Energy Economics 105 (2022) 105711

Contents lists available at ScienceDirect

Energy Economics

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

Won't Get Fooled Again: A supervised machine learning approach for screening gasoline cartels

Douglas Silveira [a,b,*], Silvinha Vasconcelos [c], Marcelo Resende [d], Daniel O. Cajueiro [e,f,g]

[a] Department of Economics, University of Alberta, Canada
[b] Territorial and Sectorial Analysis Laboratory – LATES, Brazil
[c] Department of Economics, Federal University of Juiz de Fora, Brazil
[d] Institute of Economics, Federal University of Rio de Janeiro, Brazil
[e] Department of Economics, University of Brasília, Brazil
[f] National Institute of Science and Technology for Complex Systems (INCT-SC), Brazil
[g] LAMFO-FACE, University of Brasília, Brazil

ARTICLE INFO

JEL classification:
C21
C45
C52
K40
L40
L41

Keywords:
Cartel screens
Price dynamics
Fuel retail market
Machine learning

ABSTRACT

We combine supervised machine learning techniques with statistical moments of the gasoline price distribution to detect cartels in the Brazilian retail market. Standard deviation, coefficient of variation, spread, skewness, and kurtosis are predictors that can help identify and predict anti-competitive market behavior. We evaluate each classifier and discuss the trade-offs related to false-positive (detect cartel when it does not exist) and false-negative (do not detect cartel when it does exist) predictions. The competition authority needs effective monitoring and often anticipating cartel movements. With this in mind, we test the algorithms' performance in new datasets (ex-ante screening). Our results show that false-negative outcomes can critically increase when the main objective is to minimize false-positive predictions. The models' overall average scoring rate for testing and predicting cartels in the same city is 96.22%. When we train the algorithms in one city and predict the cartel outcomes in other cities, on average, the overall scoring rate is equal to 73.75%. Our work suggests that machine learning classifiers have positive attributes and can provide valuable contributions to cartels' deterrence. In addition, we offer a policy prescription discussion for antitrust authorities regarding the pros and cons of proactive tools for inhibiting collusive agreements in retail gasoline markets.

1. Introduction

The discussion about cartel formation is relevant in several markets insofar as it guides the antitrust authorities to enforce competition laws. Given the persistence of collusive market behavior in the gasoline industry, the knowledge about the best screening methods for deterring and inhibiting cartels can guide the regulator in conducting and optimizing competition policies. On the one hand, there is an increasing number of gasoline cartels over the last decades. On the other hand, the competition authorities' financial resources to monitor and intervene in the market are scarce. Thus, statistical screening methods are a useful pro-active auxiliary tool to support and guide an investigation ex officio, once they allow us to identify those candidates most likely to have a collusive agreement (Friederiszick and Maier-Rigaud, 2008).

There is a wide variety of cartel screens that offers the regulator practical and efficient detection methods (Eckert and West, 2004; Bolotova et al., 2008; Blanckenburg et al., 2012; Harrington, 2008b; Perdiguero, 2010; Doane et al., 2015). Several studies consider retail price as the strategic variable to set up a collusive agreement. Besides being easy to measure, it discloses accurate information about how the market works. The main framework assesses anti-competitive behavior via econometric screening methods (Connor, 2005; Abrantes-Metz et al., 2006; Chouinard and Perloff, 2007; Noel, 2007; Abrantes-Metz, 2012; Jiménez and Perdiguero, 2012; Eckert, 2013; Perdiguero and Jiménez, 2020). However, there is no universal consensus on this issue.

In this paper, we combine machine learning techniques with statistical screens. By doing so, we contribute to the discussion on cartel detection and prediction. Huber and Imhof (2019) and Wallimann et al. (2020) uses a similar approach to detect bid-rigging cartels in Switzerland's civil construction sector. However, to the best of our knowledge, our work is the first to assess the performance of statistical moments of the gasoline retail price distribution combined with machine learning

* Corresponding author at: Department of Economics, University of Alberta, Canada.
E-mail addresses: dsilveir@ualberta.ca (D. Silveira), silvinha.vasconcelos@ufjf.edu.br (S. Vasconcelos), mresende@ie.ufrj.br (M. Resende), danielcajueiro@unb.br (D.O. Cajueiro).

https://doi.org/10.1016/j.eneco.2021.105711
Received 24 May 2021; Received in revised form 9 October 2021; Accepted 16 November 2021
Available online 16 December 2021
0140-9883/

algorithms as predictors for anticompetitive retail market behavior. Furthermore, we apply these techniques to prices available to consumers, whose behavior differs from public tenders and bid-rigging cases. Also, as reported in Eckert (2013), given the granularity of the retail fuel market, coordination requires constant supervision and strong commitment, so the importance of good empirical screening tests.

Specifically, the statistical screens derived from the gasoline retail price distribution we use as explanatory variables are adapted from Imhof (2017) and Wallimann et al. (2020). We list them as follows: standard deviation, coefficient of variation, spread, skewness, and kurtosis. Most of them rely on the assumption that reduced retail price volatility can be associated with collusive behavior (Abrantes-Metz et al., 2006).

We are aware that smoother output gasoline prices can also be observed in competitive market situations, such as a consequence of weaker demand periods, supply shocks, or even flatter supply curves. Although relevant aspects, in this paper, rather than focusing on structural changes or even causal inference analysis of the retail gasoline markets, we aim at providing solutions to a prediction policy problem by using supervised machine learning algorithms (Kleinberg et al., 2015; Rabuzin and Modrusan, 2019; Athey and Imbens, 2019; García Rodríguez et al., 2020; Imhof and Wallimann, 2021).

Hence, we believe machine learning techniques add value over traditional econometrics frameworks in the task of helping antitrust agencies to solve prediction problems related to proactive detection of cartels' behavior. To this end, our proposal relies on the hard evidence collected by CADE, which is explicit about the price coordination strategies practiced during the cartel period.

In addition, the literature on cartel screening also teaches us that firms can benefit from periods of structural changes to coordinate the cartel actions (Harrington, 2008a). Crede (2019), and Connor (2005) points out that structural changes and causal inference analysis might not be unambiguous to predict anticompetitive retail market behavior. Thus, it is challenging to distinguish the volatility pattern found during the cartel activities from those associated with the structure of the gasoline market chain (crude oil, wholesale, and distributors). Also, it might ignore cartels' punitive strategies that could equally lead to smoother output prices (Silveira et al., 2021).[1]

Taking the Brazilian market as a case study, we evaluate the out of the sample performance of the proposed methods in a total of 1.918 observations constructed from a weekly database of gasoline selling price in the following cities where collusion was detected: Belo Horizonte,[2] Brasília, Caxias do Sul and São Luís. The study of different regional retail markets for fuels in Brazil is especially appealing as one has often observed recurring suspicions of coordinated practices among firms. The Brazilian competition authority (CADE)[3] has often examined different cases (CADE, 2014). There is also some evidence of non-negligible damages in many Brazilian regions (Cuiabano, 2019; Da Silva et al., 2014; Motta and Resende, 2019).

[1] Merenstein (2016) points that cartels affect the consumer search cost, the number of gas stations (barriers to entry), and the supplier menu costs - to a greater or lesser extent. Ramalho and Ribeiro (2019) evaluate cartel cases in the Brazilian retail gasoline markets and conclude that causal inference analysis of the gasoline price dynamics does not necessarily enhance the regulator's ability to detect cartel behavior.

[2] We also consider the municipalities of Betim and Contagem, which make up the metropolitan region of Belo Horizonte, and was also involved in the cartel agreement.

[3] Administrative Council for Economic Defense. CADE based its decisions on shreds of evidence such as wiretaps, hot documents, text messages, e-mails. Access the following links for details: (i) https://tinyurl.com/yxz8tgnr (available in English); (ii) https://tinyurl.com/y6eoamkp (available only in Portuguese).

We intend to use the history of cases already judged and condemned by CADE for cartel practice (Pinha et al., 2019). The data comprises detected cases and another of no apparent collusion, i.e., the cartel may be in full swing since not discovered. To distinguish between them, we define a binary cartel classification as a dependent variable. The classification criterion for the cartel period is based on the judgments made by CADE, in which the case records contain the exact period in which the explicit evidence that characterized the collusive agreement in each city was collected. Similarly, the non-cartel classification period is based on the time instant the regulator made public the administrative proceeding against gas stations and the operations to disrupt the gasoline cartels.

Our framework evaluates five supervised machine learning models: logit, LASSO (Least Absolute Shrinkage and Selection Operator) logistic, ridge logistic, random forest and neural network. As already mentioned, the inputs of these models are based on the four statistical moments of the gasoline selling price distribution. The output is given by a binary variable that indicates the presence or absence of cartel behavior. Each statistical screen derived from the gasoline retail price distribution in each city captures a different aspect of price dynamics. The combination of several different screens opens an avenue for a better understanding of the differences regarding price agreements in the gasoline market.

By evaluating the so-called Confusion Matrix, we distinguish the classifiers' performance between false-positive and false-negative predictions (Akouemo and Povinelli, 2016). More specifically, a false-positive classification means that the model tags price dynamics as a cartel even though no cartel happens. On the other hand, the false-negative outcome is undesirable as well — once it shows that the algorithm is unable to tag price dynamics as a cartel, even if the cartel happens. Thus, a model that produces many false-negatives can be harmful to the competitive environment. Aware of this, a desirable classification method for the antitrust authority would be capable of balancing the trade-off between false-positive and false-negative outcomes. Our results suggest that the evaluated machine learning techniques are powerful tools for cartel detection in the gasoline market. Furthermore, it demonstrates that in specific cases, the spread, skewness, and kurtosis – which are variables little exploited in the empirical analysis concerning cartel detection in retail markets – are relevant variables to minimize the classification error. The models' average scoring rate for testing and predicting cartels in the same city is 96.24%. When we train the algorithms in one city and predict the cartel outcomes in other cities, on average, the scoring rate is equal to 73.75%.

To develop this discussion, the remainder of this paper is organized as follows. Section 2 reviews the literature on implementing screens to detect cartels in the gasoline retail market. Section 3 describes our data that includes four gasoline cartel cases in the Brazilian fuel retail market and discusses the screens used as predictors for detecting collusive market behavior. Section 4 presents the machine learning techniques. Section 5 discusses the empirical results. Section 6 discusses several policy implications regarding the machine learning algorithms combined with statistical moments screen. Section 7 concludes our work.

2. Literature review

This paper contributes to the empirical literature on behavioral screening methods that discusses ex-post cartel agreements via price dynamics. Considering the framework linked to our research, collusive patterns based on retail price variations stand out. The economic intuition is that the reduced retail price variance across time or within geographical clusters is an indicator of collusion (Abrantes-Metz, 2012; Crede, 2019; Harrington, 2008b). The literature on behavioral cartel screens has grown significantly in the last decade. Most notable are the contributions of Abrantes-Metz et al. (2006) and Bolotova et al. (2008), who propose cartel screens based on the analysis of price variance in an industry.

Most of the behavioral screens so far have been specifically tailored to detect bid-rigging conspiracies, and they are regularly used in auctions (Porter, 2005). The development of behavioral screens for assessing the retail market began only recently. Blair and Sokol (2015) provide many real-world examples of cartel screening methods for detecting collusion in retail markets. The dynamic pricing methodology to generate collusive patterns is widely applied to distinguish a competitive pricing pattern from that observed in cartel agreements (Maskin and Tirole, 1988; Harrington, 2005; Zitzewitz, 2012).

Noel (2007) uses a Markovian regression model to assess the gasoline retail dynamic pricing in Canada. The outcome shows that price cycles prevail when there are many small firms. Wang (2009) evaluated firms' dynamic pricing strategies in the Australian gasoline market before and after implementing a law that constrains firms to set prices simultaneously and only once per day. The Edgeworth price cycle approach captured the oligopoly equilibrium dynamics. In summary, all these results highlight the importance of price commitment in collusive agreements. Clark and Houde (2013) uses official records from a gasoline cartel in Canada to chart firms' colluding price strategies. The cartel leaders compensate low-efficient firms by systematically allowing high-efficient firms to make the last move during coordinated price-increase episodes. Clark and Houde (2014) uses weekly gas station-level data from before and after the cartel's breakdown to compare retail pricing patterns in gas stations affected and unaffected by the ex officio investigation. Among other factors, the results indicate that collusion is associated with asymmetric price adjustments.

Other behavioral issues of economic agents may affect the price variance, significantly impacting gas stations' profits. Accordingly, developing a better understanding of the stochastic prices driving oil and gasoline prices has value for private interests and policymakers (Wilmot and Mason, 2013). Firms in collusion can practice parallel prices. From a theoretical perspective, this strategy would lead to identical price patterns, reducing the price variance (Athey et al., 2004; Harrington and Chen, 2006).

On this subject, many contributions come from the analysis of the Spanish gasoline retail market. Jiménez and Perdiguero (2012) emphasizes the coefficient of price variation as a useful screen to capture the relationship between market structure and price rigidity, a remarkable feature of collusive markets. García (2010) uses a dynamic model based on tacit collusion price strategy to find symmetric behavior on the way companies absorb price changes in the final gasoline price in the Spanish market. For a different period, Contín-Pilart et al. (2009) also reveal that retail prices in Spain respond symmetrically to variations in the wholesale price via the multivariate error correction model. More recently, Perdiguero and Jiménez (2020) shed light on the price coordination capacity of dominant oil operators in the Spanish gasoline market and point out ways for antitrust authorities to increase competition in the gasoline sector.

