CESifo Working Papers

8835
2021

January 2021

Won’t Get Fooled Again: A Supervised Machine Learning Approach for Screening Gasoline Cartels

Douglas Silveira, Silvinha Vasconcelos, Marcelo Resende, Daniel O. Cajueiro

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8835*

Won’t Get Fooled Again: A Supervised Machine Learning Approach for Screening Gasoline Cartels

Abstract

In this article, we combine machine learning techniques with statistical moments of the gasoline price distribution. By doing so, we aim to detect and predict cartels in the Brazilian retail market. In addition to the traditional variance screen, we evaluate how the standard deviation, coefficient of variation, skewness, and kurtosis can be useful features in identifying anti-competitive market behavior. We complement our discussion with the so-called confusion matrix and discuss the trade-offs related to false-positive and false-negative predictions. Our results show that in some cases, false-negative outcomes critically increase when the main objective is to minimize false-positive predictions. We offer a discussion regarding the pros and cons of our approach for antitrust authorities aiming at detecting and avoiding gasoline cartels.

JEL-Codes: C210, C450, C520, K400, L400, L410.

Keywords: cartel screens, price dynamics, fuel retail market, machine learning.

*Douglas Silveira**
Department of Economics
University of Alberta / Canada
dsilveir@ualberta.ca

Silvinha Vasconcelos
Department of Economics
Federal University of Juiz de Fora / Brazil

Marcelo Resende
Institute of Economics
Federal University of Rio de Janeiro / Brazil
mresende@ie.ufrj.br

Daniel O. Cajueiro
Department of Economics
University of Brasilia / Brazil
danielcajueiro@unb.br

*corresponding author

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 & Maier-Rigaud, 2008).

There is a wide variety of cartel screens that offers the regulator practical and efficient detection methods. (Eckert & West, 2004; Bolotova et al., 2008; Blanckenburg et al., 2012; Harrington, 2008; 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 disclosures 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 & Perloff, 2007; Noel, 2007; Abrantes-Metz, 2012; Jiménez & Perdiguero, 2012; Eckert, 2013; Perdiguero & Jiménez, 2020). However, there is no universal consensus on this issue.

In order to contribute to the discussion on cartel detection and prediction, this paper uses machine learning techniques with statistical screens. Huber & Imhof (2019) and Imhof (2017) use 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 algorithms as screens.

Taking the Brazilian market as a case study, we evaluate the out of the sample performance of the proposed methods in a total of 1.920 observations constructed from a weekly database of gasoline selling price in the following cities where collusion was detected: Belo Horizonte[1], 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)[2] 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 & Resende, 2019).

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

[1]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.

[2]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).

and the operations to disrupt the gasoline cartels.

With this in mind, we extend the approach proposed by Huber & Imhof (2019) and Huber & Imhof (2018) to analyze cartel behavior in the gasoline market. Our framework evaluates four supervised machine learning models: Random Forest, Lasso (least absolute shrinkage and selection operator) Logistic, Ridge Logistic, and Neural Network. 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 & 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 was 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 machine learning techniques are powerful tools for cartel detection in the gasoline market. Furthermore, it demonstrates that in specific cases, the 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.

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.

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, 2008). 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 & Sokol (2015) provide many real-world examples of cartel screening methods for detecting collusion in retail markets. As reported in Harrington (2005) and Zitzewitz (2012), economists widely apply this dynamic pricing methodology to generate collusive patterns. Then, the goal is to distinguish a competitive pricing pattern from that observed in cartel agreements (Maskin & Tirole, 1988).

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 & 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 & 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 in the market under analysis, 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 policy-makers (Wilmot & 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) and (Harrington & Chen, 2006). On this subject, many contributions come from the analysis of the Spanish gasoline retail market. Jiménez & 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 & 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.

However, the vast majority of studies cited above use econometric techniques rather than supervised machine learning algorithms. Besides, to the best of our knowledge, there are very few papers systematically investigating the screening performance based on patterns derived from all four statistical moments via computational methods, especially in the gasoline market. In this sense, our research dialogues with the incipient literature on implementing screens to detect bid-rigging cartels (Huber & Imhof, 2019).

3. Database and Gasoline Industry in Brazil

The fuel supply in Brazil is made by oil companies, refineries, distributors, and retailers. 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[3] and SBDC − Brazilian Competition Defense System. The role of ANP is to regulate products and firms and provide the SBDC

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

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[4], 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[5], it is worth remembering that the gas stations' market behavior

[4]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).

[5]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.

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 local market power to the fuel retailers. This contributes even more to anti-competitive actions in the sector (CADE, 2014). With this in mind, 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 & 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 and collected weekly. 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 summarizes the number of the cartel and non-cartel observations. The first case we evaluate happened in the metropolitan region of Belo Horizonte, including the neighboring municipalities of Betim and Contagem.[6] As described in the administrative procedure[7] started in 2014, anonymous complaints date back to the early 2000s. The hard evidence was collected by the antitrust authority 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 the period between January 2014 and April 2019 as the non-cartel period.

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 gathered, involving distributors and resellers.[8] 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.

In Caxias do Sul[9], the antitrust agency confirmed the evidence that fuel distributors had organized a cartel to fix and standardize prices practiced in fuel resale.

[6] 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.

[7] 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.

[8] 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`.

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

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, causing immense losses to final consumers. 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.

	Cartel Obs	**Perc. (%)**	**Non-Cartel Obs**	**Perc. (%)**	**Total**
Belo Horizonte	221	45	276	55	497
Date	**01/2004 - 04/2008**		**01/2014 - 04/2019**		
Brasília	309	63	179	37	488
Date	**11/2009 - 11/2015**		**12/2015 - 04/2019**		
Caxias do Sul	178	37	306	63	484
Date	**01/2004 - 07/2007**		**03/2013 - 04/2019**		
São Luís	215	48	236	52	451
Date	**01/2010 - 10/2014**		**11/2014 - 04/2019**		
Total	702	48,25	721	51,75	1920

Table 1: Number of Cartel and Non-Cartel observations.

