average size of price changes*, we find that out of over 13,000 candidate stations, 11,264 experience a statistically significant structural break. Once again, over 25% of the best candidate breaks occur in the summer of 2017 (see Appendix A and Figure A5 for the full distribution and additional discussion).

We also find that that the changes in pricing strategies captured by these breaks are substantial. For the number of *price changes*, on average, a station that experienced a structural break changes their prices 6 times a day before the break and 10 after the break. *Rival response time* decreases

[25] The QLR statistic for these stations are greater than or equal to the 5% critical value 8.68

[26] We take the structural break as the date with the highest associated F-statistic. There may be some concern 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 investigate the distribution of F-statistics for stations in the appendix and find that generally stations do not have significantly different dates that may be identified as a structural break.

from 62 minutes to 51 minutes on average after a structural break, a drop of about 20%. Similarly, the *average size of price changes* falls from 4.1 cents to 3.7 cents. Additional summary statistics and discussion are in Appendix A and Tables A1, A2 and A3.

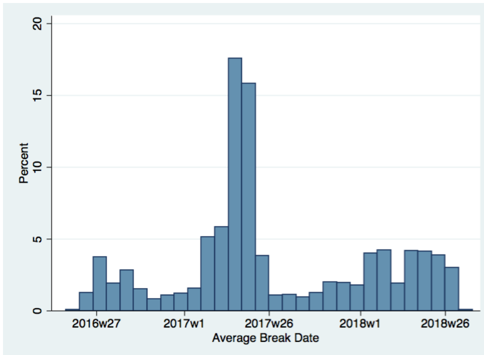
Classification: Since there are many factors that may influence a single adoption measure on its own, to confidently identify a station's decision to adopt algorithmic-pricing software, we look for stations that experienced structural breaks for at least two measures of pricing behaviour within a short period of time. In our main specification, we label a station as an adopter if it experiences a structural break in at least two of the three measures within 8 weeks.[27] Currently, our analysis of adoption is based on data for Super E10 gasoline,[28] and we exclude observations from weekends and holidays. Our results are robust to alternative stricter definitions of adoption.[29]

We observe 3,104 stations break in both number of price changes and average size of price change within 8 weeks. Of the stations that break in rival response time, there are 1,062 that break in the number of price changes within 8 weeks and 812 that break in average size of price change within 8 weeks. There are also 250 stations that break in all 3 measures within an 8 week period. Together, we classify 4,441 stations as adopters as some stations break in more than one pair of measures. This constitutes approximately 30% of stations in our final estimating sample.

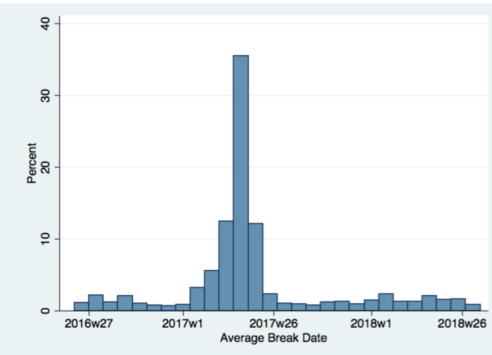
[27] Any combination of two measures will label a station as an adopter.

[28] As a robustness check, we look at structural breaks in each measure for Diesel gas and find the resulting break period distributions are similar for those using E10 gas, see Figure A11. Additionally, we find that the majority of stations that experience structural breaks, experience breaks in both E10 and Diesel gas (between 80-95% of stations experience a structural break in either gas type, experience structural breaks in both types). The timing of structural breaks for each gas type is similar, with the median time difference for E10 and Diesel gas structural breaks being 1 (number of price changes), 2 (rival response time), and 3 (average price change size) weeks.

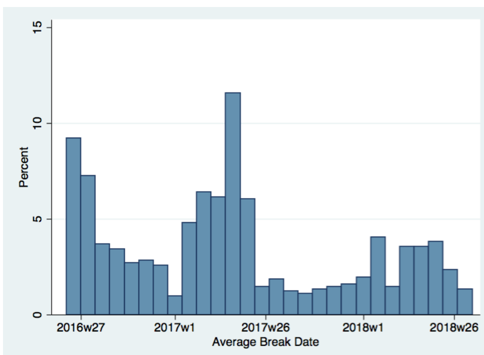
[29] In Section 7.3.2 we change the definition of"a short period of time," requiring stations to experience structural breaks in at least two of the three measures within 4 or 2 weeks. We also include an additional definition that only labels stations as adopters if they experience multiple structural breaks in *both* E10 and Diesel. See Section 7.3.2 for more.



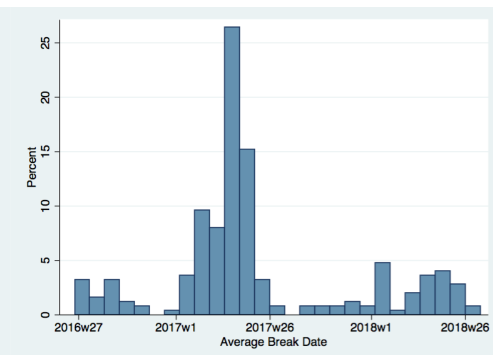
(a) Number of Price Changes and Average Size of Price Change (3,104 stations)



(b) Number of Price Changes and Rival Response Time (1,062 stations)



(c) Average Price Change Size and Rival Response Time (812 stations)



(d) All 3 Measures (250 stations)

Figure 1: Frequency of Average Break Date for Measures Breaking Within 8 Weeks

Figure 1 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 break date. For each measure pair, the largest frequency of average break dates occur in mid-2017. We do see a number of stations that break in both average price change size and rival response time in mid-2016, however about 45% of these stations belong to ARAL and Shell, so it is likely that these breaks may be related to prevailing effects of Shell's 2015 price matching policy. Overall, the fact that we see the largest frequency of multiple measure breaks in mid-2017, the suspected period of large scale adoption, provides some confidence that these measures accurately represent behavioural changes relating to the adoption of algorithmic pricing.

In addition to station-level, our main empirical analysis considers the effects of market-level adoption in isolated duopoly markets. These markets can have zero adopters, one adopter or two adopters. Out of over 1,300 duopoly ZIP markets in our sample, approximately 700 markets had no adopters at any point in our sample. 450 markets have only at most one adopter station throughout the sample period. Approximately 120 of duopoly markets had one adopter station followed by subsequent adoption by the second duopolist. A concern with our definition of adoption and our adoption measures is that a non-adopter responding to an adopting rival will also be labelled as an adopter, for example if they reduce the size of their price changes. Given the large number of markets where one station adopted without its competitor adopting, this is not a substantial problem.[30]

Adopter and non-adopter stations are different. In Table A4 we compare market characteristics for adopter and non-adopter stations before any adoption takes place (in 2016). We find statistically significant differences between the two. Adopter stations are located in denser areas with different demographic profiles. Adopter stations also face more competition. This suggests that the adoption decision is likely endogeneous, with stations deciding to adopt in response to market conditions. If adopters and non-adopters are dissimilar in their observables, it is likely that they are also dissimilar in their unobservable characteristics (e.g., managerial quality, demand and cost shocks). These findings confirm the need to use an IV strategy to address the endogeneity.

6.2 Brand-Level Adoption

In this section we display the distribution of adoption dates for different brands and the evolution of the share of adopting stations by leading brands over time. Table 3 present summary statistics regarding the share of a brand's stations that adopt by the end of each calendar year. We separately show statistics for the largest 5 brands and smaller brands (those with two stations or more).

This table confirms that AI adoption does not happen instantaneously across stations belonging to a given brand. Although we find evidence of station-level adoption as early as 2016 and brand-level adopter shares grow over time, by the end of 2018 only slightly more than 33% of stations belonging to the 5 largest brands in Germany are classified as AI adopters. Summary statistics also capture heterogeneity in adoption based on brand size. On average, a smaller share of non top-5 brand stations adopts than of the top 5 brands.[31] The mean adopter share in 2018 for non top-5 brands is 25%. This likely reflects the better support that larger brands can provide to their stations, which

[30]There are also approximately 40 ZIP codes where both duopolists are labelled as adopters in the same month. This could potentially reflect such concerns about mis-labelled adoption. Practically, this is a small number of markets and they are not driving our main results. We replicated our analysis from Section 7.2 and 8 without these markets and the results remain qualitatively and quantitatively the same.

[31]We find similar patterns in the distribution of structural breaks for individual adoption measures (i.e., number of price changes) in Figure A8 in Appendix A.

Table 3: Share of Stations Classified as Adopters

	Mean	Std. Dev.	Median
	December 2016		
Top 5 Brands	0.053	0.034	0.040
Other Brands	0.037	0.082	0
	December 2017		
Top 5 Brands	0.264	0.098	0.250
Other Brands	0.197	0.207	0.180
	December 2018		
Top 5 Brands	0.337	0.110	0.350
Other Brands	0.246	0.231	0.250

would reduce their cost of adoption. Within non-top 5 brands, there is substantial heterogeneity in adoption shares. By the end of 2018, 20% of non-top 5 brands have adoption rates of over 40%, which is higher than adoption rates for top 5 brands. At the same time, there are 25% of non-top 5 brands with zero AI adoption. This is especially the case for very small brands. The median brand with less than 10 stations has zero AI adoption in 2018.

In addition to heterogeneity across large and small brands in terms of their adoption speed, there is also heterogeneity within the large brands. Figure 2 shows the evolution of the share of adopting stations for the Top 5 brands in our data throughout our sample period. Notably, none of these brands have adoption rates over 50% by the end of the sample period. Adoption clearly happens at a staggered rate that is different across brands. All brands experience spikes in adoption patterns that happen around early/mid 2017. This likely reflects the increased availability of the technology. However, some adoption happens beforehand or afterwards. Total overtakes the relatively early adopter Aral in 2017 and 2018. Esso's adoption rate increases at a steadier (if slower) pace as compared to other brands. The heterogeneity in adoption rates across brands suggests that there is a brand-specific component to AI adoption. As mentioned in Section 5.1, it is likely that some brands were more likely to support the new technology (or adopt at the "HQ" level).