Finally, there are studies related to cartel agreements in the Brazilian fuel retail market. Amann et al. (2011) focus on white flag gas stations to investigate anti-competitive behavior. The outcomes reveal a negative relationship between retail margins and price dispersion. This might suggest a lack of competition as the white flag gas stations would combine prices with the aim to increase margins. Da Silva et al. (2014) discusses the regional aspects of the asymmetric price transmission. The results shows that it is not a national problem, as the authors find different patterns for each of the Brazilian regions. Nascimento Filho et al. (2018) use the average gasoline retail margin to assess the spatial cross-correlation. By evaluating many Brazilian regions, the study find evidence of tacit collusion in São Paulo and Rio de Janeiro. Assuming ethanol as a perfect substitute for gasoline, Cuiabano (2019) estimates the fuel retailer cartel damage in the south of Brazil (in the city of Londrina, Paraná) via reduced and structural forms for supply and demand. The overcharges range between 6.6% (gasoline) and 12% (ethanol). Silveira et al. (2021) evaluate two different econometric screens for identifying anti-competitive behavior in the fuel market. Using the gasoline cartel judged and condemned in Brasília as a benchmark, the results indicate that the model based on the correlation of the resale price margin and price variability may provide biased likelihood and incorrect identification of cartel behavior over time. The model relying solely on the log deviation of the average gasoline sales price showed better accuracy in cartel detection.

However, the vast majority of works cited above use econometric techniques. Formulated on the nature of each case evaluated in this paper, we provide a performance comparison of different machine learning methods based on gasoline sales price patterns derived from all four statistical moments as a cartel screen. In that sense, the contributions of our research also help to fill a gap in the energy economics literature (Ghoddusi et al., 2019).

3. Database and gasoline industry in Brazil

Oil companies, refineries, distributors, and retailers are responsible for the fuel supply in Brazil. Petrobras, a state-owned mixed economy company, is the largest player in this market, supplying around 97% of type A gasoline and Diesel volumes. In any case, except for fuel retailing, the entire remaining production chain is also very concentrated and regulated by ANP.[4] The role of ANP is to regulate products and firms and provide the SBDC,[5] the Brazilian Competition System, with all the necessary information for any antitrust proceedings initiated (ANP, 2020).

The distribution of gasoline, in turn, is controlled by a small number of distributors,[6] where the four largest companies hold 75% of the market share. But even with controlled input prices, they are free to set the offer price on the market. In this context, the distribution segment is frequently investigated by the SBDC, associating both with cartel formation and increased concentration in the sector. (CADE, 2014). On the policy of gasoline and diesel prices, there is a need for periodic adjustments. The adherence of domestic prices to the international market in the short term triggered a considerable rise between 2017 and 2018, provoking a strike by large truckers in the country, leading the government to subsidize diesel oil prices. Simultaneously, the ANP increased price monitoring to check the amount charged by both producers and distributors of fuels (ANP, 2020).

By its turn, fuel retailing in Brazil is quite atomized and is expanding. The number of gas stations authorized by the ANP in 2014 was 39,763 to 40,970 in 2019 (up 3% in five years) (ANP, 2020). In addition to the structural characteristics of the gasoline industry,[7] it is worth remembering that the gas stations' market behavior places the fuel sector as one of the most investigated by the Brazilian competition authority. Collusive agreements and cartel formation may be the most relevant element in the definition of gasoline selling prices. Paradoxically, although atomized, the geographic market confers considerable

[4] National Agency of Petroleum, Natural Gas and Bio-fuels: http://www.anp.gov.br.

[5] SBDC (Portuguese acronym for Sistema Brasileiro de Defesa da Concorrência). As per defined by Law 12.529/11, the Brazilian Competition System is a set of institutions responsible for competition promotion, prevention and condemnation of anti-competitive practices.

[6] In addition to distributors having the price as the key decision variable, the larger ones also have practices of products differentiation (additive and premium fuels, for example), investment in the brand (advertising), creation of loyalty programs (rewards), and investment in the expansion of scale (construction/attraction of new gas stations; supply to the so-called white flag gas stations; and processes of merger/acquisition of competing distributors). Many of these actions also work as structural barriers to the entry of new distributors and even induce the exit of old distributors (ANP, 2020).

[7] Such as homogeneous products, similar cost structures, government pricing control, local unions, exclusive vertical contracts, barriers to entry, absence of perfect substitutes goods, and the low price elasticity of demand.

Table 1
Number of Cartel and Non-Cartel observations.

	Cartel Obs	Perc. (%)	Non-Cartel Obs	Perc. (%)	Total
Belo Horizonte	220	44.35	276	55.65	496
Date	**01/2004–04/2008**		**01/2014–04/2019**		
Brasília	311	63.60	178	36.40	489
Date	**11/2009–11/2015**		**12/2015–04/2019**		
Caxias do Sul	178	36.78	306	63.22	484
Date	**01/2004–07/2007**		**03/2013–04/2019**		
São Luís	213	47.44	236	52.56	449
Date	**01/2007–03/2011**		**10/2014–04/2019**		
Total	922	48	996	52	1918

local market power to the fuel retailers. This contributes even more to anti-competitive actions in the sector (CADE, 2014). We describe our database in the next section.

3.1. Sample description

ANP is responsible for planning and collecting the retail fuel price database. In this paper, to preserve transparency in our analysis, we use the same database that underlies CADE's decisions on cartel conviction. ANP outsources the prices collection service, as stated in Pedra et al. (2010) and Freitas and Balbinotto Neto (2011). It is divided into the following steps: (a) a weekly collection of the retail prices; (b) quality control of the information; (c) data entry into the system; (d) creation of a database containing the information specified through contracts; (f) forwarding the results to ANP. The field planning within each municipality is based on a geographical identification of the resale points. The weekly collection routes are carried out based on the registration data of resellers in the sample design. The main objective is to optimize the geographical representation of each of them. Finally, a random sample selection is made, containing weekly time-series data for each city in the ANP's records. In the selection procedures, it must observe the geographic coverage of the municipality to guarantee randomness. Given this sampling plan, we have sufficient information to estimate the city-level statistical moment of the gasoline price distribution, such as the average price, the variance, the skewness, and the kurtosis.

3.2. The cartel cases

Table 1 shows the number of the cartel and non-cartel observations for each evaluated city. The first case we analyze happened in the metropolitan region of Belo Horizonte, including the neighboring municipalities of Betim and Contagem.[8] As described in the administrative procedure[9] started in 2014, anonymous complaints date back to the early 2000s. CADE collected the hard evidence between March 2007 and April 2008. Therefore, we consider the period from January 2004 to April 2008 as the cartel phase. To evaluate the regulatory agency performance, we assume January 2014 until April 2019 as the non-cartel period. From Table 2, we see that in Belo Horizonte, out of 996 gas stations registered in the ANP's records, on average, 158 are selected on a weekly basis to collect prices during the evaluated time interval.

Since November 2009, the Brazilian competition authority collects information related to the fuel market in Brasilia. During that time, a considerable amount of economic evidence of cartel formation was

Table 2
Descriptive statistics of the gas stations weekly selected by ANP to monitor the gasoline price.

Number of gas stations in the ANP sample					
	Total	Mean	Std. Dev.	Min	Max
Belo Horizonte	996	158	99.36	30	609
Brasília	411	63	20.94	22	80
Caxias do Sul	200	26	12.05	4	116
São Luís	208	42	11.84	4	127

gathered, involving distributors and resellers.[10] In November 2015, CADE decided to enforce a preventive measure in the administrative investigation regarding the gasoline cartel in Brasília. Thus, we consider November 2009 until November 2015 as a cartel period. The non-cartel period runs from December 2015 to April 2019. The database in Brasília contains 411 different gas stations. On average, ANP selects approximately 63 to monitor the weekly gasoline price.

In Caxias do Sul,[11] the antitrust agency confirmed the evidence that fuel distributors had organized a cartel to fix and standardize prices practiced in fuel resale. The cartel aimed to increase resale margins and eliminate competition, as well as charging excessively high prices. As a result, the municipality's resale margins were much higher than those in other neighboring cities in the state. CADE concluded that there was a violation of the economic order and that the gas stations and their managers adopted a uniform and concerted commercial conduct. The cartel was endowed with a high degree of organization, which is why it lasted, at least, between 2004 and 2007. The conviction was concluded in 2012. Thus, we consider the period between January 2004 and July 2007 as the cartel phase and the period between March 2013 and April 2019 as the non-cartel period. Caxias do Sul has 200 different gas stations in its database. On average, ANP chooses around 26 of them to weekly monitor retails' prices.

In São Luís,[12] intercepted conversations revealed that the owners of gas stations combined prices and induced other stations that sold the cheaper product to increase their values to strengthen the cartel. Such irregularities would have occurred between January 2007 and March 2011. The investigation also has economic evidence resulting from analyses carried out by the ANP on the São Luís fuel resale market. Frequently, these analyses pointed to the existence of elements that would indicate the possibility of concerted conduct between the gas

[8] Resende (2012) had studied the case of Belo Horizonte in terms of the assessment of price synchronization patterns across different fuel stations both for gasoline and ethanol.

[9] All information collected is available at http://en.cade.gov.br/, in the session Procedure Search. The record of the administrative process related to the gasoline cartel in Belo Horizonte is given by 08012.007515/2000-31.

[10] Administrative Process No. 0800.024581/1994-77 and No. 08012.008859/2009-86, available at http://en.cade.gov.br/ and at https://tinyurl.com/us8yffd.

[11] Administrative Process No. 08012.010215/2007-96, available at http://en.cade.gov.br/.

[12] Administrative Process 08700.002821/2014-09 started in October 2014, after receipt of transcripts of telephone interceptions duly authorized by the Judiciary of Maranhão, and other evidence forwarded by the local Public Ministry to the competition authority, conducted a criminal investigation concerning the same offense. The document is available at http://en.cade.gov.br/.

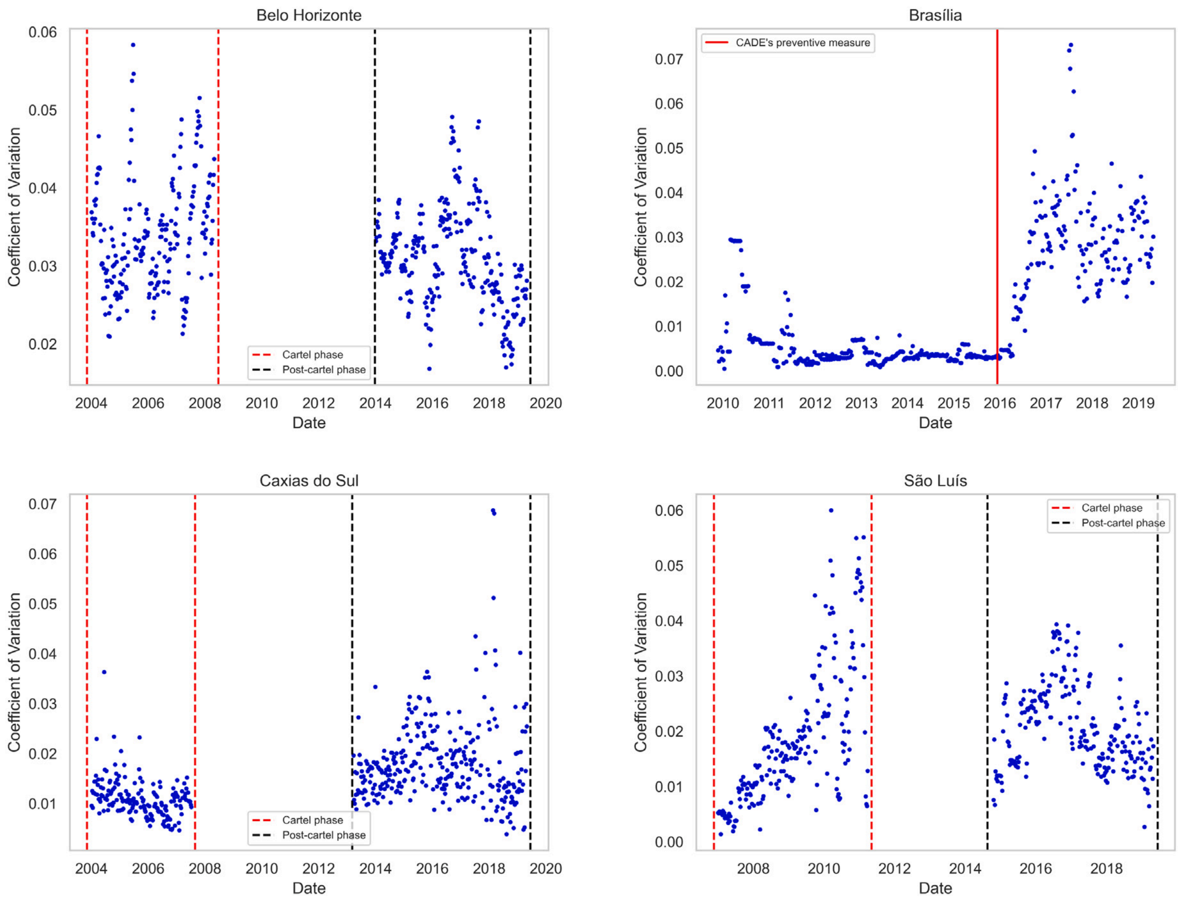


Fig. 1. Coefficient of variation.

station owners in the municipality. Besides, the investigation pointed that the union coordinated a market division to facilitate the operationalization of the illegal agreement. The legal authorities found (at the union headquarters) a map dividing the city into corridors where firms were charging the same price. We consider the period between January 2007 and March 2011 as the cartel phase and the period between October 2014 and April 2019 as the non-cartel period. The database in São Luís contains 208 different gas stations. On average, 42 are selected for assessing the weekly retail price.

3.3. Statistical screens

The statistical screens are calculated from gas station-level data on a weekly basis via a non-overlapping rolling window procedure. Taking the number of weeks for each city in Table 1, we consider the following inputs: standard deviation, coefficient of variation, spread, asymmetry, and kurtosis. These screens derive from the standardized moments of the weekly retail price distribution. Scale-invariant variables enable us to seek different price patterns and distinguish the collusive behavior from non-collusive behavior. The normalized moments calculated in Table 3 allow us to compare the shape of different probability distributions across the cartel and the non-cartel periods via the Kolmogorov–Smirnov and Mann–Whitney tests.