	Obs	Mean	Std. Dev.	Min	Max		Obs	Mean	Std. Dev.	Min	Max
Belo Horizonte						**Caxias do Sul**					
Cartel Periods						**Cartel Periods**					
Standard deviation	221	0.0748018	0.0161794	0.041401	0.11666	Standard deviation	178	0.0284347	0.0091526	0.0125238	0.0866679
Variance	221	0.0058559	0.0025664	0.001714	0.01360	Variance	178	0.0008918	0.0007337	0.0001568	0.0075113
Skewness	221	0.7554747	0.7215891	-1.9099710	3.35897	Skewness	178	-0.7019769	1.125306	-4.499262	3.311451
Kurtosis	221	4.7893610	4.6096630	1.0000000	24.96786	Kurtosis	178	5.136701	3.351867	1.387655	25.00011
Coefficient of Variation	221	0.0340085	0.0070236	0.020943	0.05835	Coefficient of Variation	178	0.0109407	0.0081338	0.0038722	0.0687003
Non-Cartel Periods						**Non-Cartel Periods**					
Standard deviation	276	0.1111709	0.0215818	0.0591717	0.179358	Standard deviation	306	0.0679517	0.0335223	0.018549	0.2813645
Variance	276	0.0128231	0.0051184	0.0035013	0.0321694	Variance	306	0.0057375	0.0078758	0.0003441	0.079166
Skewness	276	0.7070966	0.5847417	-1.978601	2.5789970	Skewness	306	-0.6077758	1.111237	-3.228627	0.079166
Kurtosis	276	3.8759010	1.6697230	2.0357790	14.300410	Kurtosis	306	4.208592	2.292335	1.307984	12.00068
Coefficient of Variation	276	0.0307957	0.0061329	0.016822	0.0491287	Coefficient of Variation	306	0.0181618	0.0081338	0.0038722	0.0687003
Brasília						**São Luís**					
Cartel Periods						**Cartel Periods**					
Standard deviation	309	0.0156584	0.0159	0.0000000	0.0805304	Standard deviation	215	0.0486387	0.0311685	0.0037796	0.1431614
Variance	309	0.0004972	0.0012938	0.0000000	0.0064851	Variance	215	0.0033327	0.0041652	0.0000143	0.0204952
Skewness	309	-0.7774293	2.370778	-7.862468	6.802973	Skewness	215	1.112635	1.505716	-3.749028	4.110874
Kurtosis	309	10.5154700	12.19631	1.053223	64.64728	Kurtosis	215	7.13466	4.259128	1.000000	24.14746
Coefficient of Variation	309	0.0054248	0.005926	0.0000000	0.0294542	Coefficient of Variation	215	0.0189299	0.012681	0.0013737	0.0599945
Non-Cartel Periods						**Non-Cartel Periods**					
Standard deviation	179	0.1063629	0.0489471	0.0105688	0.2383494	Standard deviation	236	0.072602	0.0255238	0.0107529	0.138873
Variance	179	0.0136954	0.0107811	0.0001117	0.0568104	Variance	236	0.0059197	0.0040704	0.0001156	0.0192857
Skewness	179	-0.089362	1.466399	-4.033528	3.740933	Skewness	236	0.2761049	1.108964	-3.285877	4.698436
Kurtosis	179	5.0314610	4.233466	1.416789	20.87988	Kurtosis	236	4.503887	3.429616	1.092339	26.91653
Coefficient of Variation	179	0.0267536	0.01313	0.0000000	0.0731427	Coefficient of Variation	236	0.0205283	0.0075034	0.0026962	0.0393634

Table 2: Descriptive Statistics of the four evaluated cities. As already reported, the Belo Horizonte cartel involved the neighboring municipalities of Betim and Contagem, which were duly incorporated into our analysis.

In São Luís[10], intercepted conversations revealed that the owners of gas stations

[10] 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/`.

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 2010 and October 2014. 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 station owners in the municipality. Besides, the investigation conducted by the Maranhão Public Prosecutor's Office pointed to a market division among the cartel's participants, to facilitate the operationalization of the illegal agreement, under the coordination of the union. It was also found at the union headquarters a map dividing the city into "corridors", where the same price was established.

3.3. Statistical screens

To check whether a market is more likely to practice cartel, 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, variance, asymmetry, and kurtosis. Note that all 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. From the weekly price dynamics, we calculate each of the predictors (inputs) detailed described in equations (1) - (4). The normalized moments calculated in Table 2 allow us to compare the shape of different probability distributions across the cartel and the non-cartel periods. Then, for each city, we report the Kolmogorov Smirnov and Mann Whitney tests for the predictors derived from the price distribution for the cartel and the non-cartel periods in Table 3.

3.3.1. Standard Deviation & 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 also evaluate the coefficient of variation defined as follows as a statistical screen:

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

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.

3.3.2. Variance

We also consider the variance $\sigma_{c,w}$ of the weekly gasoline selling price within a given city as a screen for detecting cartels. There are theoretical justifications for a variance screen for collusion if it is costly to coordinate price changes or if the cartel must solve an agency problem. There is also some empirical evidence of a decrease in the variance of price during collusion (Abrantes-Metz et al., 2006).

$$s^2_{c,w} = \frac{\sum_{i=1}^{n}(P_{c,w} - \bar{m}_{c,w})^2}{n-1}. \tag{2}$$

3.3.3. Skewness

Price manipulation may affect the symmetry of the distribution of the weekly gasoline selling price. Thus, for a sample of size n, the methods of moments estimator of the skewness yields:

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

where $m_{3c,w}$ is the sample third central moment of the weekly retail gasoline price within a given city c.

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:

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

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

3.4. Descriptive statistics

We now evaluate the descriptive statistics by separating them between cartel periods and "non-cartel" periods in each evaluated city. Note from Table 2 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 variance, skewness, and kurtosis. Furthermore, in some cities, the difference between the statistical moments is quite noticeable. Typically, during cartel periods, it is common to see less variance in price distribution. Besides, we assess the expected pattern concerning the other statistical moments on a case-by-case basis, as follows.

In Belo Horizonte, the mean of the standard deviation screen is approximately 70% lower during the cartel. Thus, prices are more similar in collusive than in competitive periods. This same intuition fits on the standard deviation screen. The variance is 45% lower in cartel 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 in cartel periods than in non-cartel periods, suggesting that prices converge when there is a cartel in the retail gasoline market. 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 0.005, compared to 0.013 for non-cartel periods. The variance reveals to be almost double in non-cartel periods. This behavior provides shreds of evidence that prices are more similar in cartels. The price distribution is highly asymmetric in cartel periods. The kurtosis amounts to 5.0314 in non-cartel periods and more than doubles in cartel periods (10.515).

When compared with Brasília, we observed some similarities concerning the coefficient of variation and variance patterns in Caxias do Sul. During the cartel period, these screens show a much lower variation than that observed in the non-cartel period, matching with the cartel practice. Besides, the price distribution is more asymmetric in collusive periods. Although the behavior of the retail price of gasoline in Caxias do Sul is not as diverse as that observed in Brasília, it is almost 22% greater during the cartel period. In São Luís, the coefficient of variation and variance is slightly low during the cartel periods. The spread of the skewness is higher in cartel periods, with a standard deviation of 1.5057. During the non-cartel period, the standard deviation is equal to 1.1089. The mean of the kurtosis amounts to 7.1346 in cartel periods. It is almost 60% higher than the non-cartel periods (4.5038).