The pattern in Figure 2 is similar to previous patterns of technology adoption in gasoline markets. In Appendix B, we use data from Ontario (Canada) to look at the adoption of electronic payments

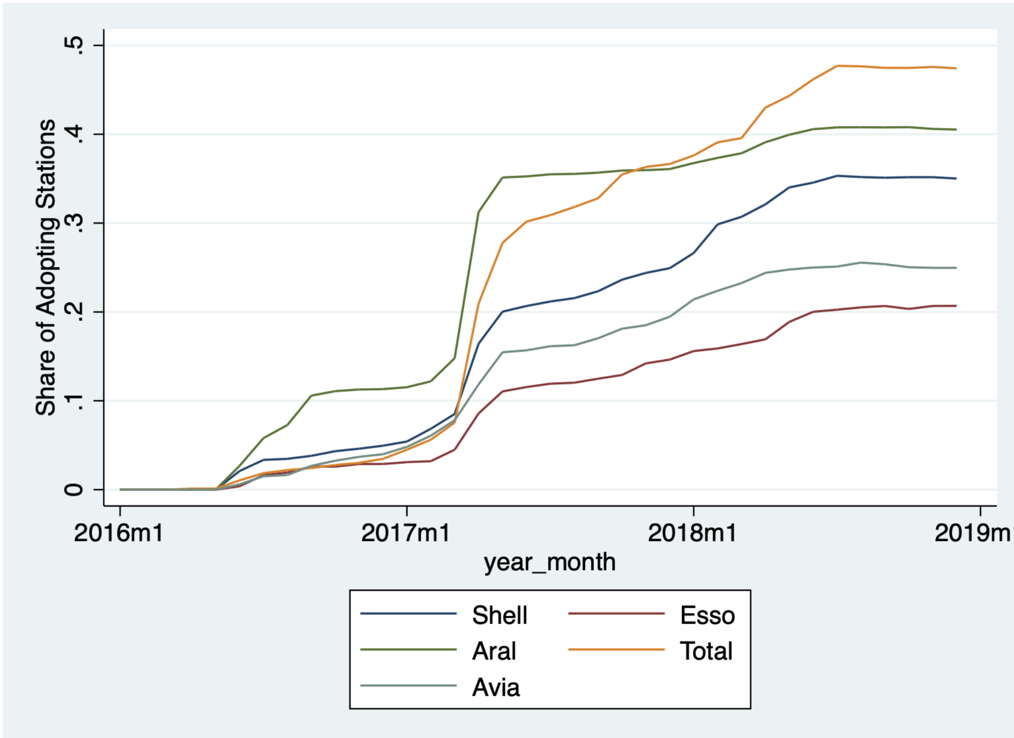


Figure 2: Share of AI Adopters Among Top 5 Brands

by retail gasoline stations in the 1990s. Figure B1 shows the share of adopting stations from 1991 to 2001 for the Top 5 brands in Ontario. Despite the differences in time, geography and technology, the patterns are remarkably similar. In that setting we also find a staggered pattern of technology adoption that appears to be highly brand specific. This suggests that our AI adoption measures do indeed capture technology adoption.

A reasonable question is whether the heterogeneity in brand-level adoption probability can be explained by observable brand characteristics. We find that unlike for 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 A5 shows that conditional on the number of stations in the brand, the share of brand adopters is uncorrelated with the average demographic characteristics of a brand's stations. It is also uncorrelated with a brand's stations competitors. This makes intuitive sense. Brands with more than a few stations will also be spread out across different local 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 our IV estimates presented below.

7 Results – Effects of AI Adoption

7.1 Impact of Adoption on Station Outcomes

We use OLS and 2SLS regressions to measure the impact of algorithmic-pricing adoption on mean, minimum and maximum monthly station margins. For each station and day in our sample we compute an average station/day price and subtract the average price of crude oil, German gasoline taxes and VAT.[32] This provides us with daily station-level margins. We define mean station-level margins by taking an average of these daily station-level margins for each station for each month. We define minimum margins by finding the *lowest* daily margin for each station within a month. We define maximum monthly margins by finding the *highest* daily margin for each station within a month. We label a station as an adopting station if (1) it experiences a structural break in at least two out of three of (i) number of price changes, (ii) average size of price change, and (iii) rival response time, and if (2) these breaks occur within a period of 8 weeks. Non-adopters break in zero or 1 of the three measures.[33] For the time period of AI-adoption, we use the average year-week between the break weeks for each measure.

Station-level estimates are presented in Table 4. In each regression we control for the number of competitors in the station's ZIP code, as well as region/year demographics and the number of stations in station i's brand. Columns (1)-(3) display OLS regression results. Column (4) shows the first stage of the IV regression, and Columns (5)-(7) show the 2SLS estimates. Results from the OLS specification reveal a small negative impact on margins from adoption. Mean margins fall by 0.1 cents. Minimum margins (the lowest daily margin observed in a month) fall by 0.4 cents and maximum margins fall by 0.1 cents.

OLS estimates suggest that AI adoption was not profitable for stations but this is likely an artefact of the endogeneity of the adoption decision, which would attenuate the estimated coefficients downward.[34] To correct for this endogeneity problem, we turn to our IV/2SLS estimates. The first stage of the 2SLS regression is very strong. A 10% increase in the number of other stations affiliated

[32] Data on German gasoline taxes, 0.65 cents per litre, comes from the EU (europa.eu). Data on VAT, 0.247 cents per litre, comes from the OECD.

[33] This is a conservative approach. We may be "missing" some adopters, either due to measurement errors in our 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 estimates towards zero and under-state the true effects of adoption.

[34] As discussed in Section 6.4 and shown in Table A4, adopter and non-adopter stations are very different in their local market demographics and in their competitive environment. They are also similarly likely to be different in their unobservable characteristics.

Table 4: Station-Level Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome:	Mean Margin OLS	Min Margin OLS	Max Margin OLS	Adopter 1st Stage	Mean Margin 2SLS	Min Margin 2SLS	Max Margin 2SLS
Adopter	-0.001***	-0.004***	-0.001***		0.015***	0.035***	0.004
	(0.000)	(0.001)	(0.000)		(0.002)	(0.008)	(0.002)
Share Brand Adopters				0.656***			
				(0.032)			
N Competitors in ZIP	-0.003***	-0.009***	-0.003***	0.006	-0.003***	-0.009***	-0.003***
	(0.000)	(0.001)	(0.000)	(0.006)	(0.000)	(0.001)	(0.000)
Non-Adopter Mean Outcome	0.127	0.073	0.160		0.127	0.073	0.160
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
N Brand Stations Control	YES	YES	YES	YES	YES	YES	YES
Observations	480,235	480,235	480,235	457,083	456,975	456,975	456,975

Notes: Sample is gas station/month observations from January 2016 until January 2019. Mean Margin is the monthly average of daily differences of pump price for station j in month t and crude oil price. Min Margin is the lowest daily difference of pump price and crude oil price for station j in month t. Max Margin is the highest daily difference of pump price and crude oil price for station j in month t. "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 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 j that adopted in period t. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j. 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. Standard errors are clustered at gas station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

with station i's brand that adopt by period t increases the probability that i adopts by period t by 67%. This is consistent with our intuition that adoption of algorithmic pricing is at least in part a brand-level decision.

The 2SLS stage estimates show considerable differences relative to OLS estimates. Column (5) shows that mean margins increase by 1.5 cents on average after AI adoption. This is an increase of about 12% relative to the average non-adopter margin of 0.127 cents. Column (6) shows that minimum margins increase by 3.5 cents after algorithmic adoption. From Column 7 we can see that maximum margins remain the same for adopters and non-adopters. Results suggest substantial endogeneity concerns with OLS estimates that are at least partially solved by IVs.

Algorithmic pricing can increase station margins through a reduction in competition and increased market power. But there can also be other reasons for such changes. An algorithm could better understand underlying fluctuations in crude/wholesale prices, or identify how price elasticity of demand changes over the day or the week and adjust prices accordingly. We test for these different explanations by allowing for heterogeneous effects across different market structures. We separate our sample into two - one sub-sample of stations that are monopolists in their ZIP code, and one sub-sample of stations that are not monopolists. Results of our 2SLS regression for the two subsamples are presented in Table 5.[35] We find that non-monopolist stations are driving the increase in margins, with mean margins increasing for non-monopolist adopters by 1.6 cents post-adoption (12%), in comparison to a negligible and non-statistically significant increase for monopolist adopters. Non-monopolist adopters also experience an increase in both minimum and maximum margins, with increases of 3.2 and 0.6 cents respectively post-adoption. The entire distribution of non-monopolist adopters' margins shifts right.

Relatively large margin changes in non-monopoly markets suggests that there is an increase in market power/reduction in competition after algorithmic adoption. While monopolist adopters do not experience a statistically significant change in mean margins post-adoption, there appears to be a compression effect with minimum margins increasing by 5.6 cents and maximum margins falling by 1.2 cents in comparison to pre-adoption levels. These outcomes could be consistent with the algorithm better understanding how price elasticity changes over the day or the week and avoiding too-high or too-low prices.

There is a concern about the impact of Shell's 2015 price matching guarantees. The introduction of price matching in 2015 appears to have changed pricing strategies (Cabral et al 2018). These changes in strategies may still be ongoing in 2016 and would confound our results. We deal with this concern with two robustness checks. First, we drop all ZIP codes where the price matching guarantee would be relevant. This includes all Shell stations and stations that are in the same ZIP codes as Shell stations. Second, we drop all observations from 2016. Results from these two robustness checks are in Table A7. They suggest that despite dropping around half of the observations in both cases, point estimates remain qualitatively and quantitatively the same.

7.2 Impact of Adoption on Market-Level Outcomes

Table 6 presents estimates of regressions at the ZIP market-level, looking at all duopoly (two-station) ZIP codes.[36] Column (1) shows OLS estimates of Equation (4), using mean market-level margins as the outcome. Market-level margins are calculated as the monthly average of mean daily differences between pump prices for stations in market m in month t and crude oil price. Columns (2)-(4) show 2SLS estimates of Equation (4) using the instruments defined in Equation (5). First-stage estimates of

[35]Results using the alternative 1KM radius market definition are in Table A6 in the Appendix. See additional discussion of alternative market definitions in Section 7.3.1.

[36]Results at the 1km radius market-level are in Table A8 in Appendix A. See Section 7.3.1 for additional discussion of alternative market definitions.

Table 5: Station-Level Results by ZIP Market Structure

Outcome:	(1) Mean Margin 2LSL	(2) Min Margin 2SLS	(3) Max Margin 2SLS
	Sample: Monopoly ZIP Stations		
Adopter	0.009	0.055**	-0.011*
	(0.005)	(0.027)	(0.006)
Non-Adopter Mean Outcome	0.126	0.068	0.160
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	YES
Observations	64,757	64,757	64,757
	Sample: Non-Monopoly ZIP Stations		
Adopter	0.016***	0.032***	0.006**
	(0.002)	(0.008)	(0.002)
N Competitors in ZIP	-0.003***	-0.009***	-0.003***
	(0.000)	(0.001)	(0.000)
Non-Adopter Mean Outcome	0.127	0.074	0.160
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	YES
Observations	392,121	392,121	392,121

Notes: Sample is gas station/month observations from January 2016 until January 2019, 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 j in month t and crude oil price. Min Margin is the lowest daily difference of pump price and crude oil price for station j in month t. Max Margin is the highest daily difference of pump price and crude oil price for station j in month t. "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 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 j that adopted in period t. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j. 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. Standard errors are clustered at gas station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

the 2SLS are available in Table A9 in Appendix A. As was the case with the station-level instruments, the partial correlation between market-level instruments and the endogenous variables is strong.

2SLS estimates suggest that AI adoption by **one** station in a duopoly market **does not** change the distribution of market-level margins relative to a duopoly market where no stations adopted. In contrast, AI adoption by **both** stations in a duopoly market results in **statistically significant margin distribution differences** relative to a duopoly market where no stations adopted. The entire distribution of margins shifts to the right (increases).[37] Mean market-level margins increase by 3.4 cents after market-wide AI adoption. This is a substantial increase of 29% relative to the baseline. The minimum and maximum market-level average margins also shift up. Minimum daily margins increase by nearly 8.2 cents.