3.3.1. Standard deviation and coefficient of variation

Price coordination might affect gasoline price dispersion within a city. We thus consider the standard deviation of the gasoline selling price as a screen. Besides, we use the coefficient of variation as predictor:

$$CV_{c,w} = \frac{s_{c,w}}{\bar{m}_{c,w}}, \tag{1}$$

in which the terms $s_{c,w}$ and $\bar{m}_{c,w}$ represents the standard deviation and the mean of the gasoline selling price ($P_{c,w}$), respectively, in a given city c during the week w. Fig. 1 shows the prices' coefficient of variation.

3.3.2. Spread

There are both theoretical and empirical justifications for a variance screen for collusion if it is costly to coordinate price changes or if the cartel must solve an agency problem (Abrantes-Metz et al., 2006). According to Wallimann et al. (2020), during cartel periods, the spread of prices might be lower than non-cartel periods. Thus, we define $SPD_{c,w}$ as follows:

$$SPD_{c,w} = \frac{Pmax_{c,w} - Pmin_{c,w}}{Pmin_{c,w}}, \tag{2}$$

where $Pmax_{c,w}$ and $Pmin_{c,w}$ are the maximum and minimum gasoline retail price during the week w in the city c. Fig. 2 shows the weekly spread dynamics in each evaluated city.

3.3.3. Skewness

Price manipulation may affect the symmetry of the distribution of the weekly gasoline selling price. The skewness aims at capturing pricing strategies to define possible retail market divisions. In other words, gasoline cartel members might agree to take turns by temporarily deviating from the cartel scheme, charging lower prices. By doing so, the deviant cartel member can increase its market share in this interim. In this situation, the retail price becomes more asymmetric, i.e., the skewness statistic tends to be lower (or even negative skewed) during the cartel periods. Thus, for a sample of size n, the methods of moments estimator of the skewness yields:

$$skew_{c,w} = \frac{m_{3_{c,w}}}{s^{3}{}_{c,w}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(P_{c,w} - \bar{m}_{c,w})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(P_{c,w} - \bar{m}_{c,w})^2\right]^{3/2}}, \tag{3}$$

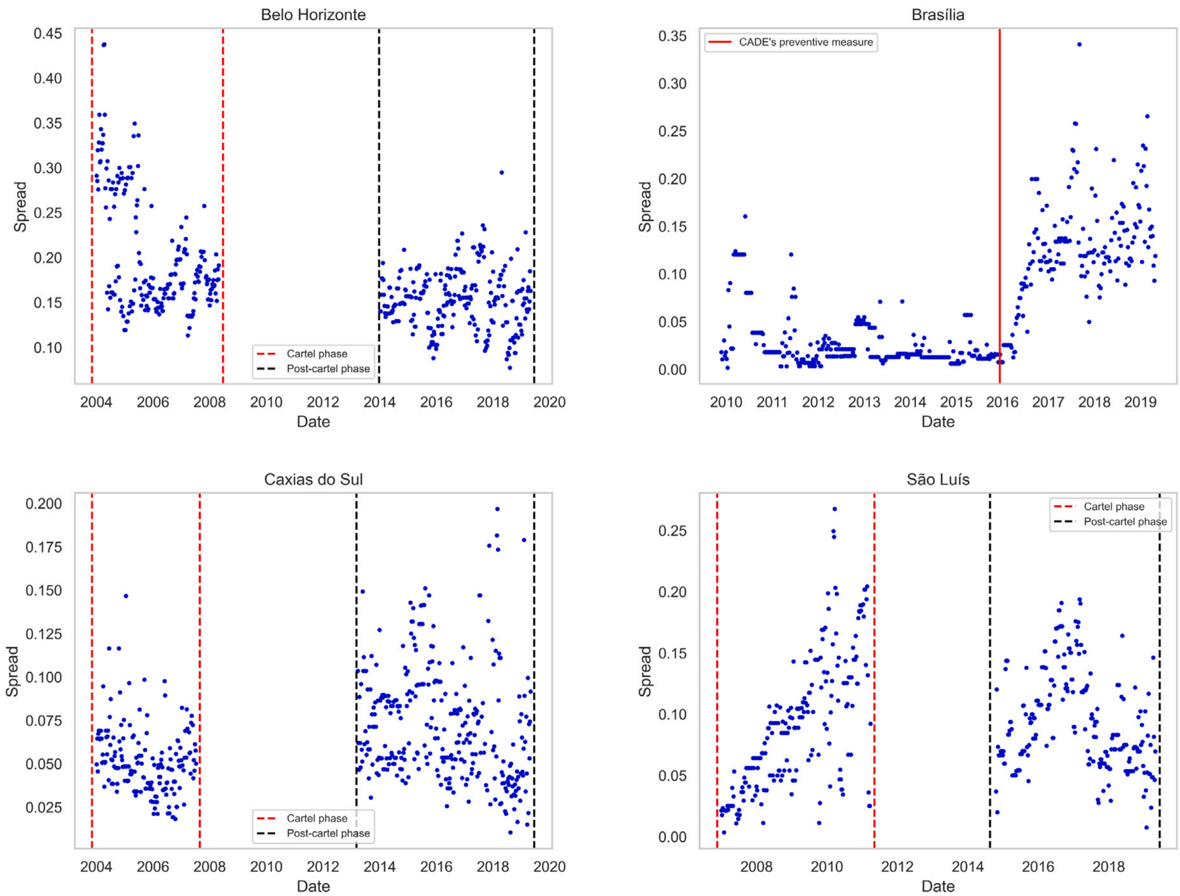


Fig. 2. Spread.

where $m_{3_{c,w}}$ is the sample third central moment of the weekly retail gasoline price within a given city c. Fig. 3 illustrates the weekly skewness observed in each city.

3.3.4. Kurtosis

Finally, we also investigate whether the cartel affects the "tailedness" of the weekly retail gasoline price distribution through coordination. Thus, we have the following expression for the kurtosis:

$$kurt_{c,w} = \frac{m_{4_{c,w}}}{s_{c,w}^2} - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(P_{c,w} - \bar{m}_{c,w})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(P_{c,w} - \bar{m}_{c,w})^2\right]^2} - 3, \tag{4}$$

where $m_{4_{c,w}}$ is the fourth sample moment of the sample variance. Fig. 4 depicts the kurtosis.

3.4. Descriptive statistics

To complement the illustrations as given in Figs. 1–4, we now evaluate the descriptive statistics by separating them between cartel periods and "non-cartel" periods in each city. Note from Table 3 that most screens show fluctuation in the coefficient of variation and standard deviation of prices. Although in different proportions, this same behavior can be seen for the spread, skewness, and kurtosis. In some cities, the difference between the statistical moments is quite noticeable. Typically, during cartel periods, we observer lower spread, standard deviation, and coefficient of variation. Besides, we assess the expected pattern concerning the other statistical moments on a case-by-case basis.

In Belo Horizonte, the mean of the standard deviation screen is approximately 36% lower during the cartel. Thus, prices are more similar in collusive than in competitive periods. On average, both skewness and kurtosis have proven to be higher during the cartel period. This pattern leads to a more compressed distribution of prices during the cartel than in non-cartel periods, suggesting that prices converge in retail gasoline cartels.

In contrast, we notice a considerable difference in terms of the means and standard deviation across the periods in Brasília. The spread of the coefficient of variation is lower in cartel periods with a standard deviation of 0.006, compared to 0.013 for non-cartel periods. The spread predictor is almost six times greater during the non-cartel period. This behavior shows that prices are more similar in cartels, where the price distribution is also highly asymmetric. The kurtosis amounts to 5.0314 in non-cartel periods and more than doubles in cartel periods. Skewness suggests a market division strategy in Brasília.

Compared with Brasília, we observe similarities concerning the standard deviation and spread patterns in Caxias do Sul. In the cartel, these screens show a lower variation than that observed in the non-cartel period, matching with the cartel practice. On average, in collusion, the coefficient of variation is small. The price distribution is more asymmetric. The retail gasoline price skewness in Caxias do Sul is not as diverse as the pattern observed in Brasília. During the anticompetitive period, it is almost 15% more negative.

In São Luís, the spread, coefficient of variation, and standard deviation are slightly low during the cartel phase. On average, the skewness is higher in cartel periods, with a standard deviation of 1.51. The mean of the kurtosis amounts to 7.20 in cartel periods – approximately 60% higher than the non-cartel phase (4.50).

In Table 4, we report the Mann–Whitney and the Kolmogorov–Smirnov test for the predictors in each city. The Mann–Whitney test allows us to investigate whether two independent samples were selected

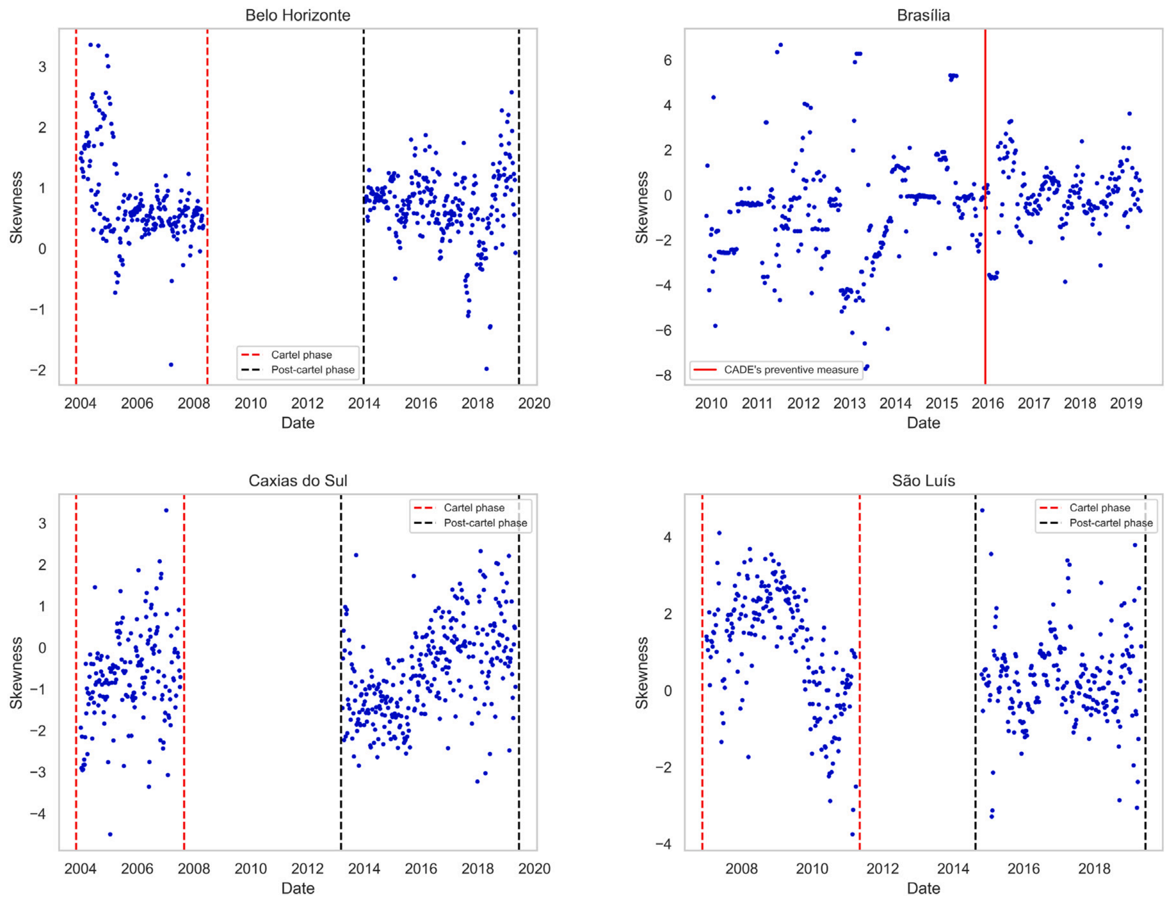


Fig. 3. Skewness.

Table 3
Descriptive statistics of the four evaluated cities.

	Obs	Mean	Std. Dev.	Min	Max		Obs	Mean	Std. Dev.	Min	Max
Belo Horizonte						*Caxias do Sul*					
Cartel period						**Cartel period**					
Spread	220	0.1991847	0.0636429	0.1135867	0.4376187	Spread	178	0.0517836	0.019282	0.0182861	0.1467318
Standard deviation	220	0.0747239	0.0161747	0.0414011	0.11666	Standard deviation	178	0.0284347	0.0091526	0.0125238	0.0866679
Coefficient of variation	220	0.033984	0.0070302	0.020943	0.05835	Coefficient of variation	178	0.0109407	0.0038203	0.0046505	0.0363194
Skewness	220	0.7589087	0.7214225	−1.909971	3.35897	Skewness	178	−0.7019769	1.125306	−4.499262	3.311451
Kurtosis	220	4.806586	4.613041	1.878587	24.96786	Kurtosis	178	5.136701	3.351867	1.387655	25.00011
Non cartel period						**Non cartel period**					
Spread	276	0.1536458	0.0320374	0.0774137	0.294847	Spread	306	0.0727111	0.0328841	0.0105285	0.1968609
Standard deviation	276	0.1111709	0.0215818	0.0591717	0.1793583	Standard deviation	306	0.0679517	0.0335223	0.018549	0.2813645
Coefficient of variation	276	0.0307957	0.0061329	0.016822	0.0491287	Coefficient of variation	306	0.0181618	0.0081338	0.0038722	0.0687003
Skewness	276	0.7070966	0.5847417	−1.978601	2.578997	Skewness	306	−0.6077758	1.111237	−3.228627	2.32932
Kurtosis	276	3.875901	1.669723	2.035779	14.30041	Kurtosis	306	4.208592	2.292335	1.307984	12.00068
Brasília						*São Luís*					
Cartel period						**Cartel period**					
Spread	311	0.027667	0.027835	0.0018349	0.1606426	Spread	213	0.0919739	0.054454	0.0036364	0.2675925
Standard deviation	311	0.0157087	0.0159007	0.0012894	0.0805304	Standard deviation	213	0.0484679	0.0307085	0.0037796	0.1431614
Coefficient of variation	311	0.0054423	0.0059275	0.0004724	0.0294542	Coefficient of variation	213	0.0188849	0.0125611	0.0013737	0.0599945
Skewness	311	−0.7631132	2.325062	−7.714273	6.67475	Skewness	213	1.123083	1.508901	−3.749028	4.110874
Kurtosis	311	10.51547	12.19631	1.053223	64.64728	Kurtosis	213	7.192262	4.23709	1.224377	24.14746
Non cartel period						**Non cartel period**					
Spread	178	0.1236792	0.0577609	0.0077128	0.3410441	Spread	236	0.0936033	0.0395152	0.0075586	0.1938511
Standard deviation	178	0.1063629	0.0489471	0.0105688	0.2383493	Standard deviation	236	0.0727284	0.0256627	0.0107529	0.15
Coefficient of variation	178	0.0269039	0.0130117	0.0028036	0.0731427	Coefficient of variation	236	0.0205591	0.0075181	0.0026962	0.0393634
Skewness	178	−0.0849258	1.404586	−3.853352	3.620465	Skewness	236	0.2761049	1.108964	−3.285877	4.698436
Kurtosis	178	5.03146	4.233462	1.416794	20.87988	Kurtosis	236	4.503887	3.429616	1.092339	26.91653

from populations having the same distribution. In other words, it tests the hypothesis of a zero-median difference between two independently sampled populations.