In Table 3, 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 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 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.

Screens	**z-statistic**	**p-value MW**	**Ksa**	**p-value KS**
Belo Horizonte				
Standard deviation	15.996	$<$ .0001	0.7022	$<$ .0001
Variance	15.996	$<$ .0001	0.7022	$<$ .0001
Skewness	2.105	0.0353	0.2168	$<$ .0001
Kurtosis	4.348	$<$ .0001	0.3012	$<$ .0001
Coefficient of variation	-4.845	$<$ .0001	0.2033	$<$ .0001
Brasília				
Standard deviation	16.994	$<$ .0001	0.8115	$<$ .0001
Variance	16.996	$<$ .0001	0.8117	$<$ .0001
Skewness	4.729	$<$ .0001	0.2684	$<$ .0001
Kurtosis	-6.315	$<$ .0001	0.2857	$<$ .0001
Coefficient of variation	15.653	$<$ .0001	0.797	$<$ .0001
Caxias Do Sul				
Standard deviation	16.656	$<$ .0001	0.8024	$<$ .0001
Variance	16.656	$<$ .0001	0.8024	$<$ .0001
Skewness	0.503	0.6151	0.1013	0.198
Kurtosis	-3.164	0.0016	0.3012	$<$ .0001
Coefficient of variation	-4.845	$<$ .0001	0.1375	0 .0280
São Luís				
Standard deviation	9.224	$<$ .0001	0.4354	$<$ .0001
Variance	9.224	$<$ .0001	0.4354	$<$ .0001
Skewness	-6.977	$<$ .0001	0.4234	$<$ .0001
Kurtosis	-7.772	$<$ .0001	0.3749	$<$ .0001
Coefficient of variation	3.781	0 .0002	0.2845	$<$ .0001

Table 3: 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.

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 variance, the kurtosis, and the coefficient of variation. The skewness is significant at 5%. In Brasília, these differences are statistically significant at 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 variance, and the coefficient of variation. In turn, regarding the Kolmogorov-Smirnov test, the coefficient of variation is only statistically significant at the 5% level. The kurtosis is statistically significant at 5% level for the Mann-Whitney test and 1% level for the Kolmogorov-Smirnov test. However, the skewness is not statistically significant at the 5% level. In São Luís, only the coefficient of variation is not statistically significant at the 1% level for the Mann-Whitney test.

4. The Supervised Machine Learning Algorithms

We evaluate the predictions based on several machine learning methods and assess the out of sample performance. In order to avoid overfitting, we use both the Lasso and Ridge regularized versions of the logit model (Tibshirani, 1996). Random Forest consists of a large number of individual decision trees that operate as an ensemble. Its foundation is based on the so-called wisdom of crowds. In other words, the Random Forest algorithm uses a large number of relatively uncorrelated models (trees) operating as a committee capable of outperforming any of the individual constituent models (Ho, 1995; Breiman, 2001). Neural Networks are models inspired by the human brain that process information in a parallel fashion and are useful tools for clustering and classifying data (Hjort, 1996; Ripley, 2007). We use both the cross-validation and random splitting approaches to split the database between the training and testing data.[11] 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. Figure 1 illustrates our classification modeling.

Figure 1: 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.

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.

[11]The training sample estimates the model parameters for a given city and contains 75% of the total of observations. The testing sample calculates the out-of-sample predictions and consists of 25% 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. Typically, this is the standard strategy for determining the optimal penalty level both for the Lasso and the Ridge Logistic regressors. To parsimoniously assess the trade-off between bias and variance, we repeat these steps 100 times to estimate classifiers' accuracy.

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; an area of 0.5 represents a low-quality classifier.[12]

4.1. 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. Decision trees are 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 Figure 1 $-$ 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

[12] 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.

training algorithm for random forests applies the general technique of bootstrap aggregating[13], or bagging, to tree learners. Given a training set $X = x_1, ..., x_n$ with responses $Y = y_1, ..., y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to the following samples:

For $b = 1, ..., B$:

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':

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(x'), \tag{5}$$

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':

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B} (f_b(x') - \hat{f})^2}{B - 1}} \tag{6}$$

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

[13]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.

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.[14]. From the input vector of features, the random forest algorithm selects the standard deviation, variance, and coefficient of variation as predictors of cartel behavior in Belo Horizonte, Brasília, and Caxias do Sul. Only in São Luís, the selected features are the standard deviation and the coefficient of variation.

4.2. Lasso and ridge logistic regression

Lasso and Ridge imposes a penalty term to the logistic regression model. Typically, the logistic regression with binary classifier is given by:

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

where $P(y_i = 1)$ is the probability of detecting collusive behavior periods in the data. 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,

[14] 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)

we obtain the parameters estimates as follows:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \Big[y_i \, x_i \, \beta - log(1 + e^{x_i \, \beta}) \Big], \tag{7}$$

Comparing equations (7) and (8), we note that the Ridge Logistic Regression 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 (Pereira et al., 2016):

$$\mathcal{L}_{\lambda}^{ridge}(\beta) = \sum_{i=1}^{n} \Big[y_i \, x_i \, \beta - log(1 + e^{x_i \, \beta}) \Big] - \lambda \sum_{j=1}^{p} \beta_j^2. \tag{8}$$

In summary, Ridge Logistic Regression 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. In summary, 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 Logistic Regression 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 Logistic Regression model. Hastie et al. (2009) introduces the penalized version of the log-likelihood function to be maximized as follows:

$$\mathcal{L}_{\lambda}^{lasso}(\beta) = \sum_{i=1}^{n} \Big[y_i \, x_i \, \beta - log(1 + e^{x_i \, \beta}) \Big] - \lambda \sum_{j=1}^{p} |\beta_j|. \tag{9}$$

The Lasso Logistic regression 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. By doing so, 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.[15]

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:

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

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 $i = 1, ...M_0$ represents the inputs of the neural network. Regarding the target variable, we have that $y_i^{nl} = y_i^0$, in which $i = 1, ...M_{nl}$ represents the output of the neural network. Thus, the neural network has $M_{nl} = M_0$ outputs. In our study, the inputs

[15] 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 Regression by using the training sample. Finally, we predict the cartel probability in the testing sample.

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 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 data set is unbalanced (Fawcett, 2006; Sammut & 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 497 weeks (observations), of which 221 labeled as the cartel period. The testing sample for the machine learning algorithms performances includes 25% 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 Table 4, we evaluate the average of the predictions based on 125 observations.