One reason for not seeing any market-level effects when only one station adopts and the other does not is simultaneous changes in both stations' margins. An adopter's margins can increase and a non-adopter's margins can fall, cancelling each other out on average. This would mean that competition is not the driver of margin changes. We test this hypothesis by looking at non-adopter stations in duopoly markets and comparing their monthly mean, minimum and maximum margins before and after their rival adopts (compared to duopoly stations in markets where no one adopted). The results of these regressions are presented in Table A11 in the Appendix. We do not see a decrease in margins following a rival's AI adoption.

These results serve as a direct test of theoretical hypotheses about the effects of AI adoption on market outcomes. In Calvano et al (2019), algorithms set prices directly and learn collusive strategies. In Miklos-Thal and Tucker (2019), algorithms do not set prices directly but provide better demand predictions for price-setting human agents. Their results suggest that, in most cases, improving demand prediction allows duopolists to maintain higher collusive prices.[38] We cannot be sure exactly which algorithms station-owners are using in Germany and whether they fully turn over pricing decisions to algorithms.[39] We also do not observe the "counterfactual" competitive or purely collusive prices for a given market. Nonetheless, the complete lack of margin changes from partial adoption and the large increase in margins after complete adoption is suggestive of algorithms facilitating tacit-collusion. As well, the magnitude of margin increases in duopoly 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 algorithms affect margins in Section 8.

[37] We similarly find that mean prices increase in two-adopter markets relative to no-adopter markets. Table A10 in the Appendix shows estimates of the price (rather than margin) effects of AI adoption in duopoly markets.

[38] The exception is moving from *very good prediction* to *excellent/perfect* prediction.

[39] 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 (2019).

Table 6: ZIP Duopoly Market Results

Outcome:	(1) Mean Mkt Margin OLS	(2) Mean Mkt Margin 2SLS	(3) Min Mkt Margin 2SLS	(4) Max Mkt Margin 2SLS
Market Definition:		ZIP Code		
One Station Adopted	-0.001	-0.005	0.010	-0.010
	(0.001)	(0.009)	(0.033)	(0.008)
Both Stations Adopted	0.001	0.034***	0.082**	0.016
	(0.001)	(0.012)	(0.036)	(0.012)
Zero-Adopter Mean Outcome	0.125	0.125	0.061	0.160
IVs	NO	YES	YES	YES
ZIP FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
N Brand Stations Controls	YES	YES	YES	YES
Observations	40,497	36,079	36,079	36,079

Notes: The sample includes duopoly market/month observations from January 2016 until January 2019. A duopoly market is defined as a ZIP code with two gas stations. 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 crude oil price. Min Market Margin is the minimum observed mean market daily difference of pump prices for stations in market m in month t from crude oil price. Max Market Margin is the maximum observed mean market daily difference of pump prices for stations in market m in month t from crude oil price. "One Station Adopted " is a dummy equal to 1 in month t if one of the two stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. "Both Stations Adopted" is a dummy equal to 1 in month t if *both* stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. Regressions in Columns (2)-(4) instrument for adoption using the "share of brand adopters" of the two stations in the market. 1st stage regression results are in Table A9 in the Appendix. 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 two stations at time t. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

7.3 Robustness

7.3.1 Alternative Market Definitions

There are many possible geographic definitions of "markets." A commonly used definition takes advantage of existing geographic designations such as Census tracts, DMAs, or ZIP codes. This is the definition we use in Table 5, but another commonly used definition in the literature looks at the direct distance between stations. Table A6 replicates the analysis in Table 5 but using the following definition of a monopoly: a station that has no competitors within a 1KM radius. Non-monopoly stations then are those that have one ore more competitors within a 1KM radius. Using this alternative definition yields results that are qualitatively similar to the results in Table 5. The most substantial difference is that under the 1KM monopoly market definition, monopoly stations increase their prices after adopting AI (although the average non-monopoly station increase is almost twice as large).

The differences are because the two market definitions label different stations as "monopolists." While a 1KM definition does not vary across different regions, rural area ZIPs are larger than urban area ZIPs. The ZIP definition is therefore more conservative. In rural areas, there are many stations that do not have a competitor within 1KM (on the same intersection, or an intersection away), but that do have a competitor somewhere nearby (in the ZIP code). Table A12 shows a comparison of the number of stations that are labelled as monopolists. Using the 1KM definition over 6,000 stations are classified as monopolists, whereas only 2,300 stations are when using the ZIP definition. Approximately 1,800 of those stations overlap, meaning that the 1KM definition classifies many stations as "monopolist" that have some competitors nearby (perhaps 1.5 or 2KM away). If there is some competition at ranges beyond 1KM, this definition is too lax and would over-state effects for 1KM monopolist stations.

To a lesser extent, this is also the case for duopoly markets. Approximately 3,000 stations are classified as belonging to a duopoly market based on their ZIP codes. The alternative definition based on a 1km radius around each station defines a duopoly market as two stations that are within 1km of one another and that have no other stations within 1km. 3,800 stations are labelled as belonging to a 1km-radius duopoly market. Only 1,100 stations belong to a duopoly market according to both definitions.

Table A8 in the Appendix replicates the analysis in Table 6 using this definition of duopoly markets. Mean margin results are qualitatively and quantitatively similar to the ZIP code definition.[40] This is likely because gasoline stations compete closely with their nearest rivals. Our results in Section 8 suggest that algorithmic adoption by both stations in a duopoly market increases margins by reducing competition. This would also be the case in markets where there are two stations within 1km of one another and other stations further away.[41] The two nearby stations compete more with one another than with stations that are further away. If both nearby competitors adopt, they'll be able to compete less aggressively and increase margins. In that sense, even though the 1km duopoly market definition may be too lax for many (particularly rural) markets, it confirms our baseline results suggesting that what we find is an effect of AI adoption on competition.

[40] Maximum and minimum margin results are qualitatively similar but statistically weaker for the 1KM radius market definition.

[41] For example, consider a ZIP code with two stations within 500m of each other, and two other stations 3-5km away. This is not a duopoly ZIP market, but the two stations within 500m of each other would constitute a 1km duopoly market.

7.3.2 Alternative Adoption Definitions

Tables A13 and A14 replicate Column (5) from Table 4 and Column (2) from Table 6 using alternative definitions of AI adoption. We consider three alternative definitions. Recall that our baseline definition classifies a station as an adopter if, within a period of 8 weeks, it experienced a structural break in at least two out of three measures - number of price changes per day, average size of price changes per day and the speed of response to a rival's price change. Our first two alternative definitions are stricter: we label a station as an adopter if they experienced a structural break in any two out of three measures but within a period of (i) 4 weeks or (ii) 2 weeks. Results for these definitions are qualitatively and quantitatively similar to baseline results.

We also consider a stricter alternative adoption measure that relates to a station experiencing multiple structural breaks in different fuel types. Under this definition, a station has to experience structural breaks in at least two out of our three adoption measures in **both** E10 and Diesel within a period of 8 weeks. As market structure and demand for E10 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 E10 and the Diesel adoption date. Column (6) in each of Tables A13 and A14 present results using this definition of adoption. We find that the results are qualitatively and quantitatively the same as the baseline results at the station level. At the market level, the results are qualitatively the same, although the point estimate for the "Both Stations Adopted" effect is larger and noisier than for the baseline results.[42]

8 Mechanism

There are two main explanations for why pricing algorithms could reach margins above competitive levels. According to the first explanation algorithms *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. The second possibility is that algorithms do fully incorporate rivals' prices and best respond to them, but they *learn how not to compete* (i.e. tacitly collude). For example, algorithms may learn to punish competitors for reducing prices or other

[42]It is likely that taking the average adoption date, rather than the earliest adoption date generates additional noise in "Adopter" variable and in the instruments, since we are effectively mis-labelling some dates where both stations adopted as dates when they did not adopt.

tacitly-collusive strategies. These two explanations have very different implications for competition policy, which should mostly be concerned with algorithms actively learning *not to compete*.

These two explanations also have different predictions regarding the timing of high prices and margins after algorithmic adoption. If the first explanation holds we would expect to see high margins throughout the post-adoption period.[43] If the second explanation holds, we would expect to see no initial effects followed by an eventual convergence towards tacitly-collusive price levels and increased margins. Echoing statements made by AI experts, Calvano et al (2019) point out that it takes a long time for algorithms to train and converge to stable strategies. Without "offline" training, their simulations suggest that training should take several years. Even with offline training, they suggest that it takes up to a year for their algorithms to converge to stability.

We provide some evidence in favour of the second explanation by examining the timing of adoption effects. We present our findings in Figure 3, which shows estimates of time-specific effects for the effect of both stations adopting on mean market margins (i.e. the "Both Stations Adopted" dummy variable, T^2_{mt}, from Equation 4), in a regression that includes market and time FE and controls for whether one station adopted (T^1_{mt}). The time-specific adoption variables are instrumented by time-specific versions of IV^2_{mt} from Equation (5). We start the timing in the month both stations adopt and go until the latest post-adoption period we observe in the sample - 29 months after adoption.

In Panel (a) of Figure 3 we show 2SLS coefficient estimates of the "Both Stations Adopt" variable on average monthly market-level margins. Consistent with Calvano et al (2019), we find that for roughly the first year after both duopolist stations in a ZIP code market adopt AI there are no statistically significant changes in average market margins at the 95% confidence level.[44] However, starting 12 months after both duopolists adopt we find a persistent upward trend in average market-level margins. This trend peaks 20 months after both duopolists adopt at a 3.2 cent (26%) increase in margins above the baseline. Margins fluctuate and become noisier afterwards (there are few markets where we observe more than 20 months following the adoption of both duopolists), but point estimates never fall below the point estimate from 12 months after 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 their prices and margins. Clark, Hortsmann and Houde (2020) also show a lengthy adjustment period to high prices for a Canadian bread cartel.

[43] Or high initial margins followed by lower margins if algorithms learn how to compete more effectively over time.

[44] There are some statistically significant margin increases at the 90% confidence level. 7 months after adoption, for example, margins increase by about 1 cent (8%) above the average outcome for markets where no stations adopted.