The Kolmogorov–Smirnov test is a nonparametric test of the equality of one-dimensional probability distributions and allows us to compare two samples. Both tests quantify a distance between the empirical

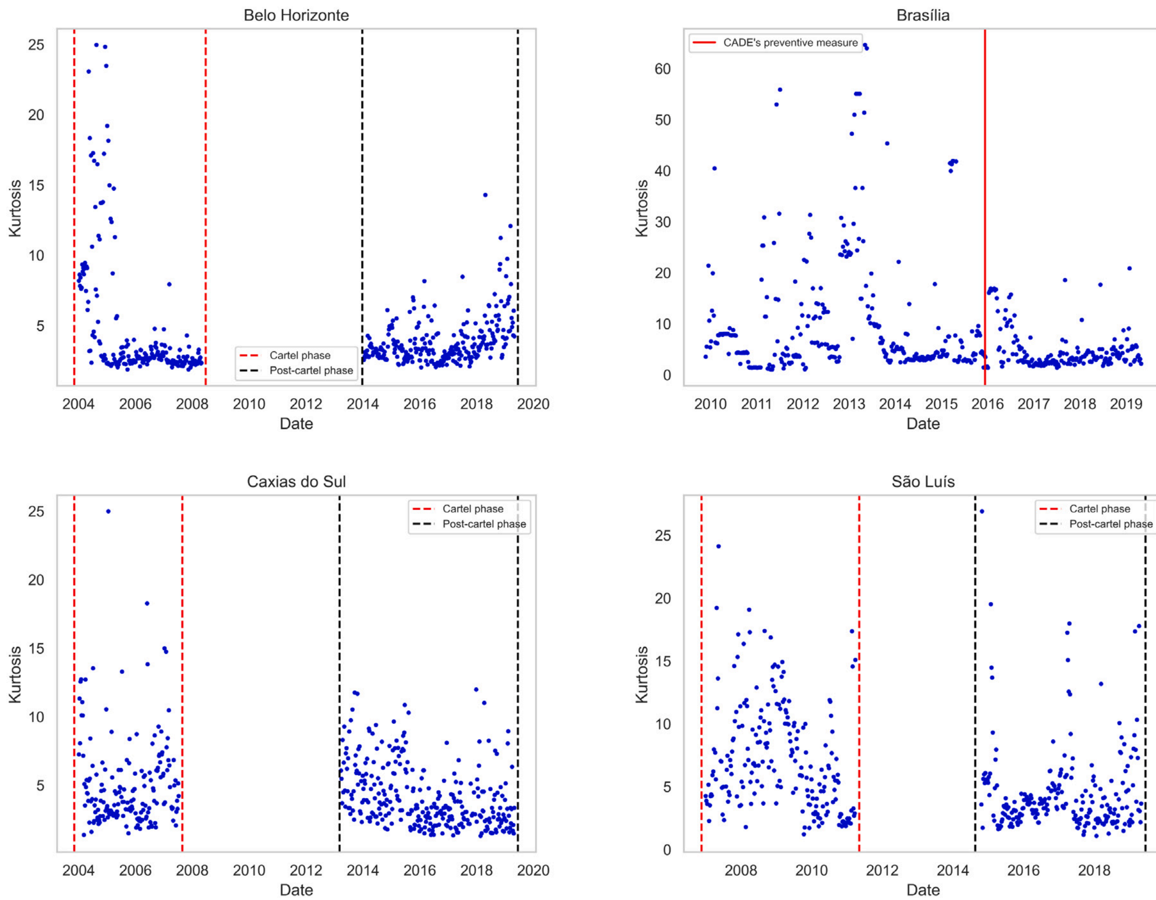


Fig. 4. Kurtosis.

Table 4
Statistical tests for the screens. Screens, z-statistic, p-value MW denote the screens tested, the z-statistic of the Mann–Whitney test and the p-value of the Mann–Whitney test, respectively. KSa and p-value KS denote the asymptotic Kolmogorov–Smirnov statistic and the p-value of the Kolmogorov–Smirnov test, respectively.

Screens	z-statistic	p-value MW	Ksa	p-value KS
Belo Horizonte				
Spread	−8.42	<.0001	0.3148	<.0001
Standard deviation	16.001	<.0001	0.7057	<.0001
Coefficient of variation	−4.787	<.0001	0.2013	<.0001
Skewness	2.042	0.0411	0.2153	<.0001
Kurtosis	4.275	<.0001	0.2989	<.0001
Brasília				
Spread	15.571	<.0001	0.7712	<.0001
Standard deviation	16.98	<.0001	0.8113	<.0001
Coefficient of variation	15.798	<.0001	0.8016	<.0001
Skewness	4.758	<.0001	0.2709	<.0001
Kurtosis	−6.315	<.0001	0.2857	<.0001
Caxias do Sul				
Spread	7.353	<.0001	0.3559	<.0001
Standard deviation	16.656	<.0001	0.8024	<.0001
Coefficient of variation	12.552	<.0001	0.5731	<.0001
Skewness	0.503	0.6151	0.1013	0.198
Kurtosis	−3.164	0.0016	0.1375	0.028
São Luís				
Spread	1.221	0.2222	0.2087	<.0001
Standard deviation	9.302	<.0001	0.4373	<.0001
Coefficient of variation	3.835	0.0001	0.2829	<.0001
Skewness	−7.063	<.0001	0.4287	<0.001
Kurtosis	−7.998	<.0001	0.3809	<0.001

distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null hypothesis assumes samples derived from the same distribution.

Accordingly, in Belo Horizonte, the differences observed between the cartel and non-cartel periods are statistically significant at the 1% level for the standard deviation, the spread, the kurtosis, and the coefficient of variation. The skewness is significant at 5% for the Mann–Whitney test and 1% for the Kolmogorov–Smirnov Test. In Brasília, these differences are statistically significant at a 1% level for all screens.

In Caxias do Sul, the difference between the cartel and non-cartel periods is significant at the 1% level for the standard deviation, the spread, and the coefficient of variation. The kurtosis is statistically significant at a 1% level for the Mann–Whitney test and a 5% level for the Kolmogorov–Smirnov test. However, the skewness is not statistically significant.[13] In São Luís, the spread is not statistically significant. All other predictors are statistically significant at the 1% level for both the MW and KS tests.

[13] As machine learning algorithms seek to find generalizable predictive patterns rather than statistical inference, we still consider the skewness as a predictor in Caxias do Sul. With this in mind, its statistical non-significance does not necessarily compromise our classification task. Also, all other predictors are statistically significant – which points to the attractiveness of screens for predicting retail cartels. This reasoning is in line with Huber and Imhof (2019), which also consider statistical non-significant predictors in their bid-rigging cartel analysis.

Fig. 5. Schematic representation of the statistical screens integrated with supervised machine learning algorithms to classify the gasoline selling price data for each city as cartel and non-cartel behavior.

4. The supervised machine learning algorithms

We evaluate the predictions of the following supervised machine learning algorithms: logit, LASSO, ridge, random forest, and neural network. We use both the cross-validation and random splitting approaches to split the database between the training and testing data.[14] We define accuracy as the gap between actual cartels and correctly predicted cartels. Then, the dependent variable is equal to 1 if the algorithm classifies the cartel probability in a threshold greater than or equal to 0.5 and becomes 0 otherwise.

To assess the performance of out of sample prediction, we consider the following measures: first, the so-called null accuracy, which measures the accuracy that could be achieved by always predicting the most frequent outcome in the database. Second, the so-called score, which measures the proportion of correct classification. Third, miss-classification errors. Fourth, the precision, which measures how often the prediction of cartels is accurate. The fifth is the area under the curve (AUC). The AUC measures the relationship between the share of true-positive predictions against the fraction of false-positive predictions at various threshold settings. An area of 1 represents a perfect prediction and 0.5 represents a low-quality classifier.[15]

4.1. Logit, LASSO and ridge classifiers

Let the dependent variable y_i follow a Bernoulli distribution. The probability of detecting collusive behavior periods in the data is $P(y_i = 1)$ and depends on the vector of explanatory variables. Thus, we have a binary classification model to explain the probability of classes $y_i = 0$ or $y_i = 1$ using the statistical predictors presented throughout Section 3. Typically, the logit classifier is given by:

$$P(y_i = 1) = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}}.$$

For a matrix with n observations and a column of ones to accommodate the intercept, β corresponds to the slope coefficients, x is the vector of predictors p, and i indexes an observation in our database. By maximizing the log-likelihood function, we obtain the parameters estimates as in Hastie et al. (2019) and Pereira et al. (2016):

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \left[y_i\, x_i\, \beta - \log\left(1 + e^{x_i\, \beta}\right) \right]. \tag{5}$$

LASSO and ridge imposes a penalty term to the logit model. Comparing Eqs. (7) and (8), we note that the ridge classifier adds a fine-tuning parameter $\lambda \geq 0$ to the ordinary logistic regression log-likelihood function. Then, to estimate the coefficients, we follow a slightly modified version of the maximum likelihood function as presented in (7), with the addition of a L_2 ridge regularization penalty term:

$$\mathcal{L}_{\lambda}^{ridge}(\beta) = \sum_{i=1}^{n} \left[y_i\, x_i\, \beta - \log\left(1 + e^{x_i\, \beta}\right) \right] - \lambda \sum_{j=1}^{p} \beta_j^2. \tag{6}$$

In summary, ridge classifier includes all the predictors in the final model, adding a squared magnitude on the coefficient β as a penalty term. Hence, if $\lambda \to \infty$, it will lead to underfitting. Increasing λ decreases the variance and increases the bias, and the model becomes less accurate. We use cross-validation to select the value of λ within each evaluated city that minimizes the validation error.

The LASSO classifier model provides an alternative regularization procedure, which allows us to reduce the number of predictors in the final model. By doing so, it bypasses some of the limitations of the ridge classifier model. Hastie et al. (2009) introduces the penalized version of the log-likelihood function to be maximized as follows:

$$\mathcal{L}_{\lambda}^{LASSO}(\beta) = \sum_{i=1}^{n} \left[y_i\, x_i\, \beta - \log\left(1 + e^{x_i\, \beta}\right) \right] - \lambda \sum_{j=1}^{p} |\beta_j|. \tag{7}$$

The LASSO classifier uses a L_1 penalty term. It differs from the traditional logit model since it penalizes the original likelihood function by the absolute sum of the parameters of the model. Depending on the penalty term, the estimator sets the coefficients of less predictive variables to zero. Then, we can select the most relevant features among a possibly large set of predictors.

One drawback of the LASSO regularization is that, when there are strong correlations among terms, it arbitrarily selects which covariates to include in the model. The ridge regularization solves this problem by encouraging highly correlated features to be averaged.[16]

[14] The training sample estimates the model parameters for a given city and contains 70% of the total of observations. The testing sample calculates the out-of-sample predictions and consists of 30% of the total observations. After splitting, the cartel price pattern is estimated in the training sample as a function of a range of predictors, namely the original statistical screens. To parsimoniously assess the trade-off between bias and variance, we repeat these steps 100 times to estimate classifiers' accuracy.

[15] To compute the measures, we create a variable that takes the value 1 for predicted cartel probabilities greater than or equal to 0.5 and takes the value 0 otherwise. Then, we compare it to the actual incidence of collusion in the testing sample. We repeat random sample splitting into 75% training, and 25% test data and all subsequent steps previously mentioned 100 times. Then, we take the averages of our performance measures over the 100 repetitions.

[16] By cross-validation and randomly splitting the training sample into subsamples, we choose the λ that minimizes the average over the miss-classification error estimates. Most of the subsamples are used to estimate the LASSO coefficients under different possible values for λ. One of the subsamples represents the validation database, which we use for predicting cartels based on the different sets of coefficients related to the various penalties and for computing the miss-classification error. After that, we estimate the coefficients of the LASSO logistic by using the training sample. Finally, we predict the cartel probability in the testing sample.

4.2. Random forest

A random forest is an ensemble learning method used in classification tasks. It operates by constructing a multitude of decision trees at training time and outputting the value that appears most often, i.e., the mode, in the individual trees' classes. A Decision tree is a popular method for various machine learning tasks that divide the sample in hyper rectangles and approximates the dependent variable in this region by a constant. Random forests works by averaging multiple decision trees, trained on different parts of the same training set, intending to reduce the variance. It comes at the expense of a small increase in the bias and some loss of interpretability but generally boosts the final model performance (Breiman, 2017).