Confusion Matrix - Random Forest

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	66	1
	cartel (1)	0	58

Confusion Matrix - Lasso Logistic

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	63	4
	cartel (1)	12	46

Confusion Matrix - Neural Networks

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	65	2
	cartel (1)	5	53

Confusion Matrix - Ridge Logistic

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	59	08
	cartel (1)	48	10

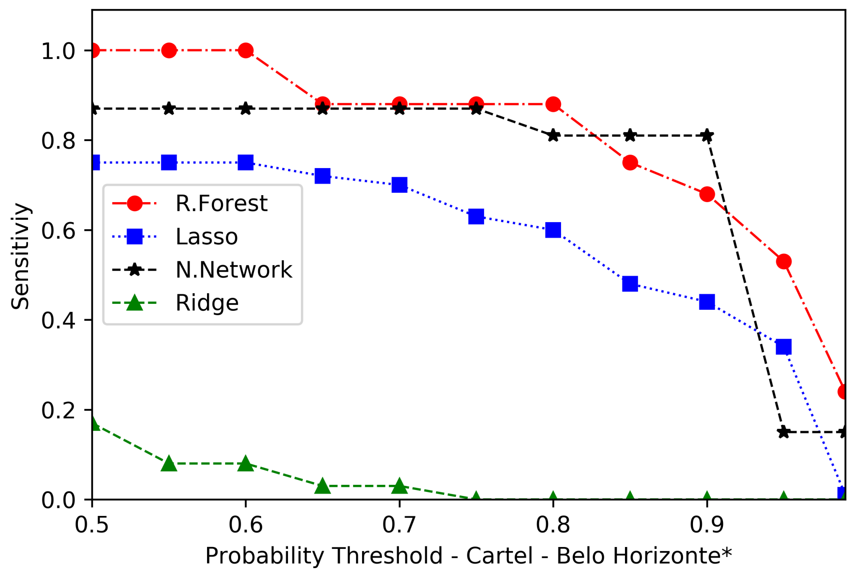
Table 4: Confusion Matrix for the machine learning classifiers considering a classification threshold equal to 0.5 - Belo Horizonte. We repeat the classification procedure 100 times and the values are based on the average of each of the metrics computed from the confusion matrix.

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.

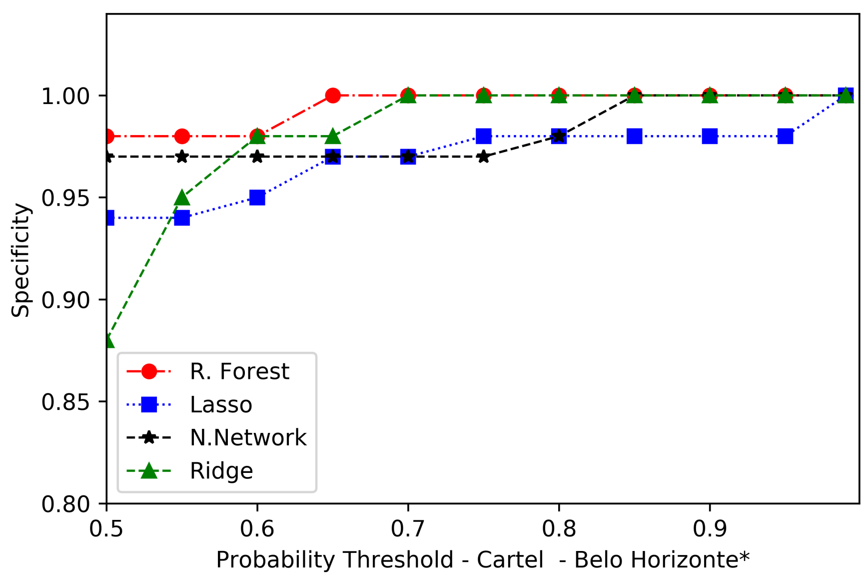
By looking at the Random Forest predictions, we see on the bottom right the number of true positives, which indicates that in 58 cases the classifier correctly predicted the cartel period. On the upper left, we observe the number of true negatives, which indicates that in 66 cases the classifier correctly predicted the non-cartel periods. On the upper right, we have the number of false-positives. Note that it indicates that only in 1 case the classifier incurred a Type I error. However, on the bottom left we see that the random forest 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 $(66 + 58)/125 = 99.2\%$. When comparing the random forest score with the results derived from the confusion matrices of the other algorithms, we see that the closest score is that of the neural network. The model with the lowest classification score was the Ridge Logistic Regression, where $(59+10)/125 = 55.2\%$. As well, 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 random forest misclassification error is given by $(0+1)/125 = 0.8\%$. This is the smallest classification error observed for Belo Horizonte. In contrast, for the Ridge Logistic Regression we have an error given by $(48 + 8)/125 = 44.8\%$.

Figure 2 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.



(a)



(b)

Figure 2: Sensitivity and Specificity results by restricting the cartel classification rule.

In the confusion matrix, sensitivity is calculated by dividing the true positives by the total of the bottom row. For the random forest classifier, we have that the sensitivity is equal to $58/(0+58) = 100\%$. In contrast, note that the Ridge Logistic Regression classifier has the lowest true positive rate. Specificity is calculated by dividing the true negatives by the total amount in the top row. Hence, for the random forest classifier, the Specificity measure is given by $66/(66+1) = 98.5\%$.

The Ridge Logistic Regression has the lowest true negative values $59/(59+8) = 88\%$. Finally, 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 Random Forest classifier has a precision equals to $58/(58+1) = 98.3\%$. The performance of the Neural Network is $53/(53+2) = 96.3\%$. Lasso and Ridge show reasonable precision rates, but relatively smaller than the others.

As expected, through Figures 2a and 2b, the correct classification rate in non-cartel periods increases in the probability threshold for the decision rule (false-positive results decrease). In contrast, the correct classification rate in cartel periods deteriorates much faster in the threshold (false-negative results increase). In other words, the antitrust agency would be able to minimize the false-positive rates ($1-$ specificity) by increasing the decision rule threshold to a value closer to 0.7.

In this scenario, the performance of the Random Forest and Ridge Logistic Regression predictors allows for minimal risk of false-positives outcomes. As well, for the Ridge algorithm, we must observe that the 0.7 classification rule, leads to a false-negative rate ($1-$ sensitivity) closer to 1. In contrast, the Neural Network and the Random Forest classifiers show approximately 15% of false-negative outcomes. In summary, the gain of reducing the risk of false-positives, therefore, induces a disproportionate increase in false-negatives.

Moreover, any further tightening of the decision rule would lead to an even more severe increase of false-negatives. At a probability threshold of 0.8, Random Forest shows the best performance. It, therefore, seems that for the gasoline cartel in Belo Horizonte, the best-suited probability threshold lies between values of 0.5 and 0.7. One advantage of combining screening methods and machine learning consists of quantifying the trade-off regarding false-positives and false-negatives so that the regulators are capable to determine the decision rule that optimally matches their needs.