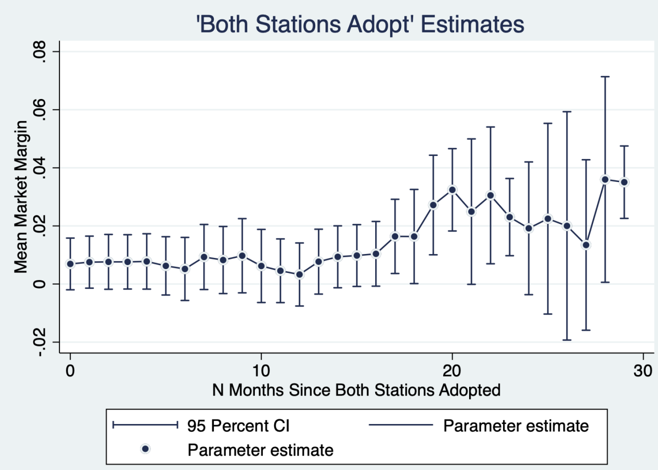
We provide additional supporting evidence of algorithms learning to *not to compete* by looking at the average number of price changes in duopoly markets. The number of price changes per day is one of our main markers of adoption, but, conditional on both stations adopting, it is possible that increases or decreases in competition will play out through changing the "speed" of interactions. Panel (b) of Figure 3 displays coefficient estimates from the same type of regression as Panel (a), but with a different outcome variable: the average number of daily price changes in market m at time t. The average number of price changes is higher in markets in which both stations adopted than in markets where none did.[45] In part, this is a mechanical effect that comes from our definition of adoption. Most interestingly, during the period when margins increase there is a *downward pattern* in the average number of daily price changes. This is not a statistically significant change but it is fairly large - the point estimates fall from 3.9 price changes per day (above the baseline) in the 13th period after both stations adopt to 3.4 price changes per day (above the baseline) in the 21st month after both stations adopt. Similarly, during a fall in average margins (24 to 27 months after adoption), there is a mirror *increase* in the number of mean daily price changes. Once again, the last pattern is driven by a very small number of markets and is quite noisy. These findings are consistent with the number of price changes capturing more or less competitive behaviour.

Brown and MacKay (2020) present a third possible explanation for why prices and margins may increase in response to algorithmic adoption that is neither about algorithms failing to learn to compete nor about algorithms learning to tacitly-collude. In their model, the adoption of algorithmic software transforms 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 without firms resorting to tacitly-collusive strategies. Our data allow us to test an important prediction from their model: in cases of asymmetric adoption (when one firm adopts technology that allows it to respond more quickly to a rival's price changes and the other one does not) the non-adopting firm should have higher prices and margins than the adopter. In general, 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 A11 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

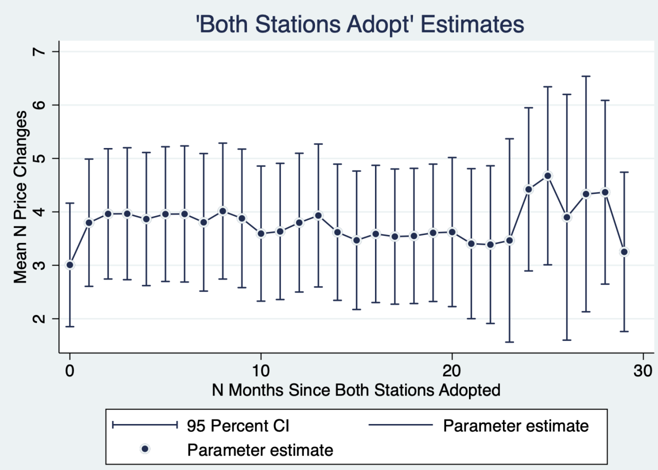
[45] Even in period 0 the number of daily price changes in a duopoly market where both adopt is larger than in a duopoly market where none adopt. This is because markets generally follow a progression - one station adopts, and then another adopts. This means that in period 0 when both stations adopt, one station has already adopted and the average number of price changes is higher than in a market where no stations adopted.

rival's adoption.[46] Although the Brown and Mackay (2020) model appears to fit well certain settings (such as cold medicine markets), in our context it does not seem to apply. Instead we find more support for the Calvano et al (2019) framework.

Overall, our results suggest that algorithmic adopters learn tacitly-collusive strategies over time.



(a) Average Market Margins



(b) Average Market Number of Daily Price Changes

Figure 3: Timing Tests for Duopoly ZIP Market Outcomes

9 Discussion: Algorithmic Collusion, Competition Law and Policy

Our results present the first systematic evidence of the effects of algorithmic pricing software adoption on competition. From the perspective of competition and antitrust authorities, they are troubling. We find that algorithmic pricing software can learn tacitly-collusive strategies, suggesting that widespread adoption of such software can facilitate tacit collusion and raise prices and markups. To the best of our knowledge, this occurs without explicit communication between competitors, making it legal according to current competition laws in many countries.

Our findings have important implications for antitrust authorities around the world and for competition law. Multiple antitrust authorities and economic organizations (OECD 2017; Competition Bureau 2018; Autorité de la Concurrence and Bundeskartellamt 2019; UK Digital Competition Ex-

[46]Point estimates on the non-adopter's mean margins are negative.

pert 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).

If antitrust authorities were to amend laws to account for tacit algorithmic collusion, they would encounter considerable difficulties in doing so. As in all cases of tacit collusion, it is difficult to establish a collusive agreement absent explicit communication. With algorithms, the definition of what constitutes a formal agreement and what constitutes communication should change. For example, repeated interactions between competing algorithms could be a form of communication. There is also an issue of liability when it comes to algorithmic collusion. Antitrust authorities must determine who is at fault in these cases: the algorithm creator, the user, or the algorithm itself.

In Germany, the Federal Cartel Office (Bundeskartellamt) is the competition authority charged with regulating and protecting competition.[47] Germany also has an independent advisory board, the Monopoly Commission (Monopolkommission), tasked with advising the German Federal Government on competition related issues.[48] The Monopolkommission's 2018 report on competition issues in Germany included a discussion on the issue of algorithms and collusion. The report states that further monitoring and evidence is needed to determine whether changes need to be made. If evidence does arise that algorithms lead to the further development of collusive markets, the report suggests that potential revisions could include (i) in cases of prohibited competitive behaviour, imposing the burden of proof that algorithmic usage has not contributed to infringement, and (ii) supplementation of the liability regimen under Article 101 of the TFEU to include review and potential regulation of third parties (i.e. those that contribute IT expertise to algorithms) in cases of collusive behaviour.

While our evidence is particular to retail gasoline markets in Germany (where high frequency pricing data are available), the same algorithmic pricing software is adopted in gasoline retail markets

[47] The Bunderskartellamt's tasks are in accordance with legal provisions provided on both a national and European level. Relevant provisions regarding collusion are articles 101 and 102 of the Treaty on the Functioning of the European Union (TFEU) and the Act against Restraints of Competition (Competition Act - GWB) (The Bundeskartellamt, 2020).

[48] The tasks of the Monopolkommission are regulated by a number of legislative acts including sections 42(5) and 44 to 47 of the GWB (Monopolkommission, 2020).

around the world. At a minimum, our results suggest that competition authorities in Germany and around the world 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.

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 (2019) 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). *PerfectPrice*, another AI retail pricing software, provides specialized solutions for airlines, car rental and vacation rental companies (PerfectPrice.com). Our results suggest that competition authorities should investigate the relationship between algorithmic pricing software adoption and competition in these and other contexts.

10 Conclusion

We investigate potential links between algorithmic pricing and competition by looking at the widespread introduction of AI-pricing 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 before and after the time of adoption.

To identify algorithmic-pricing adoption, we focus on stations that experience structural breaks in at least two out of three measures of pricing behaviour within an 8 week period. Comparing breaks in (i) the number of price changes, (ii) the average size of price changes, and (iii) rival response time, we find that the vast majority of breaks occur in mid-2017, the time at which the AI software became widely available.

Having identified adopting stations we investigate the effects of algorithmic adoption on mean, minimum and maximum margins. 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, AI-adopters with nearby competitors increase mean margins by 12% on average in comparison to pre-adoption levels. These stations also experience higher levels of both minimum and maximum margins post-adoption. In contrast, adopters that are a monopolist in their ZIP code do not see changes in their mean margins. Looking at duopoly (two station) markets exclusively, we find that there is no difference in market-level margins between markets in which no stations adopted and markets in which one of the two stations adopted. However, markets in which both stations adopted show a mean margin increase of nearly 30% and the entire distribution of margins shifts to the right (increases). These estimates are lower-bounds on the true effects, since measurement errors in the first step of the analysis likely result in labelling some AI adopters as non-adopters.

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 AI. 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 both duopolists adopt do not change for about a year after adoption and then increase gradually. 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 in markets. That said, while we identify that adoption has an effect on tacit-collusion between stations, we do not observe *which* algorithm or algorithms stations and brands adopt. We have information about the mass availability of one particular algorithm in Germany but there are other software providers active in this market and stations or brands could have adopted different algorithms. Whether our estimated effects come from multiple stations in a market adopting *the same* or *different* algorithms should have different implications for regulators and competition policy.

References

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

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

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

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

5. Borenstein, S. and Shepard, A. , 1996. Dynamic Pricing in Retail Gasoline Markets. *The Rand Journal of Economics,* 27(3), pp. 429?451.

6. Boswijk, P., Bun, J.H., and Schinkel, M. (2018), Cartel Dating, *Journal of Applied Econometrics* Vol 34, Issue 1 26-42.

7. Brown, Z. and A. MacKay (2020), Competition in Pricing Algorithms.

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

9. Byrne, D. and N. de Roos (2019), Learning to Coordinate: A Study in Retail Gasoline, *American Economic Review*, 109(2), 591-619, 2019

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

11. Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2019), Artificial Intelligence, Algorithmic Pricing and Collusion, forthcoming *American Economic Review.*

12. Chen, L., Mislove, A., and Wilson.C. (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.

13. Clark, R. and J-F. Houde. (2013), Collusion with asymmetric retailers: Evidence from a gasoline price-fixing case, *American Economic Journal: Microeconomics* Vol 5, Issue 3, 97-123.

14. 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* Vol 62, Issue 2, 191-228.

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

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

17. Cooper, W.L., Homem-de-Mello, T. and Kleywegt, A.J., 2015. Learning and pricing with models that do not explicitly incorporate competition. Operations research, 63(1), pp.86-103.

18. Crede, C.J. (2019). A structural break cartel screen for dating and detecting collusion. Review of Industrial Organization, 54(3), pp.543-574.

19. Decarolis, F. and G. Rovigatti. (2019). From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising. Mimeo.

20. 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.

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

22. Erutku, C. and Hildebrand, V. , 2010. Conspiracy at the pump. *Journal of Law & Economics*, 53(1), pp. 223?237.

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

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

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

26. 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*.

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

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

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

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

31. 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.

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

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

34. Klein, T (2019), Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing". Working paper.

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

36. Miklós-Thal, J. and C. Tucker (2019), Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers? *Management Science* 65(4), 1552-1561.

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

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

39. 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.

40. 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

41. O'Connor, J. and N. E. Wilson (2019), Reduced Demand Uncertainty and the Sustainability of Collusion: How AI Could Affect Competition.

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

43. Quandt, R (1960), Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes, *Journal of the American Statistical Association* Vol 55, Issue 290.

44. Sahuguet and Walckiers (2014), Hub-and-spoke conspiracies: the vertical expression of a horizontal desire. *Journal of European Competition Law & Practice* 5, 711?716.

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

46. Slade, M. , 1987. Interfirm Rivalry in a Repeated Game: An Empirical Test of Tacit Collusion. *Journal of Industrial Economics*, 35, pp. 499?516.

47. Slade, M. , 1992. Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies. *The Review of Economic Studies*, 59(2), pp. 257?276.