In our study, we define a vector of features (inputs), X, which is composed by the statistical screens – as summarized in Fig. 5 – that will help us to predict the behavior of our target variable y, that reveals whether the retail gasoline market in a specific evaluated city is under collusion or not (outputs). By doing so, the training algorithm for random forests applies the general technique of bootstrap aggregating,[17] or bagging, to tree learners. Given a training set $X = x_1, \ldots, x_n$ with responses $Y = y_1, \ldots, y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to the following samples:

For $b = 1, \ldots, B$:

1. Sample, with replacement, n training examples from X, Y, call these X_b, Y_b;

2. Train a classification tree f_b on X_b, Y_b.

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(x'), \tag{8}$$

or by taking the majority vote in the case of classification trees. This bootstrapping procedure leads to better model performance because it decreases the variance of the model, without increasing the bias. Additionally, an estimate of the uncertainty of the prediction can be made as to the standard deviation of the predictions from all the individual regression trees on x':

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B} (f_b(x') - \hat{f})^2}{B - 1}} \tag{9}$$

An optimal number of trees B is found using cross-validation. Another way is to observe the out-of-bag error: the mean prediction error on each training sample x_i, using only the trees that did not have x_i in their bootstrap sample. The training and test error tends to level off after some number of trees have been fit. The above procedure describes the original bagging algorithm for trees. Random forests differ in only one way from this general scheme. It uses a modified tree learning algorithm that selects, at each candidate split in the learning process, a random subset of the features.

Thus, if one or a few features are very strong predictors for the cartel, they will be selected in many of the B trees, causing them to become correlated. An analysis of how bagging and random subspace projection contribute to accuracy gains under different conditions is given by Ho (1995). Typically, for a classification problem with w features, $\sqrt{w}$ features are used in each split.[18]

[17] Bootstrap aggregating (also called bagging) is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification. See Breiman (1996) for details.

[18] In practice, the best values for these parameters will depend on the problem, and they should be treated as tuning parameters (Hastie et al., 2009).

4.3. Neural network

A neural network is composed of an n_l series of layers known as neurons. The layer l of the neural network has M_l neurons in parallel. Each neuron in layer l applies a nonlinear transformation on its M_{l-1} inputs. We can formalize the model as follows:

$$y_k^{(l)} = h^{(l)}\left(\sum_{i=1}^{M_{l-1}} \omega_{ik}^{(l)} y_i^{(l-1)} + \omega_{0k}^{(l)} \right), \quad k = 1, \ldots, n_l, \tag{10}$$

where $a_k^{(l)} = \sum_{i=1}^{M_{l-1}} \omega_{ik}^{(l)} y_i^{(l-1)}$ is the activation of the neuron k and the term $\omega_{0k}^{(l)}$ measures the bias associated to an entry $y_0^{(l-1)} = 1$. The term $h^{(l)}$ is the activation function of the neurons in layer l. By definition, we have that $y_i^0 = x_i$ where $i = 1, \ldots, M_0$ represents the inputs of the neural network. Regarding the target variable, we have that $y_i^{nl} = y_i^0$, in which $i = 1, \ldots, M_{nl}$ represents the output of the neural network. Thus, the neural network has $M_{nl} = M_0$ outputs. In our study, the inputs of the neural network are the statistical moments of the retail gasoline price. The output is our so-called target variable, i.e., the cartel predictions that take values between 0 and 1.

5. Empirical results

We start our empirical analysis by presenting the results through the confusion matrix for all machine learning techniques evaluated in each city. In predictive analytics, a confusion matrix is a table with two rows and two columns that reports the number of false-positives, false-negatives, true positives, and true negatives. In statistical hypothesis testing, a false-positive (negative) corresponds to the Type I (II) error.

Thus, each row of the matrix represents the instances in a predicted class (cartel and non-cartel periods) while each column represents the instances in an actual class. This allows a more detailed analysis than a mere proportion of correct classifications (score). A score is not a sufficient metric for the real performance of a classifier. As it does not tell us the underlying distribution of response values, it will yield misleading results if the datasets is unbalanced (Fawcett, 2006; Sammut and Webb, 2011; Powers, 2011).

In other words, it does not inform about the types of errors the classifier is making. For example, if there were 95 cartel observations and only 5 non-cartel observations in the data, a particular classifier might classify all the observations as cartels. The overall score would be 95%, but in more detail, the classifier would have a 100% sensitivity, i.e., the recognition rate for the cartel class but a 0% recognition rate for the non-cartel class.

5.1. Belo Horizonte

We first remember that the Belo Horizonte database contains a total of 496 weeks (observations), of which 220 labeled as the cartel period. The testing sample for the machine learning algorithms performances includes 30% of the total sample. As the confusion matrix in Belo Horizonte reveals, by adding all the entries for each machine learning algorithms as shown in Fig. 6, we evaluate the average of the predictions based on 149 observations. Before going deep into this analysis, it is important to highlight the following point: by convention, we describe the class encoded as 1 as the positive class (cartel) and the class encoded as 0 as the negative class (non-cartel). In that sense, the true positive (negative) represents the case in which the model correctly predicted a 1 (0) value. As well, we considered a classification threshold, i.e., the probability for the decision rule equal to 0.5.

By looking at the Logit predictions,[19] we see on the bottom right the number of true positives, which indicates that in 67 cases the classifier

[19] It is worth mentioning that the values of the output of the logit model as expected are not precisely equal to 0 or 1 as the logit line in Fig. 7 may suggest. They are either smaller than 0.01 or greater than 0.99.

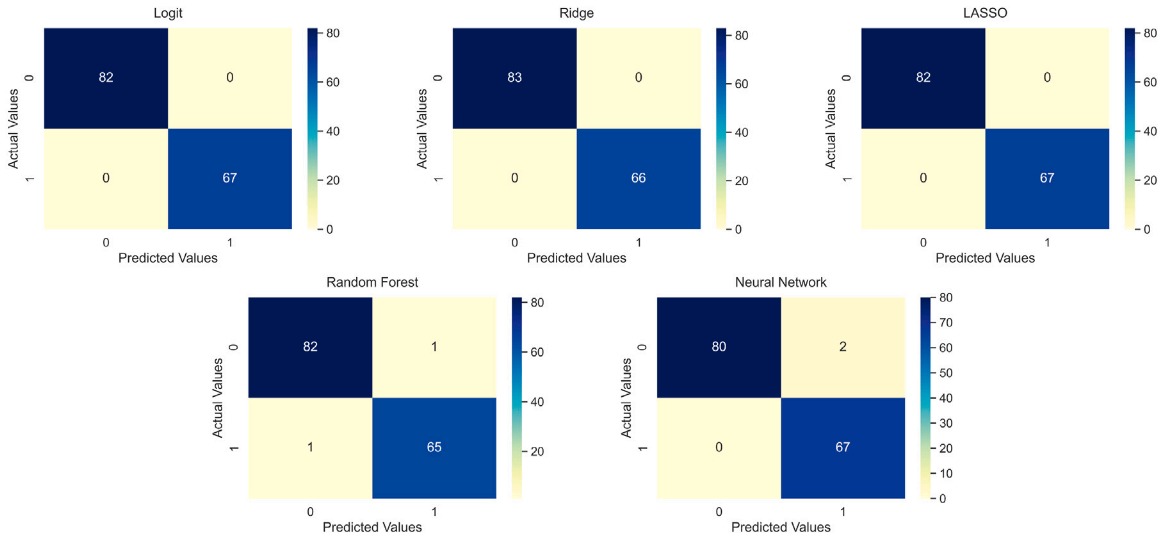
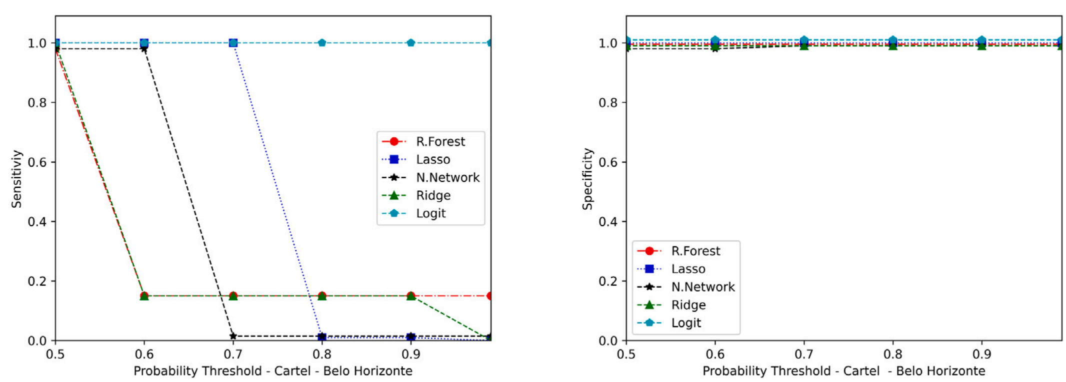


Fig. 6. Confusion matrices - Belo Horizonte.



(a) (b)

Fig. 7. Sensitivity and specificity results by restricting the cartel classification rule.

correctly predicted the cartel period. On the upper left, we observe the number of true negatives, which indicates that in 82 cases the classifier correctly predicted the non-cartel periods. On the upper right, we have the number of false-positives. It indicates that the Logit classifier does not incurred a Type I error. In addition, on the bottom left we see that the Logit model does not incur a Type II error. To compute the classification accuracy score, we first must add true-positives and true-negatives. In sequence, we divide that amount by the total number of observations, i.e., the score is equal to $(82+67)/125 = 100\%$. The Lasso and ridge classifiers has this same score. The Random Forest and Neural Network algorithms have the same classification score (98.66%).

We can assess the classification error metric by adding the false-positives and false-negatives and dividing that amount by the total number of observations. In that sense, we can infer that the Logit misclassification error is given by $(0 + 0)/125 = 0\%$. Together with the ridge and LASSO algorithms, this is the smallest classification error observed for Belo Horizonte. In contrast, for the random forest and neural network models we have an error equal to 1.34%.

Fig. 7 reports two metrics used for evaluating the trade-offs in classification accuracy. The sensitivity assesses the true positive rate and aims to measure the proportion of actual positives correctly identified. The specificity is also known as the true negative rate and measures the proportion of actual negatives correctly identified. For both sensitivity and specificity, the best possible value is 1.

In the confusion matrix, sensitivity is calculated by dividing the true positives by the total of the bottom row. Still considering a probability threshold equal to 0.5, the sensitivity of the Logit classifier is equal to $67/(0 + 67) = 1$. Note that ridge, LASSO, and neural networks classifiers have this same true positive rate (sensitivity). The random forest algorithm has a slightly lower sensitivity $(65/(1+65) = 0.985)$.

Specificity is calculated by dividing the true negatives by the total amount in the top row. Hence, for the Logit classifier, the true negative rate (specificity) is given by $82/(82 + 0) = 1$. The ridge and LASSO classifiers have this same specificity. The random forest has a specificity equal to $82/(82 + 1) = 0.988$. The true negative rate for the neural network is given by $80/(80 + 2) = 0.976$. From the confusion matrix, we can calculate the precision metrics by dividing true positives by the total of the right column. By doing so, the Logit classifier has a precision equals to $67/(67+0) = 100\%$. It is the same precision observed for both the ridge and LASSO algorithms. The precision of the random forest is $65/(65 + 1) = 98.48\%$. The neural network precision is slightly smaller than the others $(67/(67+2) = 97.10\%)$.

Figs. 7a and 7b shows two opposite forces regarding the probability thresholds and the true positive/negative rates. On the one hand, when increasing the probability threshold to 0.7, the true positive rates of the random forest, ridge, and neural networks algorithms drastically decrease. We observe this same behavior for the LASSO when assuming a probability threshold greater than 0.75. The sensitivity of the Logit model remains unchanged even when considering a probability threshold closer to one. On the other hand, the true negative rates remain practically unchanged for all evaluated models, even when increasing the probability threshold to a value closer to one. By observing these

Table 5
Performance of the machine learning algorithms. Null accuracy captures the accuracy by always predicting the most frequent outcome in the database. The score measures how often the classifier is correct. Error denotes the miss-classification errors regarding the predicted cartel probabilities in the total sample. The precision measures how often the prediction of cartels is correct. AUC captures the relationship between the share of true positive predictions against the share of false-positive predictions at various threshold settings.

	Null accuracy (%)	Score (%)	Error (%)	Precision (%)	AUC (%)
Belo Horizonte					
LASSO	55.65	100	0	100	100
Logit	55.65	100	0	100	100
Neural networks	55.65	98.66	1.34	97.10	99.98
Random forest	55.65	98.66	1.34	98.48	99.96
Ridge	55.65	100	0	100	100
Brasília					
LASSO	63.6	96.6	3.4	94.90	98.64
Logit	63.6	98.64	1.36	97.89	99.04
Neural networks	63.6	94.44	5.56	86.67	90.61
Random forest	63.6	97.96	2.04	96.88	99.62
Ridge	63.6	86.39	13.61	87.64	90.42
Caxias do Sul					
LASSO	63.22	92.46	7.54	87.30	98.05
Logit	63.22	100	0	100	100
Neural networks	63.22	93.15	6.85	87.5	98.49
Random forest	63.22	96.58	3.42	98.04	94.62
Ridge	63.22	93.83	6.17	90.91	96.24
São Luís					
LASSO	52.56	97.77	2.23	95.52	100
Logit	52.56	98.51	1.49	96.96	100
Neural networks	52.56	94.07	5.93	91.78	99.38
Random forest	52.56	88.89	11.11	91.53	97.42
Ridge	52.56	97.77	2.23	95.83	97.69
Overall average	58.76	96.22	3.78	94.75	98.01

outcomes for Belo Horizonte, the antitrust agency would choose the Logit classifier – it minimize the false-positive rates (1 - specificity = 0) without increasing the false-negative rate (1 - sensitivity = 0) in the task of detecting cartels.

We conclude the performance of our classifiers by assessing the area under the curve (AUC) metrics. It provides information regarding how well the classifiers are separating the cartel periods from the non-cartel periods. The AUC represents the probability that a classifier will rank a randomly chosen positive observation higher than a randomly chosen negative observation. Thus, the closer the AUC is to 1 (100%), the better the classifier. In Belo Horizonte, Table 5 shows that Logit, LASSO, and ridge have the greater AUC.