We conclude the performance of our binary classifiers by assessing the area under the curve (AUC) metrics. It provides useful information regarding how well the classifiers are separating the cartel periods from the non-cartel periods. In general, 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, the better the classifier. As Table 8 reveals, the Random Forest predictor has the greater AUC.

5.2. Brasília

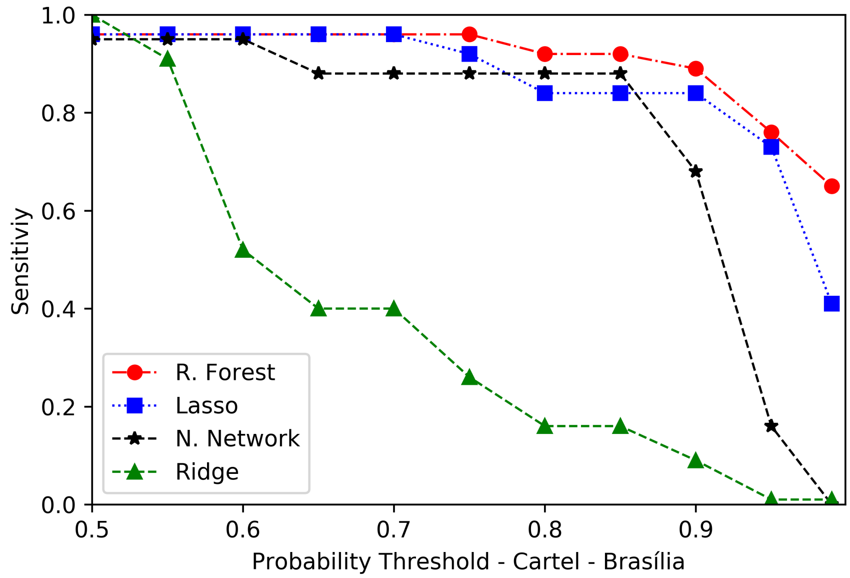
Out of a total of 488 weeks, the observations labeled as a cartel period in Brasilia represent 63% of the total sample. The testing sample for evaluating the classifiers contains 122 observations. Differently from the previous case, Table 5 reveals that the Lasso Regression shows the best score index $(44 + 73)/122 = 95.9\%$. Besides, it presents a classification error equals to $(3 + 2)/122 = 4.1\%$. The Random Forest algorithm also shows a reasonable performance. In terms of sensitivity and specificity, when considering a classification threshold equals to 0.5, the Lasso Regression classifier shows the best prediction outcomes. The true positive and true negative rates are given by $73/(2 + 73) = 97.3\%$ and $44/(44 + 3) = 93.6\%$. The precision index of the Lasso Regression is slightly higher $73/(73+6) = 96.1\%$ than the Random Forest.

Confusion Matrix - Random Forest				Confusion Matrix - Lasso Logistic			
		Predicted				**Predicted**	
		Non-cartel (0)	cartel (1)			Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	44	3	**Actual**	Non-cartel (0)	44	3
	cartel (1)	3	72		cartel (1)	2	73

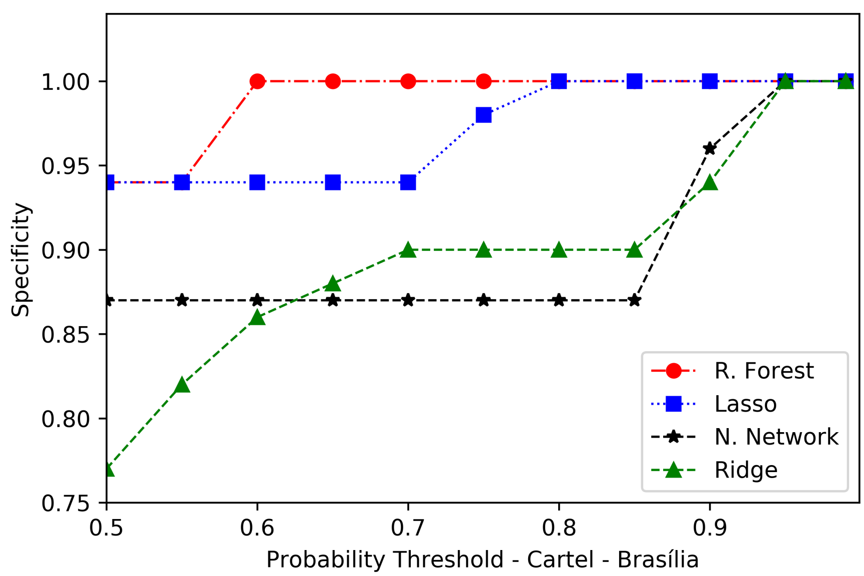
Confusion Matrix - Neural Networks				Confusion Matrix - Ridge Logistic			
		Predicted				**Predicted**	
		Non-cartel (0)	cartel (1)			Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	40	7	**Actual**	Non-cartel (0)	33	16
	cartel (1)	4	71		cartel (1)	0	73

Table 5: Confusion Matrix for the machine learning classifiers considering a classification threshold equal to 0.5 - Brasília. We repeat the classification procedure 100 times and the values are based on the average of each of the metrics computed from the confusion matrix.

Figure 3 summarizes the trade-offs in classification accuracy for the gasoline cartel in Brasília. As expected, through Figures 3a and 3b, the observe that false-positive results decreases in the probability threshold. However, the false-negative rate in cartel periods increases much faster in the threshold. To minimize the false-positive rates ($1-$ specificity), the optimum decision rule threshold should be greater than 0.6. In this case, the performance of the Random Forest minimizes the false-positive rate. Ridge Logistic Regression predictors allow for minimal risk of false-positives outcomes. This same condition is true for Lasso Regression when the threshold is greater than 0.8. Yet, we must observe that a classification rule greater than 0.75, leads to a false-negative rate ($1-$ sensitivity) closer to 10% for the Random Forest. The Lasso predictors show approximately 15% of false-negative outcomes.



(a)



(b)

Figure 3: Sensitivity and Specificity results by restricting the cartel classification rule.

As before, the benefits of reducing the risk of false-positives is unreasonable for the increase in false-negatives. Therefore, at a probability threshold of 0.5, Lasso Regression shows the best performance. When we increase the decision rule by considering a threshold greater than 0.75, Random Forest proves to be the best algorithm for classifying the gasoline cartel in Brasília. Judging by the AUC criterion, both predictors have a satisfactory classification rate, but the Ridge predictor shows the best performance in this regard (AUC = 88.3%). On the other hand, taking into account all the evaluation metrics, from Table 8, we can conclude that LASSO regression, on average, performs subtly better than Random Forest.