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

49. 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(4), 630-660

50. Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. Management Science, 54(12), pp. 2024-2038.

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

52. Van Cayseele and Miegielsen (2014), Hub-and-spoke collusion: Some nagging questions raised by economists. *Journal of European Competition Law & Practice* 5, 164-168.

53. Wang, Z. , 2008. Collusive Communication and Pricing Coordination in a Retail Gasoline Market. *Review of Industrial Organization*, 32(1), pp. 35?52.

54. Wang, Z. , 2009. (Mixed) Strategy in Oligopoly Pricing: Evidence from Gasoline Price Cycles Before and Under a Timing Regulation. *Journal of Political Economy*, 117(6), pp. 987?1030.

A Appendix A

A.1 Algorithms in Gasoline Retail Markets

Intelligente Preisstellung leichtgemacht

„Daten sind das neue Öl", sagen Wirtschaftsexperten, und Datenwissenschaftler aus vielen Branchen fördern mit Hilfe von künstlicher Intelligenz (KI) und neuronalen Netzen immer mehr Wissen über Zusammenhänge aus großen Datenmengen (Big Data) zutage. Dieses Wissen nutzen führende Unternehmen für ihren wirtschaftlichen Erfolg, indem es ihnen zu schnelleren und besseren Entscheidungen verhilft – z. B. beim Pricing im Einzelhandel.

„a2i systems A/S" hat den Wert von dynamischem Pricing von Kraftstoffen mit Hilfe von neuronalen Netzen schon 2011 als Erster bei „OK Benzin" in Dänemark durch eine 40 Tankstellen umfassende Untersuchung dokumentiert. Seitdem nutzt „OK Benzin" dieses Verfahren überaus erfolgreich an allen ihren 700 Tankstellen und zwar mithilfe des innovativen Pricing-Systems „PriceCast Fuel®" von „a2i". Weitere Tankstellenbetreiber in Skandinavien und Mitteleuropa sind nach und nach den gleichen Weg gegangen und haben das Pricing in Teilen Europas bereits verändert und neue Maßstäbe für das strategische Pricing mit Wettbewerbsautomatik gesetzt.

„PriceCast Fuel®" unterscheidet sich funktionell deutlich von herkömmlichen Pricing-Systemen, indem es das Kundenverhalten abbildet und danach den optimalen Preis bestimmt, statt nur auf die Preise der Wettbewerber zu schauen und daraufhin zu (re-)agieren.

Das Nachfrageverhalten und die Preisempfindlichkeit der Käufer in Abhängigkeit von Tageszeit, Wochentagen, Feiertagen, Ferienzeiten sowie auch aktuellen Preisen der Wettbewerber werden von „PriceCast Fuel®" stündlich analysiert. Fortgeschrittene Algorithmen berechnen daraus optimale Preisempfehlungen für die jeweils nächsten 24 Stunden. Die Schnelligkeit und die Genauigkeit des Systems, was Timing und Preisgefüge betrifft, ist dabei unübertroffen im Vergleich zu bisherigen regelbasierten Methoden.

Je nach Ihrer Positionierung am Markt und Ihren Geschäftszielen können Sie mit „PriceCast Fuel®" unterschiedliche Strategien einstellen. Am übersichtlichen Kontrollzentrum (Dashboard) des Systems haben Sie jederzeit alle Ihre Tankstellen im Blick: am Ampelsystem – grün, gelb, rot – erkennen Sie sofort, ob Ihre Tankstellen ihre Kennzahlenvorgaben erreichen und welche Ihre Aufmerksamkeit benötigen.

Seit vergangenem Sommer bietet „a2i systems A/S" nun auch unabhängigen mittelständischen Tankstellenbetreibern in Deutschland das hoch innovative Pricing-System als einfach zu nutzende Cloud-Lösung an. Die Beratung, die Anbindung und der Support des Systems werden von „WEAT Electronic Datenservice GmbH" geleistet. Damit ist ein branchenbekanntes und vertrautes Unternehmen als Distributionspartner an Bord, das mit technischer Kompetenz, bundesweiter Reichweite und sehr reaktionsfähigem, deutschsprachigem Support glänzt.

Mit einem 14-wöchigen Proof-of-Concept Programm lässt sich der Effekt und der Nutzen von „PriceCast Fuel®" an einer Auswahl von Tankstellen überprüfen und nachvollziehen. In der Regel ist der Zugewinn an Marge während des Programms sogar höher als der Preis für das Proof-of-Concept Programm. In jedem Fall kann man damit erfahren, wie mit „PriceCast Fuel®" das tägliche Pricing einfacher, transparenter und effizienter wird.

♦ a2i systems A/S

Torben Lauridsen, Business Consultant
Mobil: +49 160 9770 6739
Mail: tbl@a2isystems.com

28 INDUSTRIEPARTNER

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

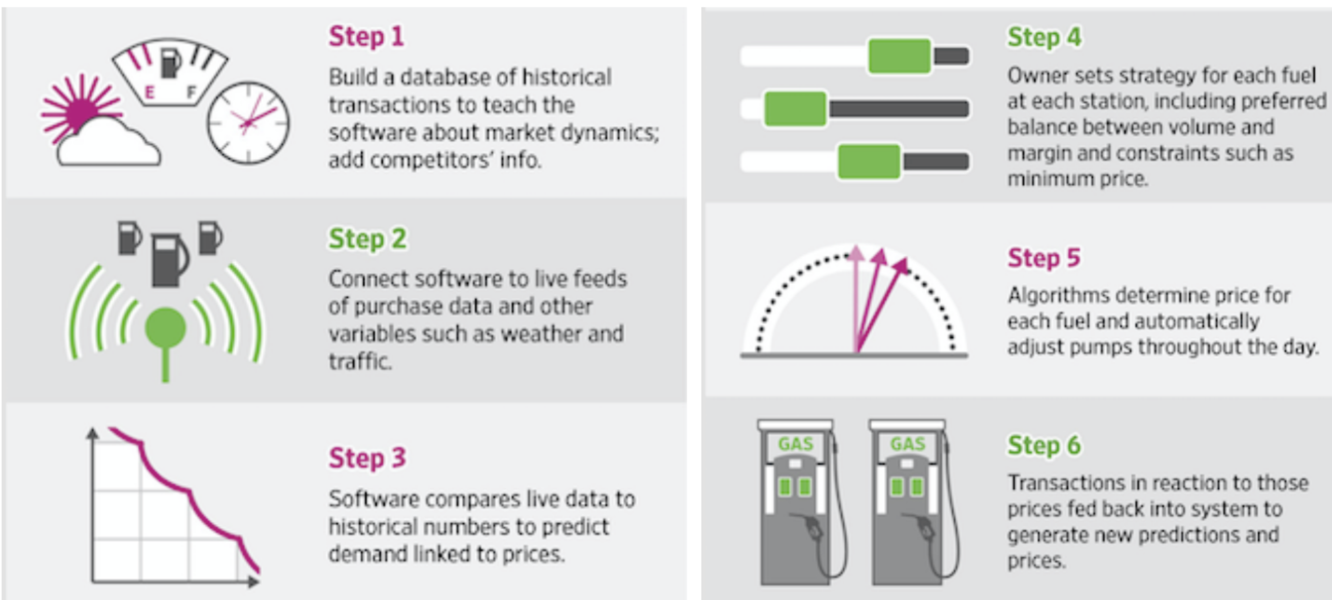


Figure A2: How Algorithms Work (wsj.com)

A.2 Tests for Structural Break in the Three Adoption Markers

A.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.[49] Figure A3 shows the overall distribution of significant breaks.

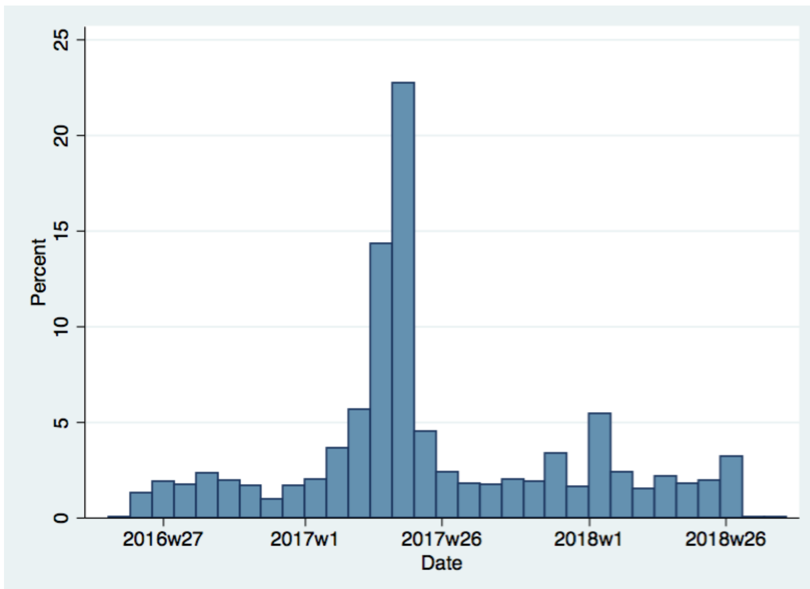


Figure A3: Frequency of Significant Structural Breaks in Number of Price Changes (12,402 stations included)

We compare the number of price changes before and after structural breaks for stations that experienced structural breaks in Table A1. We also include stations that did not experience structural breaks. Adopting stations change their prices much more frequently than non-adopting stations, suggesting that our structural breaks do manage to pick out large changes in pricing strategy. On average, a station that experienced a structural break changes their prices 6 times a day before the break and 10 after the break. There is also a general rightward shift in the distribution of the number of daily price changes after the break.[50] Stations that do not experience breaks are similar to stations that experience breaks in the "pre-break" period.

[49] Any stations that do not have a weekly observation for average number of price changes in every week of 2017 are dropped.

[50] At the 5th percentile of number of price changes, a station only changes their prices once per day before the break but four times a day after the break. At the 95th percentile, a station changes their prices 17 times per day after the break but 10 times before the break.

Table A1: Daily Number of Price Changes

	Mean	Std. Dev.
Post Structural Break Stations	10	4
Pre Structural Break Stations	6	2
No Structural Break Stations	5	4

A.2.2 Rival Response Time

Currently, a rival for station i is defined as any station j that is within a 1KM radius of station i but that belongs to a different brand.[51] 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 j changes its price more than once before station i makes a price change, rival response time is taken as the average of the time gaps between each of station j's price changes and station i's subsequent change. When testing for structural breaks in rival response time, we take into account the fact that changes in response time will be mechanically impacted by changes in number of price changes. To identify structural changes separately from this mechanical effect, we control for the number of price changes when running regression (1) for this measure. Figure A4 shows the overall distribution of significant breaks.

[51] This reflects the average distance of stations in the data, although some stations in rural areas may have a "close" competitor outside of a 1km radius.