5.2. Brasília

Out of a total of 489 weeks, the observations labeled as a cartel period in Brasilia represent 63% of the total sample. As we can see from Fig. 8, the testing sample for evaluating the classifiers contains 147 observations. The Logit classifier has a scores index given by $(52 + 93)/147 = 98.64\%$. The classification error is $(0 + 2)/147 = 1.36\%$. Random forest's score is equal to $(51 + 93)/147 = 97.96\%$, and a classification error of 2.05%.

From Fig. 9a, we see that the Logit, LASSO, and random forest classifiers have a sensitivity rate equal to 1 when the threshold probability is lower or equal to 0.7. After this range, i.e., for values greater than 0.7, the random forest algorithm shows the best performance in terms of the true positive rate, especially when the threshold probability is greater than 0.8. In Fig. 9b, we can see that the specificity of all classifiers shows reasonable outcomes even when increasing the probability threshold. The Logit, random forest, and LASSO classifiers have the best performance concerning the true negative rates.

Logit's precision index is slightly higher $93/(93 + 2) = 97.89\%$ than the one obtained by the random forest $93/(93 + 3) = 96.88\%$. Fig. 9 summarizes the trade-offs in classification accuracy for the gasoline cartel in Brasília. Observe that false-positive results (1− specificity) decreases in the probability threshold. However, the false-negative rate (1− sensitivity) in cartel periods increases much faster in the threshold. To minimize both the false-positive and false-negative rates, the antitrust authority's optimum decision rule threshold lies between 0.5 and 0.7. Within this range, the Logit, LASSO and random forest algorithms have the best performances. However, when the regulator sets a threshold probability greater to 0.8 to predict cartel behavior, the random forest provides the lowest false-negative rate.

Judging by the AUC criterion, the random forest algorithm shows the best performance (AUC = 99.62%). Thus, taking into account all the evaluation metrics as provided by Table 5, we can conclude that random forest, on average, performs subtly better than LASSO and logit classifiers. Therefore, random forest proves to be the best algorithm for classifying the gasoline cartel in Brasília.

5.3. Caxias do Sul

Caxias do Sul has 178 weeks labeled as cartel and 306 weeks labeled as non-cartel. Then, we have 484 observations, from which 30% (146 observations) are used for testing the classifiers. Fig. 10 detail the confusion matrices. Note that the logit classifier provides the best score index $(92 + 54)/146 = 100\%$. The classification error is given by equals to $(0+0)/146 = 0\%$. Fig. 11 shows that when considering a classification threshold ranging from 0.5 to 0.7, the logit still provides the best prediction outcomes. In this range, the true positive and true negative rates are given by $54/(0 + 54) = 1$ and $92/(92 + 0) = 1$. In addition, the precision index of the logit classifier is the largest $54/(0 + 54) = 100\%$.

In addition, evaluating Fig. 11, we can understand how sensitivity and specificity react together to an increase in the probability threshold. By comparing Figs. 11a and 11b, we see that the logit's false-negative rate (1− sensitivity) is equal to zero for the threshold probability lower or equal to 0.7. Besides, the logit's false-positive rate (1 - specificity) remains equal to zero even when considering a threshold probability closer to one. In other words, when we narrow the probability rule, we only affect the logit's sensitivity. Note that if the antitrust authority adopts a probability threshold greater than 0.6, it is not recommended to adopt the random forest classifier to identify the cartel in Caxias do Sul.

With this in mind, the antitrust authority would not adopt the ridge model in Caxias do Sul as well. Notice that the ridge classifier has a high specificity rate. However, it is not a sufficient condition to minimize classification errors. The ridge algorithm has a poor performance regarding the false-negative metrics, which incorrectly classifies observations as belonging to cartel periods. This reasoning is analogous to the neural network algorithm. Thus, the classifier that best responds to the data – indicating how many observations are correctly identified as a cartel period (sensitivity) and how many observations are correctly identified as a non-cartel period (specificity) – is the logit. The AUC criterion showed in Table 5 also allows us to conclude that, on average, the logit (AUC = 100%) has the superior performance in predicting the gasoline cartel in Caxias do Sul.

5.4. São Luís

São Luís has a total sample of 449 weeks. The period labeled as cartel behavior contemplates 47.44% of this amount. We use 135 observations to compare the machine learning classifiers. The models that best classify the cartel in São Luís are LASSO and logit, respectively.

Using the confusion matrices available in Fig. 12, we see that the LASSO's classification error 2.23% is almost five times lower than the error calculated for random forest 11.11%. For the logit, this measure is 1.49%, which is more than seven times lower than the random forest's classification error. Regarding the precision measure, the logit shows

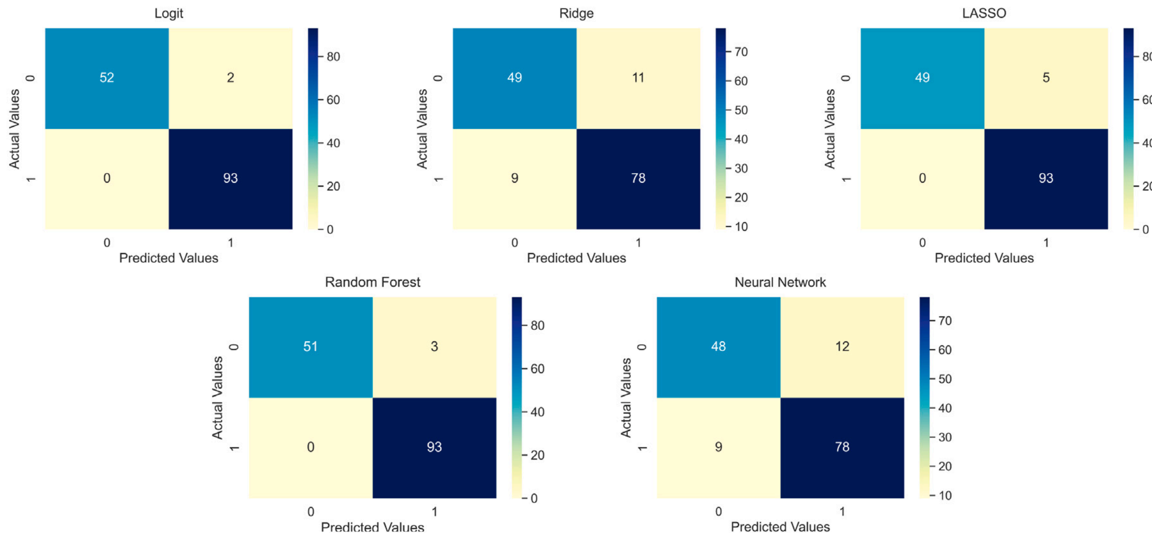
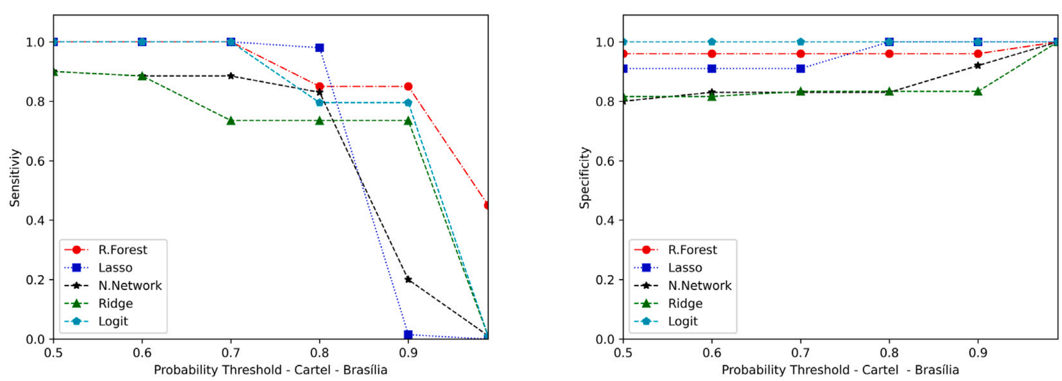


Fig. 8. Confusion matrices - Brasília.



(a) (b)

Fig. 9. Sensitivity and specificity results by restricting the cartel classification rule.

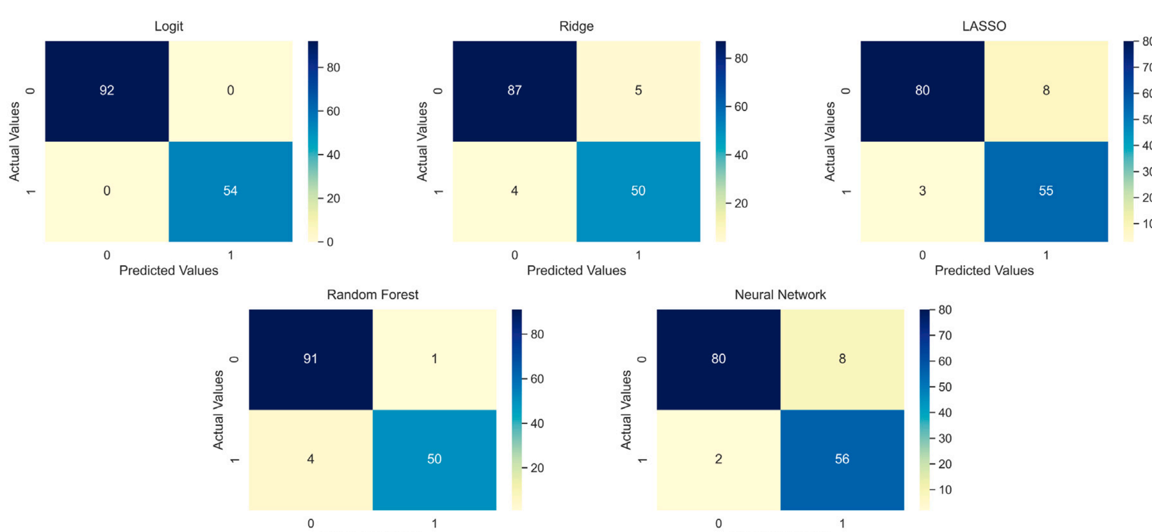
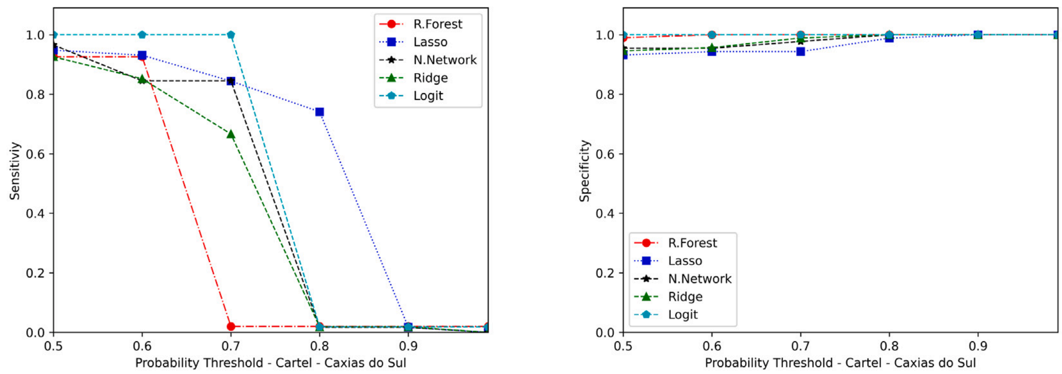


Fig. 10. Confusion matrices - Caxias do Sul.

the best performance concerning the true positives cartel observations 64/(64 + 2) = 96.97%. The models that come closest to this rate are the ridge ((69/69 + 3) = 95.83%), and LASSO ((64/64 + 3) = 95.52%) classifiers.

Considering the probability threshold equals 0.5 and judging by using all the performances metrics, on average, we observe from Fig. 14 that logit, LASSO, and ridge are more accurate. Fig. 13a shows the sensitivity. For a threshold between 0.6 and 0.95, the logit and LASSO



(a) (b)

Fig. 11. Sensitivity and specificity results by restricting the cartel classification rule.

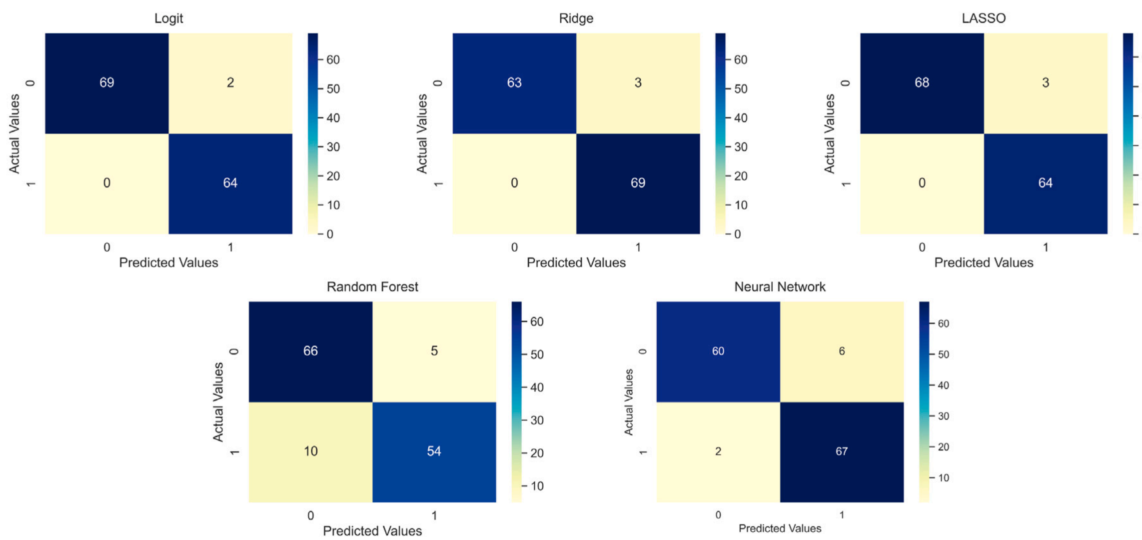


Fig. 12. Confusion matrices - São Luís.

classifiers present better performances, minimizing the false-positive rate (closer to zero). By Fig. 13b, we see that all classification models become equivalents regarding the proportion of actual negatives that are correctly identified when narrowing the probability threshold, i.e., the specificity is closer to one. However, the LASSO classifier shows a slightly lower false-positive rate when the Antitrust authority chooses a probability threshold greater than 0.8. With this in mind, the LASSO algorithm is the best model for the cartel prediction in São Luís.