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 25% (121 observations) are used for testing the classifiers. The confusion matrix in Table 6 shows that the Random Forest provides the best score index $(75 + 40)/121 = 95\%$. The classification error is given by equals to $(2 + 4)/121 = 5\%$. Considering a classification threshold equal or greater than 0.5, the Random Forest shows the best prediction outcomes. For a probability decision rule equals 0.5, the true positive and true negative rates are given by $40/(4 + 40) = 90.9\%$ and $75/(75 + 2) = 97.4\%$. The precision index of the Ridge Logistic Regression model is the largest $4/(0 + 4) = 100\%$.

Figure 4 illustrates how sensitivity and specificity react to an increase in the probability threshold. By comparing the outcomes represented in Figures 4a and 4b, we see that the false-negative rate ($1-$ sensitivity) is closer to 100% for the Ridge algorithm. The Random Forrest predictors show approximately 15% of false-negative outcomes for a threshold probability equal or lower than 0.65. When we narrow the decision rule, especially assuming values greater than 0.75 we affect both the sensitivity and the specificity of the classification algorithms. Note that for the antitrust authority, it is not interesting to adopt the Ridge model to identify the cartel in Caxias do Sul.

Confusion Matrix - Random Forest

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	75	2
	cartel (1)	4	40

Confusion Matrix - Lasso Logistic

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	73	4
	cartel (1)	7	37

Confusion Matrix - Neural Networks

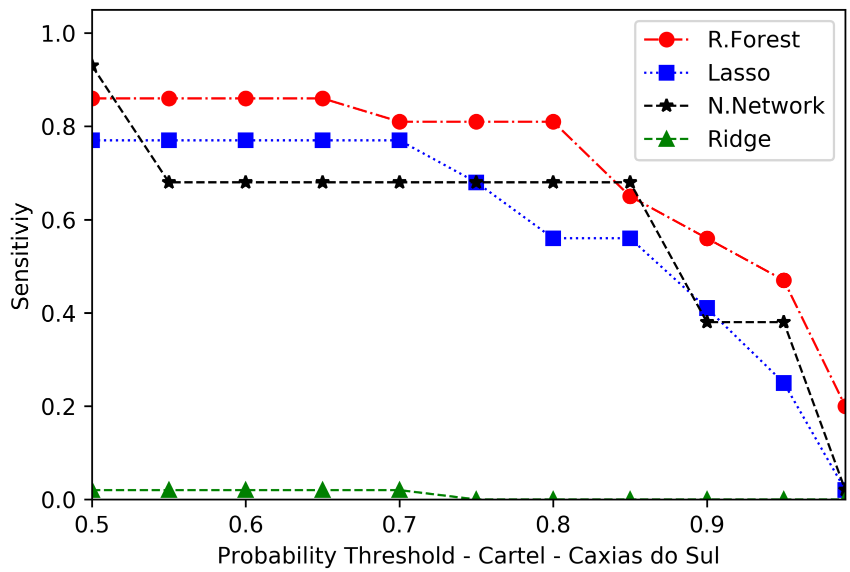
		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	73	4
	cartel (1)	3	41

Confusion Matrix - Ridge Logistic

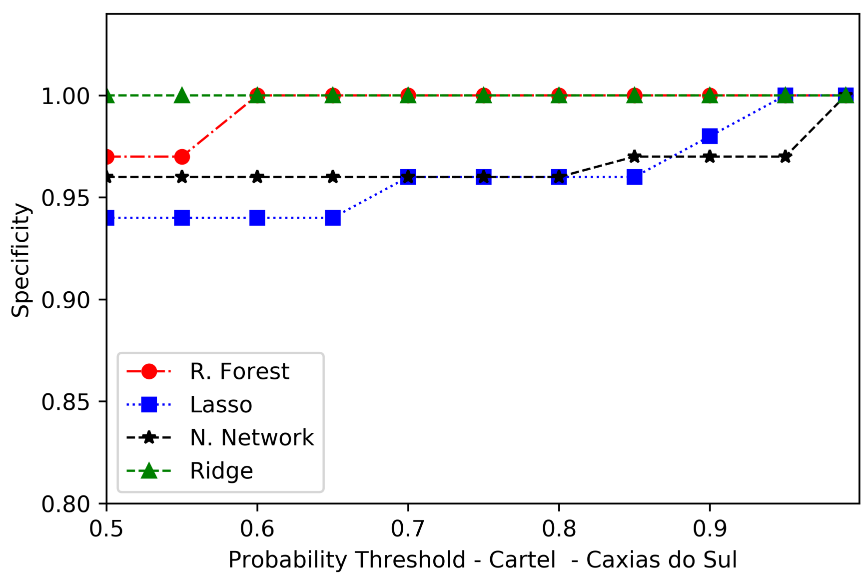
		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	77	0
	cartel (1)	40	4

Table 6: Confusion Matrix for the machine learning classifiers considering a classification threshold equal to 0.5 - Caxias do Sul. We repeat the classification procedure 100 times and the values are based on the average of each of the metrics computed from the confusion matrix.

Note that for the antitrust authority, it is not interesting to adopt the Ridge model to identify the cartel in Caxias do Sul. In other words, a high specificity rate is not a sufficient condition to minimize classification errors. To prove this, we assess the (poor) performance of the Ridge model incorrectly classifying observations as a cartel period (sensitivity). Thus, the classifier that best responds to the data - indicating how many observations were correctly identified as a cartel period (sensitivity) and how many observations were correctly identified as a non-cartel period (specificity) - is the Random Forest.



(a)



(b)

Figure 4: Sensitivity and Specificity results by restricting the cartel classification rule.

In a complementary way, we see that by the AUC criterion as in Table 8, we also conclude that, on average, the Random Forest estimators show the best performance in predicting the gasoline cartel in Caxias do Sul. It is also worth noting that, on average, the neural network performed better than the LASSO and Ridge Logistic Regressions.

5.4. São Luís

São Luís has a total sample of 451 weeks. The period labeled as cartel behavior contemplates 48% of this amount. We use 113 observations in order to compare the machine learning classifiers. The models that best classify the cartel in São Luís are Random Forest and Neural Networks, respectively.