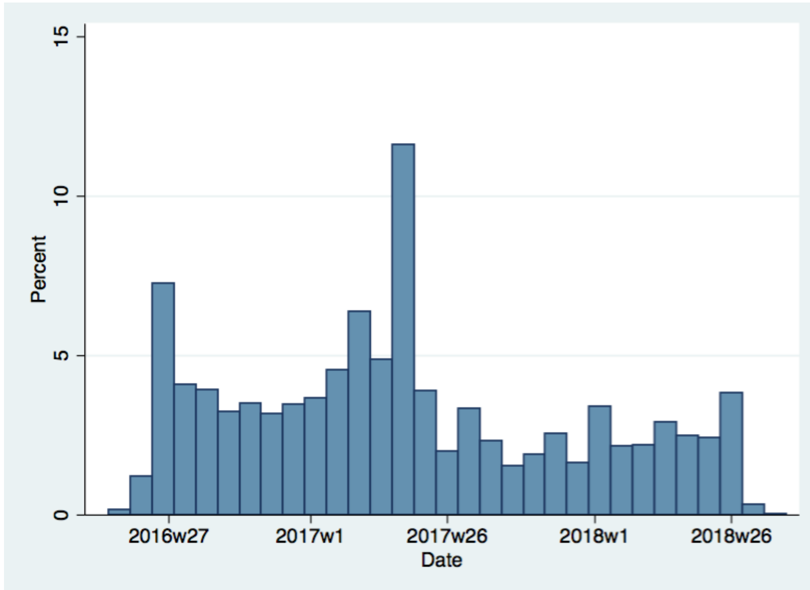


Figure A4: Frequency of Significant Structural Breaks in Rival Response Time (5,449 stations included)

As for the number of price changes, we compare average rival response times (in minutes) for stations that adopted before and after adoption in Table A2. We find that the structural break captures substantial changes in the measure. On average rival response time decreases from 62 minutes to 51 minutes after the structural break, a drop of about 20%. There are also decreases at other points in the response time distribution[52] Stations that did not experience a structural break in this measure look more like the stations in the pre-break period, having average response times of over one hour.

Table A2: Rival Response Time (Minutes)

	Mean	Std. Dev.
Post Structural Break Stations	51	41
Pre Structural Break Stations	62	48
No Structural Break Stations	68	70

[52] At the 5th percentile, response time falls from 15 minutes to 14 minutes. Median response time falls from 51 minutes to 41 minutes. At the 95th percentile, response time falls from almost two hours and a half to less than two hours.

A.2.3 Average Size of Price Change

For the average size of price changes, we calculate the average size of price changes made in a day for each station and then further average this measure to a weekly level. We look at the distribution of weekly break periods for stations with a QLR statistic that is significant at the 5% level. Results are presented in Figure A5. Although there is a spike of stations experiencing structural breaks in average price change size in Spring 2017, there is an even larger frequency of breaks in mid-2016 and a number of stations experiencing structural breaks throughout 2018. The large occurrence of structural breaks in 2016 may be due to prevailing effects of Shell's 2015 price-matching policy, which induced significant changes in pricing behaviour for some German retail gas stations. In particular, during this time, Shell and ARAL began to interrupt the previously observed Edgeworth cycles in the market with upward price jumps around midday. Medium and small retail gas brands would follow these increases, although the extent of the midday price jumps for these stations was not as significant of those of Shell and ARAL (Cabral et al. 2018). Also, as shown by Dewenter and Schwalbe (2016), this price-matching policy led to overall higher price levels of Shell gas stations as well as the average market price of stations operating in the same markets as Shell.

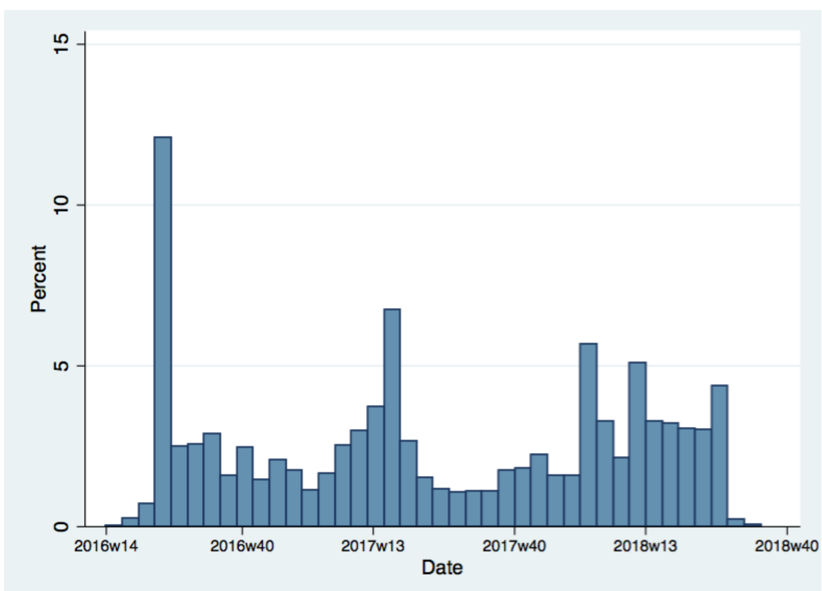


Figure A5: Frequency of Significant Structural Breaks in Average Size of Price Change (11,264 stations included)

In our analysis, if we look at stations that experience structural breaks in 2016, ARAL and Shell stations make up over 40% of these occurrences. The extent to which Shell and ARAL station breaks drives the observed spike in frequency in 2016 can be further appreciated by looking at the distribution of break periods for Shell and ARAL stations in comparison to other stations in Figure A6. While we observe a large spike of structural breaks for Shell and Aral stations in mid-2016, the frequency of breaks for other stations in this period is significantly lower. This comparison provides some support towards the mid-2016 spike in structural breaks being driven by the Shell price matching policy in 2015.

Although we observe a higher frequency of structural breaks in average price change size in 2016 and 2018, there is still a significant number of stations that experience breaks in mid-2017, with a little over 15% of stations experiencing breaks in April-May 2017. As there could be many explanations for why the average price change made by stations could structurally change, we will focus on stations that break in this measure in combination with the other pricing-behaviour measures previously discribed. We discuss this comparison and analysis of stations with breaks in multiple measures in the subsequent section.

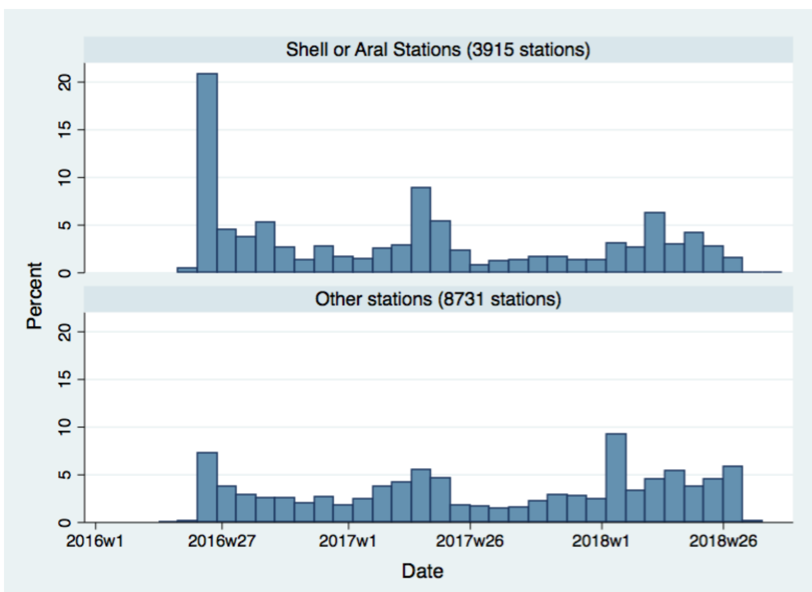


Figure A6: Frequency of Significant Structural Breaks in Average Size of Price Change for Shell and Aral stations vs. other stations)

Differences in average sizes of daily price changes (in cents) before and after the structural breaks is presented in Table A3. For stations that experience a structural break, the average size of price changes falls from 4.1 cents to 3.7 cents. The standard deviation in the average size of price changes falls as well, suggesting that price fluctuations become more targeted after the structural break.[53]

[53]Changes at other points in the distribution are less stark than for the other measures. Partly this is because

Unlike for the other two measures, the average measure for stations that did not experience a structural break is more similar to stations that have already had a structural break than to those that did not. This is not a concern. There is likely massive heterogeneity across the tens of thousands of stations in Germany and stations that adopt AI may be different than those that do not. Our analysis is performed station-by-station. The goal at this point is not to identify the cause or effect of the structural breaks but simply their existence.

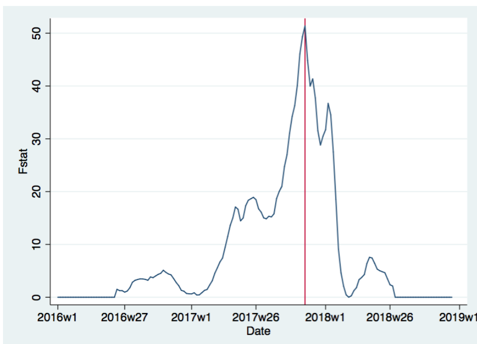
Table A3: Average Daily Price Change Size (cents)

	Mean	Std. Dev.
Post Structural Break Stations	3.7	1.8
Pre Structural Break Stations	4.1	2.4
No Structural Break Stations	3.7	3.8

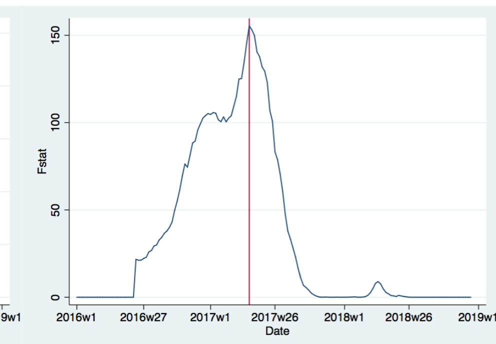
there is a lower bound to the amount by which stations can change their price (0.1 cents). Both before and after the structural break, at the bottom of the distribution stations are actually fairly close to this lower bound. Nonetheless, there are changes at the top end of the distribution after the break. At the 95th percentile, stations reduce their price change size from 7.3 cents to 5.7 cents.

A.3 Alternative Structural Breaks

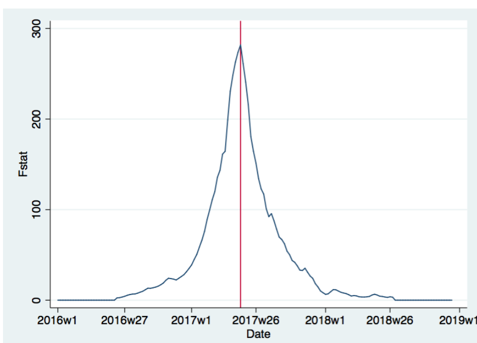
We look anecdotally at the distribution of F-statistics for structural break tests in the number of price changes for stations over the test period. We find that generally, stations display a uni-modal distribution in their F-stastistics, meaning we are unlikely to find structural breaks at a significantly different date if we were to, for example, take the second highest F-statistic rather than the maximum. A few examples are shown in A7 of what a typical distribution would look like for a station's F-statistics for structural break tests in the number of price changes.