Table 6 presents the contribution offered by each predictor based on the Permutation Feature Importance technique. It allows us to measure how much the models' score depends on each explanatory variable. Permutation Importance intuitively informs us of each predictor's contribution to the cartel classification task. Strobl et al. (2008) explains that if the observed quality decreasing is small, the information contained in the original classifier (containing all input variables) might not be sufficiently impactful in determining its predictions. In other words, it reveals that the model can still provide good outcomes when dropping some predictors. Furthermore, if the observed decrease is easily observable, the information used by the original classifier impacts its predictions.[20]

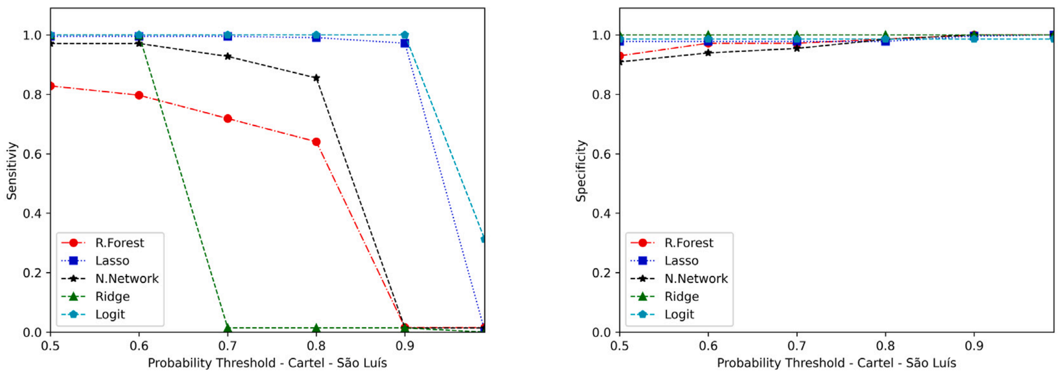
Via the Permutation Importance criterion, the Standard Deviation and Coefficient of Variation stand out for all evaluated cities. The Spread can contribute to cartel detection in Belo Horizonte, Brasília, and Caxias do Sul. By turn, the skewness is relevant for cartel inspection in Belo Horizonte, Brasília and São Luís. The kurtosis can help to identify collusive behavior only in Caxias do Sul.

6. Antitrust authorities and competition policy

Although reduced price variance across time is a collusive indicator, we still need to differentiate tacit collusion from organized cartels, especially considering that Competition Authorities have limited resources for investigation. Besides, similar prices can happen without a formal coordination mechanism in a small search costs situation. It may be the case in many regions — not being restricted to the Brazilian market, in which retailers may belong to a similar ownership structure. Then we believe that the availability of a dataset combined with qualitative information on cartel formation provides a valuable contribution to the discussion we presented here so far. In a machine learning setting, it allows policymakers and competition authorities to foresee cartel behavior in advance.

In addition, there are several questions to be addressed in order to increase the attractiveness of our method. The first one is whether our machine learning models are robust enough to be used in other

[20] See Radivojac et al. (2004) and Altmann et al. (2010) for details on the Permutation Importance implementation and interpretability.



(a) (b)

Fig. 13. Sensitivity and specificity results by restricting the cartel classification rule.

Table 6
Permutation importance.

Logit (Belo Horizonte)			Random forest Brasília		
Predictor	Weight		Predictor	Weight	
Standard deviation	0.4416	±0.0559	Standard deviation	0.4014	±0.0243
Coefficient of variation	0.1664	±0.0197	Coefficient of variation	0.0177	±0.0163
Spread	0.0268	±0.019	Skewness	0.0095	±0.0067
Skewness	0.0067	±0.0085	Spread	0.0054	±0.0102
Kurtosis	0	0.0000	Kurtosis	0	0.0000
Logit (Caxias do Sul)			LASSO São Luís		
Predictor	Weight		Predictor	Weight	
Standard deviation	0.6014	±0.053	Standard deviation	0.4948	±0.0672
Coefficient of variation	0.2466	±0.0594	Coefficient of variation	0.3481	±0.0562
Spread	0.0178	±0.0186	Skewness	0.003	±0.0119
Kurtosis	0.0082	±0.0055	Spread	0	0
Skewness	0	0.0000	Kurtosis	−0.0015	±0.0145

industries or even in other retail markets. The statistical screen approach is expected to have a good performance, even in other sectors or countries, where price dynamics may vary from those considered in this proposal.

Besides, our approach reveals some adjustability once we can create many different inputs as cartel screening predictors. Then, it can better capture some of the sensitivities conditioned to the different characteristics of markets and potentially cover different collusive price patterns. In contrast with other detection methods, especially those that require data on cost variables to detect bid-rigging cartels (Bajari and Ye, 2003).

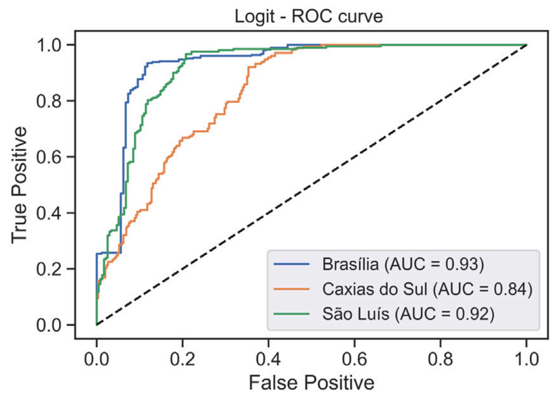
The legal and economic consistency of the cartel prosecution is a challenging objective for the competition authority. Price distribution screen-based may overcome some drawbacks of the traditional econometric approach (Huber and Imhof, 2019). Simple screens are not as time-consuming as the structural econometric methods that demand non-observable variables such as costs and produce many false-negative results when applied in real cases (Bajari and Ye, 2003). Besides, classification errors generate a very high opportunity cost for the regulator and substantially damaging their reputation (Abrantes-Metz, 2012).

We reinforce our method's attractiveness to raise the quality of the policymakers' decisions. Machine learning algorithms can easily adapt to many different situations. Then, it opens an avenue to consider little exploited variables in retail market analysis, such as the third and fourth statistical moments of the gasoline price distribution. By evaluating price dynamics, the regulator can map market behaviors that are harmful to competition and consumer welfare. In this way, the combined usage of machine learning techniques with statistical screening is promising. Mainly in the prescription of competition policies in the most varied economic sectors.

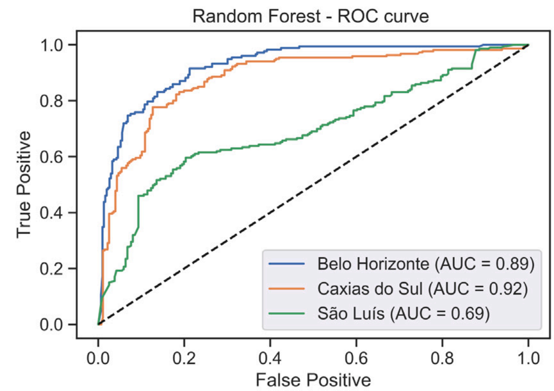
Concerning our case study, we recommend our screens by adopting some practices, as follows. First, the coefficient of variation and standard deviation reveals to be the most powerful predictors. In this way, variance-based predictors help us to infer the negative relationship between the retail gasoline price volatility and the cartel probability, as shown in Abrantes-Metz et al. (2006).

On the other hand, in some contexts, both skewness (asymmetry) and kurtosis reveal to be relevant in the correct prediction of cartel probability. Spread is also determinant in some circumstances. Thus, we can see the relevance of all statistical moments. Ultimately, we have a range of predictors that can act both in a complementary and substituting manner, increasing the contribution derived from the economic evidence on the cartel formation. Finally, regarding the trade-offs in reducing false-positive *vs.* false-negative outcomes, an appropriate strategy would be to increase the probability threshold between 0.6 and 0.75. Within this range, most of the classifiers still provide both high sensitivity and specificity. Consequently, this strategy might reduce incorrect predictions among truly non-cartel periods (false-positives) at the expense of increasing the number of actual cartel periods (false-negatives) in all evaluated cities.

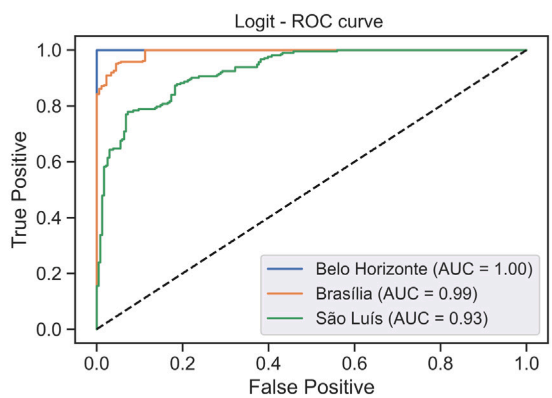
We understand that discussing the performance of proposed algorithms in new databases (ex-ante screening) can increase the antitrust authority's capacity in solving prediction policy problems. We contribute to this aim by selecting the best trained and tested models for each city, as presented in Section 5. Next, we test the performance of these algorithms when predicting anticompetitive patterns in the other cities analyzed. To assess the true-positive and false positive rates of each model, we use the ROC (Receiver Operating Characteristic) curve. It summarizes and compares the performance of the classifiers based on the positive class (cartel). As we show in Fig. 14, the axes indicate the



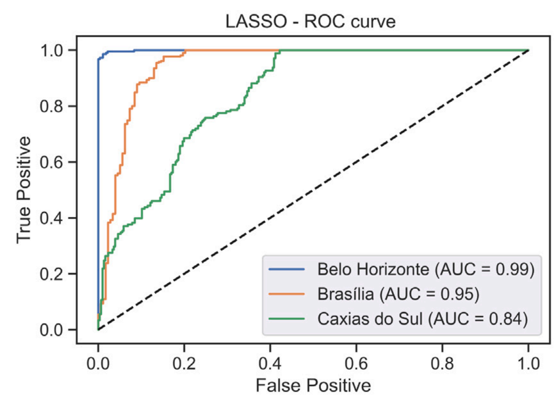
(a) Belo Horizonte



(b) Brasília



(c) Caxias do Sul



(d) São Luís

Fig. 14. The ROC curve for evaluating algorithms' performance in new datasets (ex-ante screening).

false-positive and true-positive rates, respectively, which are given by:

$$\text{True Positives Rate} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{False Positives Rate} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

Bradley (1997) explains the graph as the fraction of correct predictions for the positive class (cartel, on the y axis) versus the fraction of errors for the negative class (non-cartel, on the x axis). Ideally, we would expect the fraction of correct positive class predictions to be one and the portion of incorrect negative class predictions to be zero. It reveals that a perfect classifier would occupy the entire upper-left corner of the graph, i.e., the coordinate (0,1). As shown in the sensitivity and specificity analysis, there is a trade-off between the true-positive and false-positive rates, so that changes made to the threshold probability interfere in the accuracy of the predictions. Namely, it is possible to improve the True Positive rate at the expense of the false positive rate (and vice versa). In summary, the ROC curve allows us to evaluate the true positives and false positives for different threshold probabilities. A classifier that does not have discriminative power between classes will form a diagonal (dashed in black) line from coordinate (0,0) to coordinate (1,1). In other words, the classifier would predict all observations as belonging to the negative or positive class. Models represented below this line have little or no ability to label data.

Fig. 14a shows the ex-ante predictive power of the Logit classifier trained with the Belo Horizonte dataset. The area under the curve (AUC) for Brasília is slightly larger than that obtained for São Luís. The Logit adjusted for Belo Horizonte has a lower performance in detecting cartels in Caxias do Sul. However, the AUC is satisfactory. Fig. 14b depicts the ex-ante predictive power of the random forest classifier trained with the Brasília database. The AUC for Belo Horizonte is slightly larger than that obtained for Caxias do Sul. The lower performance of the random forest as an ex-ante screening is observed when one intends to detect cartels in São Luís.

Fig. 14c illustrates the ex-ante predictive power of the Logit classifier trained with data from Caxias do Sul. The AUC for Belo Horizonte indicates that the model can perfectly distinguish the cartel and non-cartel period in Belo Horizonte. We also see a high performance for detecting cartels in Brasília. Although smaller, the performance of the logit classifier as an ex-ante screening to detect cartels in São Luís is satisfactory as well. Fig. 14d provides the ex-ante predictive power of the LASSO classifier trained with data from São Luís. The area under the curve for Belo Horizonte is equal to 0.9992 (99.92%). It indicates that the model can almost perfectly distinguish the cartel and non-cartel period in Belo Horizonte. LASSO trained with the São Luís dataset is also satisfactory for detecting cartels in Brasília and Caxias do Sul.

Table 7 shows the algorithm's performance in new datasets (ex-ante screening). The logit classifier trained with the Belo Horizonte database shows a higher score when detecting cartel behavior in Caxias do Sul (74%). Concerning the precision, it is higher in São Luís (74%). The AUC criteria is greater when the logit classifier is tested in the Brasília database (93.3%). When it comes to the random forest classifier trained with the Brasilia data, we observe a high score rate in Belo Horizonte (79%).