Confusion Matrix - Random Forest

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	61	3
	cartel (1)	3	46

Confusion Matrix - Lasso Logistic

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	57	7
	cartel (1)	14	35

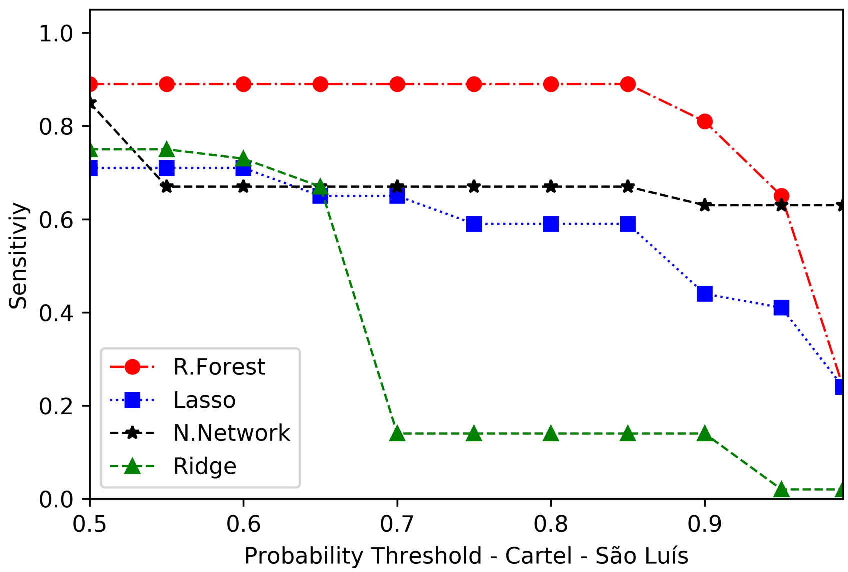
Confusion Matrix - Neural Networks

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	61	3
	cartel (1)	7	42

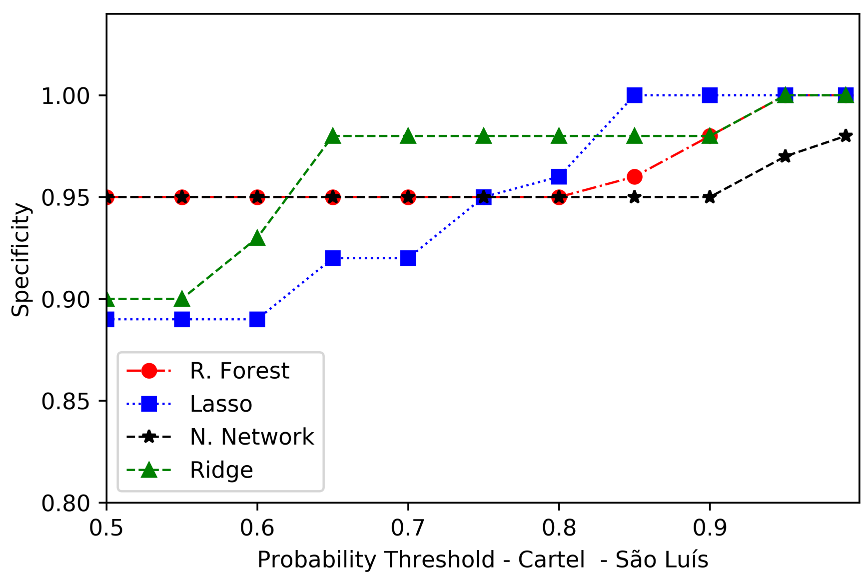
Confusion Matrix - Ridge Logistic

		Predicted	
		Non-cartel (0)	cartel (1)
Actual	Non-cartel (0)	58	6
	cartel (1)	11	38

Table 7: Confusion Matrix for the machine learning classifiers considering a classification threshold equal to 0.5 - São Luís. We repeat the classification procedure 100 times and the values are based on the average of each of the metrics computed from the confusion matrix.



(a)



(b)

Figure 5: Sensitivity and Specificity results by restricting the cartel classification rule.

Using the confusion matrix available in Table 7, we see that the power of Lasso regression, whose classification error $(14+7)/113$ is approximately 18.6%, is almost three times greater than the error calculated for Random Forest $(3+3)/113 = 5.03\%$. For Neural Network this measure is equal to $(7+3)/113 = 8.9\%$. Ridge Logistic Regression presents a classification error given by $(11+6)/113 = 15.1\%$.

Regarding the precision measure, the Random Forest shows the best performance in relation to the true positives cartel observations $46/(46+3) = 93.87\%$. The model that comes closest to this rate is the Neural Network $(42/42+3) = 93.3$. The proximity between the quality of the predictions of both models remains for all the other statistics. Thus, considering the probability threshold equals to 0.5 and judging by the set of measures, on average, we observe from Figure 5 that Random Forest is more accurate.

For a decision rule less than or equal to 0.95, as reported in Figure 5a, the Random Forest algorithm shows the lowest rate for false-negative outcomes. Withal, there is a scenario in which the sensitivity of Neural Network is equivalent to that of Random Forest. Regarding specificity, for a threshold between 0.6 and 0.8, the Ridge Logistic Regression presents better performance, being surpassed by Lasso Regression for intervals between 0.8 and 0.95. At this point, by Figure 5b, we see that the Ridge, Lasso, and Random Forest models are equivalent regarding the proportion of actual negatives that are correctly identified. Only the Neural Network has a slightly lower performance.

6. Antitrust Authorities and Competition Policy

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 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, we believe 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 & Ye, 2003), our approach does require firm-level cost information. Also, daily price databases are not necessary conditions for the antitrust agency to detect collusion via behavioral screening methods.

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 & 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 & 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, not being restricted only to bid-rigging cartels.

	Null Accuracy (%)	Score (%)	Error (%)	Precision (%)	AUC (%)
Belo Horizonte					
Random Forest	46.40	99.20	0.80	98.30	99.90
Lasso Logistic	46.40	87.20	12.80	92.00	94.30
Neural Networks	46.40	94.40	5.60	96.30	89.90
Ridge Logistic	46.40	55.20	44.80	55.50	70.80
Brasília					
Random Forest	61.50	95.10	4.90	96.00	86.80
Lasso Logistic	61.50	95.90	4.10	96.00	86.90
Neural Networks	61.50	90.90	8.20	91.00	85.90
Ridge Logistic	61.50	86.90	13.10	82.00	88.30
Caxias do Sul					
Random Forest	36.30	95.00	5.00	95.20	98.40
Lasso Logistic	36.30	90.90	9.10	90.20	94.50
Neural Networks	36.30	94.20	5.80	91.10	96.20
Ridge Logistic	36.30	66.90	33.10	100.00	73.80
São Luís					
Random Forest	43.30	94.60	5.40	93.80	98.60
Lasso Logistic	43.30	81.40	18.60	83.30	90.10
Neural Networks	43.30	91.10	8.90	93.30	96.60
Ridge Logistic	43.30	84.90	15.10	86.40	91.00

Table 8: 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.

Concerning our case study, we recommend our screens by adopting some practices, as follows. First, the coefficient of variation and the standard deviation reveals to be the most powerful predictors. In this way, they help us to infer the negative relationship between the variance of the retail price of gasoline and the cartel probability. Therefore, low price variance suggests a higher likelihood of a cartel(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. 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 piece of 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. This practice might reduce incorrect predictions among truly non-cartel periods (false-positives) at the expense of increasing the number of actual cartel periods (false-negatives).