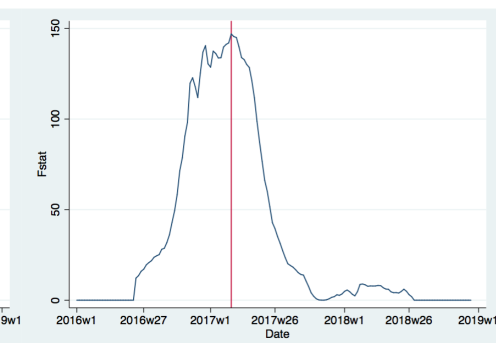
(a) Station 8 (Shell)



(b) Station 22 (Esso)



(c) Station 107 (ARAL)



(d) Station 335 (Hem)

Figure A7: 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 2 highest F-statistics for structural break tests in the number of price changes. We find that for 90% 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 5% of stations have difference of 4 or more weeks between the dates associated with the highest and second highest F-statistic.

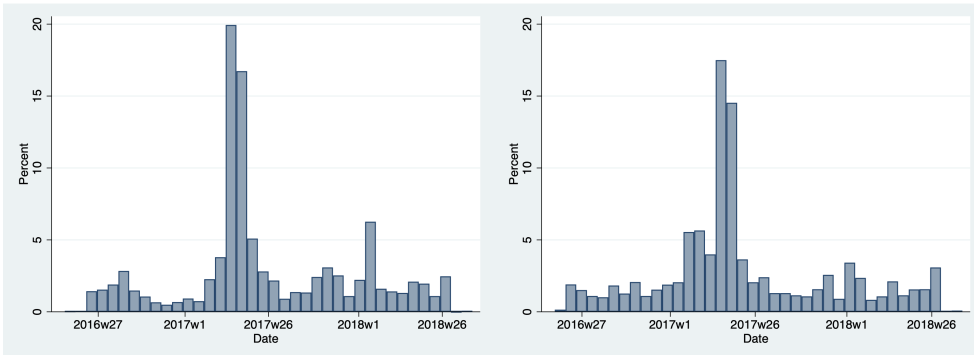
A.4 Adopter/Non-Adopter Heterogeneity

Table A4: Adopter and Non-Adopter Station Characteristics in 2016

	(1)
Outcome:	Will Station j Adopt AI?
Population Density	0.00003***
	(6.67e-06)
ln(Population)	0.011
	(0.035)
Median Population Age	0.004**
	(0.002)
Employment Share	0.079
	(0.079)
ln(region GDP)	-0.005
	(0.032)
N Competitors in ZIP	0.005***
	(0.002)
Observations	155,898

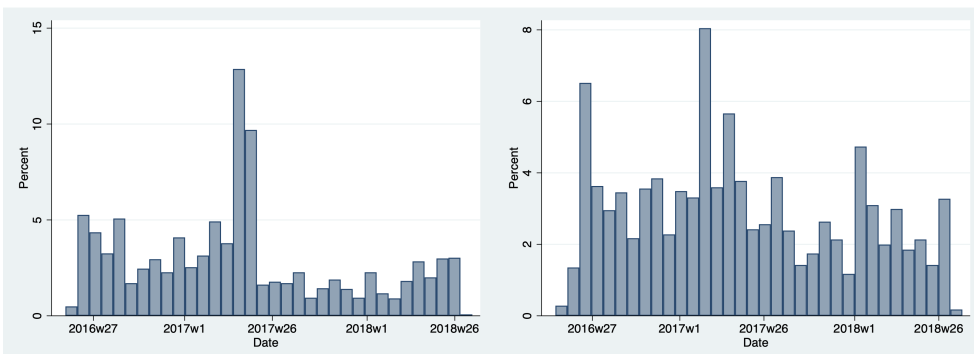
Notes: The sample for this regression includes gas station/month observations from January 2016 until December 2016 that are not labelled as AI adopters during this period. The outcome is a dummy variable equal to 1 if the station will eventually be labelled as an adopter in 2017 or 2018, and zero otherwise. Population Density, ln(Population), Median Population Age, Employment Share and ln(regional GDP) are all computed at the NUTS3-year level. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j in month t. We include month fixed effects. Standard errors clustered at the gas station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A.5 Heterogeneity in Structural Breaks/Adoption by Brand Size



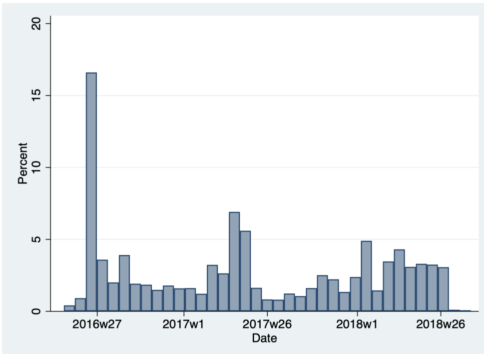
(a) Top 5 Brands (b) Non-Top 5 Brands

Figure A8: Frequency of Significant Structural Breaks in Number of Daily Price Changes by Brand Size

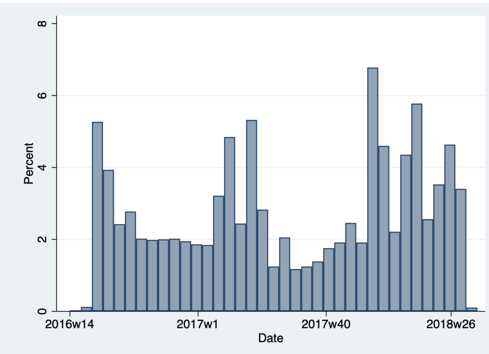


(a) Top 5 Brands (b) Non-Top 5 Brands

Figure A9: Frequency of Significant Structural Breaks in Rival Response Time by Brand Size



(a) Top 5 Brands



(b) Non-Top 5 Brands

Figure A10: Frequency of Significant Structural Breaks in Average Price Change Size by Brand Size

Table A5: Correlates to Brand-Level Adoption Probability

	(1)
Outcome:	Share Brand Adopters
Mean Brand Population Density	0.00002
	(0.00003)
Mean Brand ln(Population)	-0.052
	(0.143)
Mean Bran Median Age	0.010
	(0.008)
Mean Brand Employment Share	0.065
	(0.323)
Mean Brand ln(region GDP)	0.075
	(0.127)
Mean Brand N Competitors in ZIP	0.008
	(0.006)
N Brand Stations	0.0004***
	(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. The variable "Mean Brand X" is a simple average of variable X across all brand b stations in month t. Population Density, ln(Population), Median Population Age, Employment Share and ln(regional GDP) are all computed at the NUTS3-year level. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j 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

A.6 Diesel Gas Structural Breaks

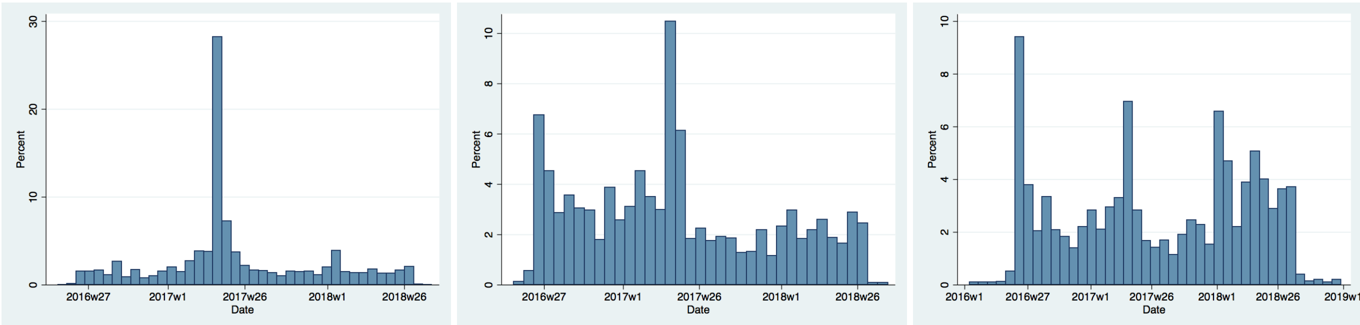


Figure A11: Frequency of Significant Structural Breaks in Number of Price Changes, Rival Response Time, and Average Size of Price Change (Diesel Gas)

A.7 Additional Estimates and Robustness

Table A6: Station-Level Results by 1km Market Structure

Outcome:	(1) Mean Margin 2LSL	(2) Min Margin 2SLS	(3) Max Margin 2SLS
	Sample: Monopoly 1km Radius Stations		
Adopter	0.011***	0.044***	-0.005
	(0.003)	(0.014)	(0.003)
N Competitors in ZIP	-0.002***	-0.003	-0.002***
	(0.001)	(0.002)	(0.001)
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	YES
Observations	191,400	191,400	191,400
	Sample: Non-Monopoly 1km Radius Stations		
Adopter	0.018***	0.031***	0.009***
	(0.003)	(0.010)	(0.003)
N Competitors in ZIP	-0.003***	-0.012***	-0.003***
	(0.000)	(0.001)	(0.000)
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	YES
Observations	265,575	265,575	265,575

Notes: Sample is gas station/month observations from January 2016 until January 2019, split up into two subsamples: one subsample only includes stations that have no competitors within a 1km radius. The other subsample includes only stations that have one or more competitors within a 1km radius. Mean Margin is the monthly average of daily differences of pump price for station j in month t and crude oil price. Min Margin is the lowest daily difference of pump price and crude oil price for station j in month t. Max Margin is the highest daily difference of pump price and crude oil price for station j in month t. "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 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 j that adopted in period t. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Standard errors are clustered at gas station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7: Station-Level Robustness

	(1)	(2)
Sample:	No Shell Markets	Dropping 2016 Data
Outcome:	Mean Margin	Mean Margin
	2SLS	2SLS
Adopter	0.009***	0.018***
	(0.003)	(0.005)
N Competitors in ZIP	-0.001***	-0.003***
	(0.001)	(0.000)
Station FE	YES	YES
Year-Month FE	YES	YES
Annual Regional Demographics	YES	YES
N Brand Stations Control	YES	YES
Observations	207,978	261,938

Notes: Sample in Column (1) includes gas station/month observations from January 2016 until January 2019 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 January 2019 (dropping 2016 data). Mean Margin is the monthly average of daily differences of pump price for station j in month t and crude oil price. "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 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 j that adopted in period t. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Standard errors are clustered at gas station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A8: 1km Duopoly Market Results

	(1)	(2)	(3)	(4)
Outcome:	Mean Mkt Margin	Mean Mkt Margin	Min Mkt Margin	Max Mkt Margin
	OLS	2SLS	2SLS	2SLS
One Station Adopted	-0.002**	-0.008	-0.000	-0.009
	(0.001)	(0.009)	(0.030)	(0.010)
Both Stations Adopted	-0.003*	0.030**	-0.004	0.024
	(0.001)	(0.015)	(0.045)	(0.016)
IVs	NO	YES	YES	YES
Market FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
N Brand Stations Controls	YES	YES	YES	YES
Observations	44,017	40,441	40,441	40,441