The predictions based on the Caxias do Sul datasets achieves higher precision (83%) and AUC (92.6%). In the ex-ante screening exercise applied to the logit classifier trained with the Caxias do Sul database, the score, precision, and AUC allow us to conclude that it can perfectly predict the cartel behavior in Belo Horizonte. Lastly, based on the score (96%), precision (96%), and AUC rates (99.92%), the LASSO classifier trained with the São Luís data almost perfectly detects the anti-competitive pattern found in Belo Horizonte.

Table 7
Algorithms' performance in new datasets (ex-ante screening).

	Score (%)	Error (%)	Precision (%)	AUC (%)
Belo Horizonte (Logit)				
Brasília	52	48	72	93.3
Caxias do Sul	74	26	73	83.63
São Luíss	68	32	74	91.68
Brasília(Random forest)				
Belo Horizonte	79	21	81	88.13
Caxias do Sul	70	30	83	92.6
São Luíss	59	41	60	69.45
Caxias do Sul (Logit)				
Belo Horizonte	100	0	100	100
Brasília	74	26	85	99.21
São Luís	84	16	85	92.79
São Luís (LASSO)				
Belo Horizonte	96	4	96	99.92
Brasília	54	46	77	94.83
Caxias do Sul	75	25	77	84.07
Overall average	73.75	26.25	80.25	90.80

7. Conclusion

In this paper, we have combined supervised machine learning techniques with statistical screens based on the gasoline retail price distribution to predict collusion. The logit classifier has shown an average overall score of 99.29% correct classifications – considering cartel and non-cartel periods. Even increasing the probability threshold, the logit algorithm remains the most stable classifier regarding sensitivity and specificity. Lasso (96.71%) and random forest (95.52%) also can be valuable classifiers for the antitrust authority when used in an ex-ante screening fashion, i.e., using new datasets to predict cartels in the gasoline market.

We also have found evidence that asymmetry, kurtosis, and spread are features that might increase the algorithms' performance. These inputs work in a complementary way with the standard deviation and coefficient of variation in the cartel prediction. Thus, we empirically reinforce the intuition by relying upon strong assumptions of the traditional econometric screening methods. In other words, the supervised machine learning classifiers evaluated in this paper show us that a structural relationship between a given screen and the probability of collusion does not assure high predictive power. Therefore, as discussed in Section 6, the regulator can take valuable information about the cartel mechanisms by assessing some descriptive statistics on pricing patterns and combining them with classifier algorithms. Typically, machine learning techniques are not as time-consuming as traditional econometric screening approaches. The competition authority needs effective monitoring and often anticipating cartel movements. On that matter, our work has shown that supervised machine learning classifiers have many positive attributes and can provide valuable contributions in detecting and fighting cartels. The need for a labeled database can impose limitations on the use of these algorithms and pose challenges and pitfalls, especially concerning the costs and damage to the antitrust authorities' reputation, inherent in the trade-off between reducing false positives *vs.* false negatives.

A concurrent approach to detect cartels is to compute the retailers' profit margin based on retail and wholesale prices. In future work, we may consider the coefficient of variation between retailer and distribution prices, as pointed by Ragazzo et al. (2012). Another extension would be to test whether price cycles and margin volatility can distinguish different price patterns, as evaluated by Atkinson et al. (2014). It can raise the power of the machine learning algorithms in predicting collusive agreements and improve our knowledge of a firm's pricing behavior. A fruitful avenue for future research would involve assessing the role of the dual-fuel system in Brazil, which includes ethanol and gasoline, in terms of the construction of cartel screens. Finally, a promising research proposal would be to develop a proactive approach, letting the algorithm itself decides which data properties' are relevant for detecting cartels. It would be possible by extending the Convolutional Neural Network (CNN) architecture proposed by Huber and Imhof (2021) to big data analytics for fighting cartels in gasoline market

CRediT authorship contribution statement

Douglas Silveira: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Data curation, Software. **Silvinha Vasconcelos:** Conceptualization, Supervision, Writing – review & editing. **Marcelo Resende:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision. **Daniel O. Cajueiro:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Acknowledgments

The authors thank Professors David Brown, Andrew Eckert, and Emilson C.D. Silva, from the University of Alberta, for their thoughtful suggestions and comments. We acknowledge the recommendations made by the two anonymous referees that were crucial to improve our manuscript. Daniel O. Cajueiro is indebted to CNPQ, Brazil for partial financial support under grant 302629/2019-0.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eneco.2021.105711.

References

Abrantes-Metz, R.M., 2012. Screens for conspiracies and their multiple applications. Compet. Policy Int. 8 (1).
Abrantes-Metz, R.M., Froeb, L.M., Geweke, J., Taylor, C.T., 2006. A variance screen for collusion. Int. J. Ind. Organ. 24 (3), 467–486.
Akouemo, H.N., Povinelli, R.J., 2016. Probabilistic anomaly detection in natural gas time series data. Int. J. Forecast. 32 (3), 948–956.
Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics 26 (10), 1340–1347.
Amann, E., Baer, W., Coes, D., 2011. Energy, Bio Fuels and Development: Comparing Brazil and the United States, Vol. 87. Taylor & Francis.
ANP, 2020. Diagnóstico da concorrência na distribuição e revenda de combustíveis automotivos. http://https://wwo.br/anp/pt-br/centrais-de-conteudo/publicacoes/livros-e-revistas/arquivos/diagnostico-sdc-2020.pdf. [Online; accessed 08-January-2021].
Athey, S., Bagwell, K., Sanchirico, C., 2004. Collusion and price rigidity. Rev. Econom. Stud. 71 (2), 317–349.
Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. Annu. Rev. Econ. 11, 685–725.
Atkinson, B., Eckert, A., West, D.S., 2014. Daily price cycles and constant margins: Recent events in Canadian gasoline retailing. Energy J. 47–69.
Bajari, P., Ye, L., 2003. Deciding between competition and collusion. Rev. Econ. Stat. 85 (4), 971–989.
Blair, R.D., Sokol, D.D., 2015. The Oxford Handbook of International Antitrust Economics, Vol. 2. Oxford University Press, USA.
Blanckenburg, K., Geist, A., Kholodilin, K.A., 2012. The influence of collusion on price changes: New evidence from major cartel cases. Ger. Econ. Rev. 13 (3), 245–256.
Bolotova, Y., Connor, J.M., Miller, D.J., 2008. The impact of collusion on price behavior: Empirical results from two recent cases. Int. J. Ind. Organ. 26 (6), 1290–1307.
Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 30 (7), 1145–1159.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.
Breiman, L., 2017. Classification and Regression Trees. Routledge.
CADE, 2014. Cadernos do cade - varejo de gasolina.
Chouinard, H.H., Perloff, J.M., 2007. Gasoline price differences: Taxes, pollution regulations, mergers, market power, and market conditions. BE J. Econ. Anal. Policy 7 (1).
Clark, R., Houde, J.-F., 2013. Collusion with asymmetric retailers: Evidence from a gasoline price-fixing case. Am. Econ. J. Microecon. 5 (3), 97–123.

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

Connor, J.M., 2005. Collusion and price dispersion. Appl. Econ. Lett. 12 (6), 335–338.

Contín-Pilart, I., Correljé, A.F., Palacios, M.B., 2009. Competition, regulation, and pricing behaviour in the Spanish retail gasoline market. Energy Policy 37 (1), 219–228.

Crede, C.J., 2019. A structural break cartel screen for dating and detecting collusion. Rev. Ind. Organ. 54 (3), 543–574.

Cuiabano, S., 2019. Competition policy evaluation through damage estimation in fuel retail cartel in londrina, Brazil [Avaliação de política de concorrência: Estimação de danos no cartel de postos de londrina]. Rev. Econ. Contemp. 23 (2).

Da Silva, A.S., Vasconcelos, C.R.F., Vasconcelos, S.P., de Mattos, R.S., 2014. Symmetric transmission of prices in the retail gasoline market in Brazil. Energy Econ. 43, 11–21.

Doane, M.J., Froeb, L.M., Sibley, D.S., Pinto, B.P., 2015. Screening for collusion as a problem of inference. In: Oxford Handbook of International Antitrust Economics, Vol. 2. pp. 523–553.

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

Eckert, A., West, D.S., 2004. Retail gasoline price cycles across spatially dispersed gasoline stations. J. Law Econ. 47 (1), 245–273.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27 (8), 861–874.

Freitas, T.A.d., Balbinotto Neto, G., 2011. Análise cross-section da dispersão dos preços para sinalização de práticas anticompetitivas no mercado de combustíveis.

Friederiszick, H.W., Maier-Rigaud, F.P., 2008. Triggering inspections ex officio: Moving beyond a passive EU cartel policy. Joclec 4 (1), 89–113.

García, J.P., 2010. Dynamic pricing in the Spanish gasoline market: A tacit collusion equilibrium. Energy Policy 38 (4), 1931–1937.

García Rodríguez, M.J., Rodríguez Montequín, V., Ortega Fernández, F., Villanueva Balsera, J.M., 2020. Bidders recommender for public procurement auctions using machine learning: Data analysis, algorithm, and case study with tenders from Spain. Complexity 2020.

Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: A review. Energy Econ. 81, 709–727.

Harrington, J.E., 2005. Detecting Cartels. Technical Report Working Papers, The Johns Hopkins University, Department of Economics.

Harrington, J.E., 2008a. Detecting cartels. In: Handbook of Antitrust Economics, Vol. 213. pp. 215–258.

Harrington, J.E., 2008b. Optimal corporate leniency programs. J. Ind. Econ. 56 (2), 215–246.

Harrington, J.E., Chen, J., 2006. Cartel pricing dynamics with cost variability and endogenous buyer detection. Int. J. Ind. Organ. 24 (6), 1185–1212.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

Hastie, T., Tibshirani, R., Wainwright, M., 2019. Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC.

Ho, T.K., 1995. Random decision forests. In: Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on, Vol. 1. IEEE, pp. 278–282.

Huber, M., Imhof, D., 2019. Machine learning with screens for detecting bid-rigging cartels. Int. J. Ind. Organ. 65, 277–301.

Huber, M., Imhof, D., 2021. Deep learning for detecting bid rigging: Flagging cartel participants based on convolutional neural networks. arXiv preprint arXiv:2104.11142.

Imhof, D., 2017. Simple Statistical Screens to Detect Bid Rigging. Technical Report, Université de Fribourg.

Imhof, D., Wallimann, H., 2021. Detecting bid-rigging coalitions in different countries and auction formats. arXiv preprint arXiv:2105.00337.

Jiménez, J.L., Perdiguero, J., 2012. Does rigidity of prices hide collusion? Rev. Ind. Organ. 41 (3), 223–248.

Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Prediction policy problems. Amer. Econ. Rev. 105 (5), 491–495.

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

Merenstein, S.G., 2016. Essays on Cartels and Market Distortions (Ph.D. thesis).

Motta, L.V., Resende, G., 2019. Mensurando os Benefícios de Combate a Cartéis: O Caso do Cartel dos Combustíveis no Distrito Federal. Documento de Trabalho N° 004/2019, Departamento de Estudos Econômicos - Conselho Administrativo de Defesa Econômica.

Nascimento Filho, A., Pereira, E., Ferreira, P., Murari, T., Moret, M., 2018. Cross-correlation analysis on Brazilian gasoline retail market. Physica A 508, 550–557.

Noel, M.D., 2007. Edgeworth price cycles, cost-based pricing, and sticky pricing in retail gasoline markets. Rev. Econ. Stat. 89 (2), 324–334.

Pedra, D.P., de Oliveira Bicalho, L.M.N., de Araújo Vilela, O., Baran, P.H., de Paiva, R.M., de Melo, T.P., 2010. Metodologia adotada pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis para a detecção de Cartéis. ANP, Rio de Janeiro.

Perdiguero, J., 2010. Symmetric or Asymmetric Gasoline Prices? A Meta-Analysis Approach. IREA–Working Papers, 2010, IR10/13, Universitat de Barcelona. Institut de Recerca en Economia Aplicada Regional i Pública.

Perdiguero, J., Jiménez, J.L., 2020. Price coordination in the Spanish oil market: the monday effect. Energy Policy 112016.

Pereira, J.M., Basto, M., da Silva, A.F., 2016. The logistic lasso and ridge regression in predicting corporate failure. Procedia Econ. Finance 39, 634–641.

Pinha, L.C., Braga, M.J., et al., 2019. Evaluating the effectiveness of the Brazilian Leniency Program. Econ. Bull. 39 (3), 1860–1869.

Porter, R.H., 2005. Detecting collusion. Rev. Ind. Organ. 26 (2), 147–167.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

Rabuzin, K., Modrusan, N., 2019. Prediction of public procurement corruption indices using machine learning methods. In: KMIS. pp. 333–340.

Radivojac, P., Obradovic, Z., Dunker, A.K., Vucetic, S., 2004. Feature selection filters based on the permutation test. In: European Conference on Machine Learning. Springer, pp. 334–346.

Ragazzo, C., et al., 2012. Screens in the gas retail market: The brazilian experience. Antitrust Chron. 3.

Ramalho, M.H.M., Ribeiro, E.P., 2019. Ex-post evaluation of mean-variance cartel filters. In: II Rio International Workshop on Advances in Competition Policy Analysis. December 5, 2019. Brazil.

Resende, M., 2012. Sincronização de preços no mercado de varejo para combustíveis: Uma nota empírica. Rev. Bras. Energ. 18, 197–204.

Sammut, C., Webb, G.I., 2011. Encyclopedia of Machine Learning. Springer Science & Business Media.

Silveira, D., Vasconcelos, S., Bogossian, P., Neto, J., 2021. Cartel screening in the Brazilian fuel retail market. EconomiA 22 (1), 53–70.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9 (1), 1–11.

Wallimann, H., Imhof, D., Huber, M., 2020. A machine learning approach for flagging incomplete bid-rigging cartels. arXiv preprint arXiv:2004.05629.

Wang, Z., 2009. (Mixed) strategy in oligopoly pricing: Evidence from gasoline price cycles before and under a timing regulation. J. Polit. Econ. 117 (6), 987–1030.

Wilmot, N.A., Mason, C.F., 2013. Jump processes in the market for crude oil. Energy J. 34 (1).

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