7. Conclusion and Policy Implications

In this paper, we combined many different supervised machine learning techniques with statistical screens based on the gasoline retail price distribution to predict collusion. Considering an average of the overall accuracy, the models correctly predicted around 87% of the cartel periods. Comparing all the four models, we highlight their predictive efficiency according to the following ranking: Random Forest, Lasso Logistic, Neural Network, and Ridge Logistic Regression. Considering all cities, the Random Forrest algorithm, on average, showed a score of 95% correct classifications − for both cartel and non-cartel periods. Even increasing the probability threshold, the Random Forest algorithm remains the most stable classifier model regarding sensitivity and specificity.

We also found evidence that both asymmetry and kurtosis are features that increases the algorithms' performance. These inputs work in a complementary way - or can even replace variance 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 showed that supervised machine learning classifiers have many positive attributes and can provide valuable contributions in detecting and fighting cartels. In contrast, we emphasize the costs and damage to the antitrust authority's reputation, inherent in the trade-off between reducing false-positives *vs.* false-negatives.

An extension of this paper would be to establish an approximation between the Edgeworth price cycle approach, passthrough of upstream cost shocks, response asymmetry, and variance screens as discussed in Eckert (2013), with machine learning algorithms. Finally, a fruitful avenue for future research would involve assessing the role of the dual-fuel system in Brazil, which includes ethanol in addition to gasoline, in terms of the construction of cartel screens.

Acknowledgements

The authors thank Professor Emilson C.D. Silva for the insightful suggestions and comments. Daniel O. Cajueiro is indebted to CNPQ for partial financial support under grant 302629/2019-0.

References

References

Abrantes-Metz, R. M. (2012). Screens for conspiracies and their multiple applications. *COMPETITION POLICY INTERNATIONAL*, *8*.

Abrantes-Metz, R. M., Froeb, L. M., Geweke, J., & Taylor, C. T. (2006). A variance screen for collusion. *International Journal of Industrial Organization*, *24*, 467–486.

Akouemo, H. N., & Povinelli, R. J. (2016). Probabilistic anomaly detection in natural gas time series data. *International Journal of Forecasting*, *32*, 948–956.

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. *The Review of Economic Studies*, *71*, 317–349.

Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. *Review of Economics and statistics*, *85*, 971–989.

Blair, R. D., & Sokol, D. D. (2015). *The Oxford handbook of international antitrust economics* volume 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. *German Economic Review*, *13*, 245–256.

Bolotova, Y., Connor, J. M., & Miller, D. J. (2008). The impact of collusion on price behavior: Empirical results from two recent cases. *International Journal of Industrial Organization*, *26*, 1290–1307.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

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. *The BE Journal of Economic Analysis & Policy*, *7*.

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

Clark, R., & Houde, J.-F. (2014). The effect of explicit communication on pricing: Evidence from the collapse of a gasoline cartel. *The Journal of Industrial Economics*, *62*, 191–228.

Connor, J. M. (2005). Collusion and price dispersion. *Applied Economics Letters*, *12*, 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*, 219–228.

Crede, C. J. (2019). A structural break cartel screen for dating and detecting collusion. *Review of Industrial Organization*, *54*, 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]. *Revista de Economia Contemporânea*, *23*.

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 Economics*, *43*, 11–21.

Doane, M. J., Froeb, L. M., Sibley, D. S., & Pinto, B. P. (2015). Screening for collusion as a problem of inference. *Oxford Handbook of International Antitrust Economics*, *2*, 523–553.

Eckert, A. (2013). Empirical studies of gasoline retailing: A guide to the literature. *Journal of Economic Surveys*, *27*, 140–166.

Eckert, A., & West, D. S. (2004). Retail gasoline price cycles across spatially dispersed gasoline stations. *The Journal of Law and Economics*, *47*, 245–273.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, *27*, 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*, 89–113.

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

Harrington, J. E. (2005). *Detecting cartels*. Technical Report Working Papers, The Johns Hopkins University, Department of Economics.

Harrington, J. E. (2008). Optimal corporate leniency programs. *The Journal of Industrial Economics*, *56*, 215–246.

Harrington, J. E., & Chen, J. (2006). Cartel pricing dynamics with cost variability and endogenous buyer detection. *International Journal of Industrial Organization*, *24*, 1185–1212.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hjort, N. (1996). *Pattern recognition and neural networks*. Cambridge university press.

Ho, T. K. (1995). Random decision forests. In *Document analysis and recognition, 1995., proceedings of the third international conference on* (pp. 278–282). IEEE volume 1.

Huber, M., & Imhof, D. (2018). *Machine learning with screens for detecting bid-rigging cartels*. Technical Report Université de Fribourg.

Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization*, *65*, 277–301.

Imhof, D. (2017). *Simple statistical screens to detect bid rigging*. Technical Report Université de Fribourg.

Jiménez, J. L., & Perdiguero, J. (2012). Does rigidity of prices hide collusion? *review of industrial organization*, *41*, 223–248.

Maskin, E., & Tirole, J. (1988). A theory of dynamic oligopoly, ii: Price competition, kinked demand curves, and edgeworth cycles. *Econometrica: Journal of the Econometric Society*, (pp. 571–599).

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 Nº004/2019, Departamento de Estudos Econômicos - Conselho Administrativo de Defesa Econômica*, .

Noel, M. D. (2007). Edgeworth price cycles, cost-based pricing, and sticky pricing in retail gasoline markets. *The Review of Economics and Statistics*, *89*, 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*, .

Perdiguero, J., & Jiménez, J. L. (2020). Price coordination in the spanish oil market: the monday effect. *Energy Policy*, (p. 112016).

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

Pinha, L. C., Braga, M. J. et al. (2019). Evaluating the effectiveness of the brazilian leniency program. *Economics Bulletin*, *39*, 1860–1869.

Porter, R. H. (2005). Detecting collusion. *Review of Industrial Organization*, *26*, 147–167.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, .

Resende, M. (2012). Sincronização de preços no mercado de varejo para combustíveis: Uma nota empírica. *Revista Brasileira de Energia*, *18*, 197–204.

Ripley, B. D. (2007). *Pattern recognition and neural networks*. Cambridge university press.

Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 267–288).

Wang, Z. (2009). (mixed) strategy in oligopoly pricing: Evidence from gasoline price cycles before and under a timing regulation. *Journal of Political Economy*, *117*, 987–1030.

Wilmot, N. A., & Mason, C. F. (2013). Jump processes in the market for crude oil. *The Energy Journal*, *34*.

Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature*, *50*, 731–69.