Notes: The sample includes duopoly market/month observations from January 2016 until January 2019. A duopoly market is defined as two stations that are within 1km of each other and have no other stations within 1km. 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 crude oil price. Min Market Margin is the minimum observed mean market daily difference of pump prices for stations in market m in month t from crude oil price. Max Market Margin is the maximum observed mean market daily difference of pump prices for stations in market m in month t from crude oil price. "One Station Adopted " is a dummy equal to 1 in month t if one of the two stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. "Both Stations Adopted" is a dummy equal to 1 in month t if *both* stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. Regressions in Columns (2)-(4) instrument for adoption using the "share of brand adopters" of the two stations in the market. 1st stage regression results are in Table A9 in the Appendix. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9: 1st Stage Results for Duopoly Markets

	(1)	(2)	(3)	(4)
Outcome:	One Station Adopted	Both Stations Adopted	One Station Adopted	Both Stations Adopted
Market Definition	ZIP Code		1km Radius	
IV1	0.732***	-0.082	0.762***	-0.007
	(0.167)	(0.096)	(0.172)	(0.097)
IV2	-0.536*	1.208***	-0.971***	0.950***
	(0.318)	(0.237)	(0.302)	(0.247)
Market/ZIP FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Annual Regional Demographics	YES	YES	YES	YES
N Brand Stations Controls	YES	YES	YES	YES
Observations	36,080	36,080	40,441	40,441

Notes: The sample includes duopoly market/month observations from January 2016 until January 2019. Columns (1)-(2) define a duopoly market as a ZIP code with two gas stations. Columns (3)-(4) define a duopoly market as two stations that are within 1km of each other and have no other stations within 1km. "One Station Adopted " is a dummy equal to 1 in month t if one of the two stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. "Both Stations Adopted" is a dummy equal to 1 in month t if *both* stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. $IV1$ and $IV2$ use the "share of brand adopters" of the two stations in the market as follows: for market m at time t, $IV1_{mt} = B_{1t}(1 - B_{2t}) + B_{2t}(1 - B_{1t})$, where B_{jt} is the share of brand adopters for station j in this market. Similarly, $IV2_{mt} = B_{1t}B_{2t}$. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10: Duopoly Market Price Results

	(1)	(2)	(3)
Outcome:	Mean Mkt Price	Min Mkt Price	Max Mkt Price
	2SLS	2SLS	2SLS
Market Definition:		ZIP Code	
One Station Adopted	-0.013	-0.028	-0.009
	(0.009)	(0.038)	(0.008)
Both Stations Adopted	0.023**	0.055	-0.009
	(0.012)	(0.039)	(0.011)
IVs	YES	YES	YES
ZIP FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Controls	YES	YES	YES
Observations	36,080	36,080	36,080
Market Definition:		1km Radius	
One Station Adopted	-0.013	-0.022	-0.001
	(0.009)	(0.033)	(0.009)
Both Stations Adopted	0.018	-0.048	-0.002
	(0.015)	(0.053)	(0.015)
IVs	YES	YES	YES
Market FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Controls	YES	YES	YES
Observations	40,441	40,441	40,441

Notes: The sample includes duopoly market/month observations from January 2016 until January 2019. The top panel defines a duopoly market as a ZIP code with two gas stations. The bottom panel defines a duopoly market as two stations that are within 1km of each other and have no other stations within 1km. Outcome variable Mean Market Margin is the average of mean market daily pump prices for stations in market m in month t. Min Market Price is the minimum observed mean market daily pump prices for stations in market m in month t. Max Market Price is the maximum observed mean market daily pump prices for stations in market m in month t. "One Station Adopted " is a dummy equal to 1 in month t if one of the two stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. "Both Stations Adopted" is a dummy equal to 1 in month t if *both* stations in the market experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. Regressions in Columns (2)-(4) instrument for adoption using the "share of brand adopters" of the two stations in the market. 1st stage regression results are in Table A9 in the Appendix. Regional demographics include GDP, total population, population density, share of population employed and median age a the NUTS3/year level. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Rival Adoption Effects

Outcome	(1) Mean Margin 2SLS	(2) Min Margin 2SLS	(3) Max Margin 2SLS
Rival Adopted	-0.005	0.002	-0.009
	(0.005)	(0.017)	(0.006)
IVs	YES	YES	YES
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	
Observations	61,545	61,545	61,545

Notes: The sample includes all station/month observations belonging to duopoly markets from January 2016 until January 2019 where zero or one of the duopolists adopted AI. Mean Margin is the monthly average of daily differences of pump price for station j in month t and crude oil price. Min Margin is the lowest daily difference of pump price and crude oil price for station j in month t. Max Margin is the highest daily difference of pump price and crude oil price for station j in month t. "Rival Adopted" is a dummy equal to 1 in month t if the duopoly rival of station j in market m experienced a structural break in any 2 of 3 relevant measures in any previous month $\{1, ..., t-1\}$. Regressions in Columns (1)-(3) instrument for a rival's adoption using 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. Standard errors clustered at station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A12: Monopoly and Duopoly Market Definition

N ZIP Monopoly Stations	2,323	N ZIP Duopoly Stations	3,093
N 1km Monopoly Stations	6,072	N 1km Duopoly Stations	3,800
N Overlap	1,857	N Overlap	1,126

Table A13: Station Level Results with Alternative "Adopter" Definitions

	(1)	(2)	(3)
Adopter Measures:	Breaks within 4 weeks	Breaks within 2 weeks	E10 + Diesel
Outcome: Mean Margin			
Adopter	0.016***	0.014***	0.019***
	(0.003)	(0.003)	(0.003)
N Competitors in ZIP	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	YES
Observations	456,975	456,975	456,975
		1st Stage Results	
Outcome: Adopter Dummy			
Share Brand Adopters	0.658***	0.634***	0.671***
	(0.035)	(0.039)	(0.038)
Station FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Control	YES	YES	YES
Observations	456,975	456,975	456,975

Notes: Sample is gas station/month observations from January 2016 until January 2019. Outcome variable Mean Margin is the monthly average of daily differences of pump price for station j in month t and crude oil price. In Column (1) "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 relevant measures within *4* weeks in any previous period. In Column (2) "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 relevant measures within *2* weeks in any previous period. In Column (3) "Adopter" is a dummy equal to 1 in month t if the gas station experienced a structural break in any 2 of 3 relevant measures for both *E10* and *Diesel* gasoline within *8* weeks in any previous period. "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 j that adopted in period t. "N Competitors in ZIP" is equal to the number of other stations present in postal code of station j. 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. Standard errors are clustered at gas station level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A14: Duopoly ZIP Market Level Results with Alternative "Adopter" Definitions

	(1)	(2)	(3)
Adopter Measures:	Breaks within 4 weeks	Breaks within 2 weeks	E10 + Diesel
Outcome: Mean Market Margin			
One Station Adopted	-0.008	0.002	0.007
	(0.015)	(0.008)	(0.013)
Both Stations Adopted	0.032**	0.047	0.058*
	(0.014)	(0.029)	(0.033)
IVs	YES	YES	YES
Market FE	YES	YES	YES
Year-Month FE	YES	YES	YES
Annual Regional Demographics	YES	YES	YES
N Brand Stations Controls	YES	YES	YES
Observations	36,079	36,079	36,079

Notes: The sample includes duopoly market/month observations from January 2016 until January 2019. A duopoly market is defined as a ZIP code with two gas stations. 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 crude oil price. In Column (1) a station is labelled as an adopter if it experienced a structural break in any 2 of 3 relevant measures within *4* weeks in any previous period. In Column (2) a station is labelled as an adopter if it experienced a structural break in any 2 of 3 relevant measures within *2* weeks in any previous period. In Column (3) a station is labelled as an adopter if it experienced a structural break in any 2 of 3 relevant measures for both *E10* and *Diesel* gasoline within *8* weeks in any previous period. "One Station Adopted " is a dummy equal to 1 in month t if one of the two stations in the market adopted in any previous period. "Both Stations Adopted" is a dummy equal to 1 in month t if *both* stations in the market adopted in any previous period. We use the "share of brand adopters" of the two stations in the market as instruments for adoption. 1st stage regression results are in Table A9 in the Appendix. 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 two stations at time t. Standard errors clustered at market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

B Appendix B

We use annual data from Kent Marketing, a leading survey company in the Canadian gasoline market.[54] 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 B1 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. Other brands, like Pioneer, adopt all at once. This likely reflects brand-specific strategies.

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

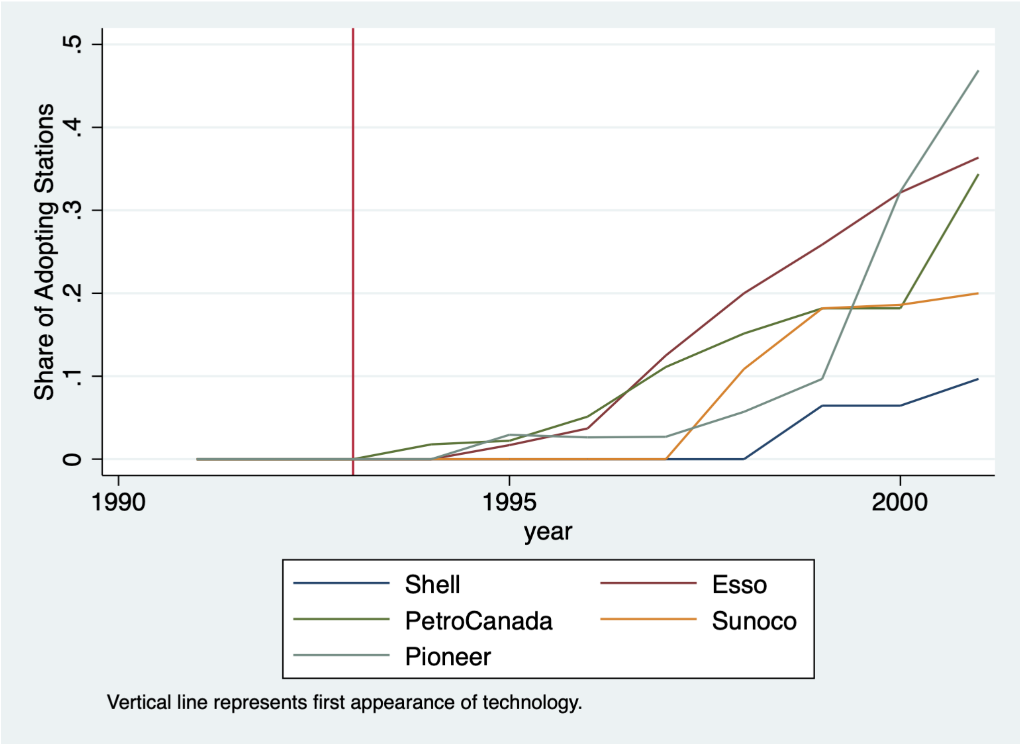


Figure B1: Share of Electronic Payment Adopters Among Top 5 Brands in Canada